Near-Maximum-Likelihood Detection of MIMO Systems using MMSE-Based Lattice-Reduction

Dirk Wübben, Ronald Böhnke, Volker Kühn, and Karl-Dirk Kammeyer
Department of Communications Engineering
University of Bremen
Otto-Hahn-Allee, D-28359 Bremen, Germany
Email: {wuebben, boehnke, kuehn, kammeyer}@ant.uni-bremen.de

***Abstract*— In recent publications the use of lattice-reduction for signal detection in multiple antenna systems has been proposed. In this paper, we adopt these lattice-reduction-aided schemes to the MMSE criterion. We show that an obvious way to do this is infeasible and propose an alternative method based on an extended system model, which in conjunction with simple successive interference cancellation nearly reaches the performance of maximum-likelihood detection. Furthermore, we demonstrate that a sorted QR decomposition can significantly reduce the computational effort associated with lattice-reduction. Thus, the new algorithm clearly outperforms existing methods with comparable complexity.**

***Index Terms*— MIMO systems, BLAST, Zero-Forcing and MMSE detection, lattice-reduction, wireless communication.**

I. Introduction

It is well-known that multiple antenna systems may provide very high data rates in rich scattering environments. In the famous V-BLAST architecture, parallel data streams are transmitted over n_T different antennas. Besides linear detection schemes based on the zero-forcing (ZF) or the minimum mean square error (MMSE) criterion, successive interference cancellation (SIC) is a popular way to detect the transmitted signals at the receiver site [1]. Unfortunately, for ill-conditioned channel matrices all these schemes are clearly inferior to maximum-likelihood (ML) detection. The latter may be accomplished by sphere detection (SD), which is an ongoing research topic [2]. However, SD requires a closest point search for each transmitted vector, which still is rather demanding. In mobile communication scenarios, where the channel remains constant for several symbol durations, it is much more preferable to spend most of the computational effort only once at the beginning of each frame. Recently, lattice-reduction (LR) has been proposed in order to transform the system model into an equivalent one with better conditioned channel matrix prior to low-complexity linear or SIC detection [3], [4]. These publications exclusively deal with ZF filtering for symbol detection. In the present work we extend the LR-aided detection schemes with respect to the MMSE criterion. To this end, we make use of an extended system model introduced in [5] and further investigated in [6], [7].

This work was supported in part by the German ministry of education and research (BMBF) under grant 01 BU 153.

The remainder of this paper is organized as follows. In Section II, the system model and notation are introduced. The fundamentals of LR are explained in Section III and different detection schemes with and without reduction of the basis are introduced in Section IV. A performance analysis is given in Section V and concluding remarks can be found in Section VI.

II. System description

We consider a multiple antenna system with n_T transmit and $n_R \geq n_T$ receive antennas. The data is demultiplexed into n_T data substreams (called layers). These substreams are mapped onto M-QAM symbols and transmitted over the n_T antennas simultaneously.

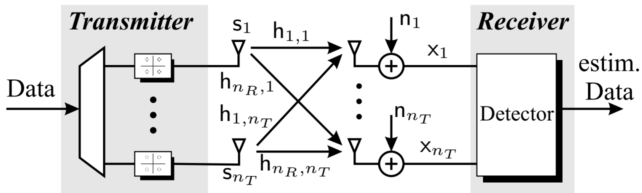

Fig. 1. Model of a MIMO system with n_T transmit and n_R receive antennas.

In order to describe the MIMO system, one time slot of the time-discrete complex baseband model is investigated. Let[1] $\mathbf{s}$ denote the complex valued $n_T \times 1$ transmit signal vector, then the corresponding $n_R \times 1$ receive signal vector $\mathbf{x}$ is given by

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

In (1), $\mathbf{n}$ represents white gaussian noise of variance σ_n^2 observed at the n_R receive antennas while the average transmit power of each antenna is normalized to one, i.e. $\mathrm{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}_{n_T}$ and $\mathrm{E}\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma_n^2 \mathbf{I}_{n_R}$. The $n_R \times n_T$ channel matrix $\mathbf{H}$ contains uncorrelated complex gaussian fading gains with unit variance. We assume a flat fading environment, where the channel matrix $\mathbf{H}$ is constant over a frame and changes independently from frame to frame (*block fading channel*).

[1]Throughout this paper, $(\cdot)^T$ and $(\cdot)^H$ denote matrix transpose and hermitian transpose, respectively. Furthermore, $\mathbf{I}_\alpha$ indicates the $\alpha \times \alpha$ identity matrix and $\mathbf{0}_{\alpha,\beta}$ denotes the $\alpha \times \beta$ all zero matrix. With $\mathcal{R}\{\cdot\}$ and $\mathcal{I}\{\cdot\}$ we denote the real part and the imaginary part, respectively.

The distinct fading gains are assumed to be perfectly known by the receiver. Treating real and imaginary part of (1) separately, the system model can be rewritten as

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \,, \tag{2}$$

with the real-valued channel matrix

$$\mathbf{H} = \begin{bmatrix} \mathcal{R}\{\mathsf{H}\} & -\mathcal{I}\{\mathsf{H}\} \\ \mathcal{I}\{\mathsf{H}\} & \mathcal{R}\{\mathsf{H}\} \end{bmatrix} \in \mathbb{R}^{n \times m} \tag{3}$$

and the real-valued vectors

$$\mathbf{x} = \begin{bmatrix} \mathcal{R}\{\mathsf{x}\} \\ \mathcal{I}\{\mathsf{x}\} \end{bmatrix} \quad \mathbf{s} = \begin{bmatrix} \mathcal{R}\{\mathsf{s}\} \\ \mathcal{I}\{\mathsf{s}\} \end{bmatrix} \quad \mathbf{n} = \begin{bmatrix} \mathcal{R}\{\mathsf{n}\} \\ \mathcal{I}\{\mathsf{n}\} \end{bmatrix} . \tag{4}$$

By defining $m = 2\,n_T$ and $n = 2\,n_R$ the dimension of the real channel matrix (3) is given by $n \times m = 2\,n_R \times 2\,n_T$. Likewise the dimension of the vectors (4) are given by $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{n} \in \mathbb{R}^n$ and $\mathbf{s} \in \mathcal{A}^m$, where $\mathcal{A}$ denotes the finite set of real-valued transmit signals. For M-QAM this set is given by $\mathcal{A} = \{\pm\frac{1}{2}a, \pm\frac{3}{2}a, \ldots, \pm\frac{\sqrt{M}-1}{2}a\}$ with $\sqrt{M}$ representing the modulation index of the corresponding real-valued ASK. The parameter $a = \sqrt{6/(M-1)}$ is used for normalizing the power of the complex valued transmit signals to 1. In the sequel we will apply this real-valued representation, as we can now interpret each noiseless receive signal as a point of a lattice spanned by $\mathbf{H}$. Additionally, the performance of successive algorithms like the V-BLAST detection can be improved by separating the real and imaginary part of each transmit signal [8].

The optimum maximum-likelihood (ML) detector searches over the whole set of transmit signals $\mathbf{s} \in \mathcal{A}^m$, and decides in favor of the transmit signal $\hat{\mathsf{s}}_{\text{ML}}$ that minimizes the euclidian distance to the receive vector $\mathbf{x}$, i.e.

$$\hat{\mathsf{s}}_{\text{ML}} = \arg\min_{\mathbf{s} \in \mathcal{A}^m} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2 \,. \tag{5}$$

As the computational effort is of order M^{n_T}, brute force ML detection is not feasible for larger number of transmit antennas or higher modulation schemes. A feasible alternative is the application of sphere detector (SD) [2], which restricts the search space to a sphere. However, the computational complexity is still high in comparison to simple but suboptimal successive interference cancellation (SIC). In the sequel, we investigate the application of lattice-reduction in order to improve the performance of SIC and linear detection. One advantage of this strategy is, that the computational overhead is only required once for each transmitted frame, so for large frame length the effort for each signal vector is very small.

III. Lattice Reduction

In the sequel, we interpret the columns $\mathbf{h}_\ell$ ($1 \le \ell \le m$) of the real-valued channel matrix $\mathbf{H}$ as the *basis* of a lattice and assume for the moment that the possible transmit vectors are given by $\mathbb{Z}^m$, the m dimensional infinite integer space. Consequently, the set of all possible undisturbed receive signals is given by the lattice

$$\mathcal{L}(\mathbf{H}) = L(\mathbf{h}_1, \ldots, \mathbf{h}_m) := \sum_{\ell=1}^{m} \mathbf{h}_\ell \mathbb{Z} \,. \tag{6}$$

The matrix $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{T}$ generates the *same* lattice as $\mathbf{H}$, if and only if the $m \times m$ matrix $\mathbf{T}$ is *unimodular* [9], i.e. $\mathbf{T}$ contains only integer entries and the determinant is $\det(\mathbf{T}) = \pm 1$:

$$\mathcal{L}(\tilde{\mathbf{H}}) = \mathcal{L}(\mathbf{H}) \Longleftrightarrow \tilde{\mathbf{H}} = \mathbf{H}\mathbf{T} \text{ and } \mathbf{T} \text{ is unimodular} \,. \tag{7}$$

The inverse of unimodular matrices always exists and contains also only integer values, i.e. $\mathbf{T}^{-1} \in \mathbb{Z}^m$. Obviously, the relation $\mathbf{H} = \tilde{\mathbf{H}}\mathbf{T}^{-1}$ holds.

For further investigations, we define the QR decomposition $\mathbf{H} = \mathbf{Q}\mathbf{R}$ with the $n \times m$ matrix $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_m]$ having orthogonal columns of unit length ($\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_m$) and the upper triangular matrix $\mathbf{R} = (r_{i,j})_{1 \le i,j \le m}$. Thus, each column vector $\mathbf{h}_k$ of $\mathbf{H}$ is given by $\mathbf{h}_k = \sum_{\ell=1}^{k} r_{\ell,k}\mathbf{q}_\ell$. The vector $\mathbf{q}_k$ denotes the direction of $\mathbf{h}_k$ perpendicular to the space spanned by $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$ and $r_{k,k}$ describes the corresponding length. Furthermore, $r_{\ell,k} = \mathbf{q}_\ell^T \mathbf{h}_k$ is the length of the projection of $\mathbf{h}_k$ onto $\mathbf{q}_\ell$. In the same way, the decomposition $\tilde{\mathbf{H}} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ is defined.

The aim of lattice-reduction is to transform a given basis $\mathbf{H}$ into a new basis $\tilde{\mathbf{H}}$ with vectors of shortest length or, equivalently, into a basis consisting of *roughly* orthogonal basis vectors. Usually, $\tilde{\mathbf{H}}$ is much better conditioned than $\mathbf{H}$. With respect to the QR decomposition, $\tilde{\mathbf{h}}_k$ is almost orthogonal to the space spanned by $\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_{k-1}$, if $|\tilde{r}_{1,k}|, \ldots, |\tilde{r}_{k-1,k}|$ are close to zero. An efficient (though not optimal) way to determine a reduced basis was proposed by Lenstra, Lenstra and Lovász [10].

Definition (Lenstra-Lenstra-Lovász-Reduced): A basis $\tilde{\mathbf{H}}$ with QR decomposition $\tilde{\mathbf{H}} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ is called *LLL-reduced with parameter* δ ($1/4 < \delta \le 1$), if [10]

$$|\tilde{r}_{\ell,k}| \le \frac{1}{2}|\tilde{r}_{\ell,\ell}| \quad \text{for} \quad 1 \le \ell < k \le m \tag{8}$$

and

$$\delta\,\tilde{r}_{k-1,k-1}^2 \le \tilde{r}_{k,k}^2 + \tilde{r}_{k-1,k}^2 \quad \text{for} \quad k = 2, \ldots, m \,. \tag{9}$$

If only (8) is fulfilled, the basis is called *size-reduced*. The parameter δ influences the quality of the reduced basis. Throughout this paper, we will assume $\delta = \frac{3}{4}$ as proposed in [10]. The whole LLL algorithm is shown in **Tab. 1**[2]. Given the QR decomposition of $\mathbf{H}$, it successively size-reduces the basis according to (8), exchanges two basis vectors if (9) is not fulfilled and adopts $\mathbf{T}$, $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{Q}}$. The output of the LLL algorithm is given by $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{R}}$, and $\mathbf{T}$.

Obviously, the complexity of the algorithm highly depends on the number of column exchanges, because in this case not only matrix multiplications are required, but also the counter k is decreased again. In the first step ($k = 2$), no exchange operation is necessary if $\delta\,\tilde{r}_{1,1}^2 \le \tilde{r}_{2,2}^2 + \tilde{r}_{1,2}^2$ holds. Consequently, $|\tilde{r}_{1,1}|$ should be as small as possible. Similar arguments hold for the remaining diagonal elements $\tilde{r}_{k,k}$. Therefore, the *Sorted QR Decomposition* (SQRD) introduced in [11] and extended

[2]Within the algorithm, $\mathbf{A}(a:b, c:d)$ denotes the submatrix of $\mathbf{A}$ with elements from rows $a, \ldots, b$ and columns $c, \ldots, d$. With $\lceil \alpha \rfloor$ we denote the nearest integer to α.

Tab. 1 LLL Lattice-Reduction Algorithm [10]

INPUT: $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{P}$ (default: $\mathbf{P} = \mathbf{I}_m$)
OUTPUT: $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{R}}$, $\mathbf{T}$

Initialization: $\tilde{\mathbf{Q}} := \mathbf{Q}$, $\tilde{\mathbf{R}} := \mathbf{R}$, $\mathbf{T} := \mathbf{P}$
$k = 2$
while $k \leq m$
  for $\ell = k-1, \ldots, 1$
    $\mu = \lceil \tilde{\mathbf{R}}(\ell,k) / \tilde{\mathbf{R}}(\ell,\ell) \rfloor$
    if $\mu \neq 0$
      $\tilde{\mathbf{R}}(1:\ell,k) := \tilde{\mathbf{R}}(1:\ell,k) - \mu\, \tilde{\mathbf{R}}(1:\ell,\ell)$
      $\mathbf{T}(:,k) := \mathbf{T}(:,k) - \mu\, \mathbf{T}(:,\ell)$
    end
  end
  if $\delta\, \tilde{\mathbf{R}}(k-1,k-1)^2 > \tilde{\mathbf{R}}(k,k)^2 + \tilde{\mathbf{R}}(k-1,k)^2$
    Swap columns $k-1$ and k in $\tilde{\mathbf{R}}$ and $\mathbf{T}$
    Calculate Givens rotation matrix $\mathbf{\Theta}$ such that element $\tilde{\mathbf{R}}(k,k-1)$ becomes zero:

$$\mathbf{\Theta} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \text{ with } \begin{array}{l} \alpha = \frac{\tilde{\mathbf{R}}(k-1,k-1)}{\|\tilde{\mathbf{R}}(k-1:k,k-1)\|} \\ \beta = \frac{\tilde{\mathbf{R}}(k,k-1)}{\|\tilde{\mathbf{R}}(k-1:k,k-1)\|} \end{array}$$

    $\tilde{\mathbf{R}}(k-1:k,k-1:m) := \mathbf{\Theta}\tilde{\mathbf{R}}(k-1:k,k-1:m)$
    $\tilde{\mathbf{Q}}(:,k-1:k) := \tilde{\mathbf{Q}}(:,k-1:k)\mathbf{\Theta}^T$
    $k := \max\{k-1, 2\}$
  else
    $k := k+1$
  end
end

in [6], [7] may provide a better starting point for the LLL algorithm than conventional QR decomposition techniques. SQRD successively minimizes $|r_{1,1}|, \ldots, |r_{m,m}|$ in the given order by permuting columns of $\mathbf{H}$, resulting in $\mathbf{QR} = \mathbf{HP}$ with a permutation matrix $\mathbf{P}$. The additional computational effort due to sorting was shown to be negligible [6]. We will see in Section V, that the application of SQRD prior to the LLL algorithm leads to a significant reduction of the computational complexity, as this decomposition already achieves a pre-sorting.

IV. Detection Algorithms

A. Common ZF and MMSE Detection Algorithms

In a zero-forcing (ZF) detector the interference is completely suppressed by multiplying the receive signal vector $\mathbf{x}$ with the Moore-Penrose pseudo-inverse of the channel matrix $\mathbf{H}^+ = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$. The decision step consists of mapping each element of the filter output vector

$$\tilde{\mathbf{s}}_{\text{ZF}} = \mathbf{H}^+\mathbf{x} = \mathbf{s} + \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{n} \tag{10}$$

onto an element of the symbol alphabet by a minimum distance quantization, which in case of M-QAM (after proper shifting and scaling) corresponds to a simple rounding operation and (if necessary) clipping to the allowed range of values. For an orthogonal channel matrix, ZF is identical to ML. However, in general ZF leads to noise amplification, which is especially observed in systems with the same number of transmit and receive antennas.

The minimum mean square error (MMSE) detector takes the noise term into account and thereby leads to an improved performance. As shown in [6], [7], MMSE detection is equal to ZF with respect to an extended system model. To this end, we define the $(n+m)\times m$ *extended* channel matrix $\underline{\mathbf{H}}$ and the $(n+m)\times 1$ *extended* receive vector $\underline{\mathbf{x}}$ by

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \sigma_n \mathbf{I}_m \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{0}_{m,1} \end{bmatrix}. \tag{11}$$

Then, the output of the MMSE filter can be written as

$$\tilde{\mathbf{s}}_{\text{MMSE}} = \left(\mathbf{H}^T\mathbf{H} + \sigma_n^2\mathbf{I}_m\right)^{-1}\mathbf{H}^T\mathbf{x} \tag{12}$$

$$= \left(\underline{\mathbf{H}}^T\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^T\underline{\mathbf{x}} = \underline{\mathbf{H}}^+\underline{\mathbf{x}}\,, \tag{13}$$

which exactly matches the structure of (10).

B. Lattice Reduction aided Linear Detection

As already mentioned, linear detection is optimal for an orthogonal channel matrix. Now, with $\tilde{\mathbf{H}} = \mathbf{HT}$ and the introduction of $\mathbf{z} = \mathbf{T}^{-1}\mathbf{s}$ the receive signal vector (2) can be rewritten as

$$\mathbf{x} = \mathbf{Hs} + \mathbf{n} = \mathbf{HTT}^{-1}\mathbf{s} + \mathbf{n} = \tilde{\mathbf{H}}\mathbf{z} + \mathbf{n}. \tag{14}$$

Note that $\mathbf{Hs}$ and $\tilde{\mathbf{H}}\mathbf{z}$ describe the same point in a lattice, but the LLL-reduced matrix $\tilde{\mathbf{H}}$ is usually much better conditioned than the original channel matrix $\mathbf{H}$. For $\mathbf{s} \in \mathbb{Z}^m$ we also have $\mathbf{z} \in \mathbb{Z}^m$, so $\mathbf{s}$ and $\mathbf{z}$ stem from the same set. However, for M-QAM, i.e. $\mathbf{s} \in \mathcal{A}^m$, the lattice is finite and the domain of $\mathbf{z}$ differs from $\mathcal{A}^m$. This is illustrated in **Fig. 2** for 16-QAM, one transmit antenna ($m = 2$) and a transformation matrix $\mathbf{T} = [1, -1; 0, 1]$.

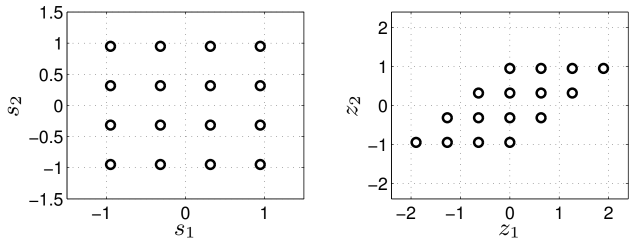

Fig. 2. Original 16-QAM symbols $\mathbf{s} \in \mathcal{A}^2$ with $\mathcal{A} = \{\pm\frac{1}{2}a, \pm\frac{3}{2}a\}$ (left) and transformed symbols $\mathbf{z} \in \mathbf{T}^{-1}\mathcal{A}^2$ (right).

The idea behind LR-aided linear detection is to consider the equivalent system model in (14) and perform the nonlinear quantization on $\mathbf{z}$ instead of $\mathbf{s}$. For LR-aided ZF this means that first

$$\tilde{\mathbf{z}}_{\text{LR-ZF}} = \mathbf{T}^{-1}\tilde{\mathbf{s}}_{\text{ZF}} = \tilde{\mathbf{H}}^+\mathbf{x} = \mathbf{z} + \tilde{\mathbf{H}}^+\mathbf{n} \tag{15}$$

is calculated, where the multiplication with $\tilde{\mathbf{H}}^+$ usually causes less noise amplification than the multiplication with $\mathbf{H}^+$ in (10) due to the roughly orthogonal columns of $\tilde{\mathbf{H}}$. Therefore, a hard decision based on $\tilde{\mathbf{z}}_{\text{LR-ZF}}$ is in general more reliable than one on $\tilde{\mathbf{s}}_{\text{ZF}}$. However, the elements of the transformed vector $\mathbf{z}$ are not independent of each other, e.g., in **Fig. 2** the range of possible values for z_1 depends on z_2. A straightforward (though suboptimal) solution is to perform an unconstrained elementwise quantization[3] $\hat{\mathbf{z}}_{\text{LR-ZF}} = \mathcal{Q}\{\tilde{\mathbf{z}}_{\text{LR-ZF}}\}$, calculate $\hat{\mathbf{s}}_{\text{LR-ZF}} = \mathbf{T}\hat{\mathbf{z}}_{\text{LR-ZF}}$, and finally restrict this result to the set $\mathcal{A}^m$. Note that due to the quantization in the transformed domain this receiver structure is not linear anymore.

Similar to Section IV-A we may apply a MMSE filter instead of the ZF solution in order to get an improved estimate for $\mathbf{z}$. One obvious way is given by the MMSE-solution of the lattice-reduced system (14)

$$\tilde{\mathbf{z}}^{(\mathbf{H})}_{\text{LR-MMSE}} = \left(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} + \sigma_n^2\mathbf{T}\mathbf{T}^{-1}\right)^{-1}\tilde{\mathbf{H}}^T\mathbf{x} \qquad (16)$$

$$= \mathbf{T}^{-1}\tilde{\mathbf{s}}_{\text{MMSE}} \; . \qquad (17)$$

Again, this corresponds to simply replace $\tilde{\mathbf{s}}_{\text{ZF}}$ in (15) by $\tilde{\mathbf{s}}_{\text{MMSE}}$ from (12). A better alternative is to perform the LR for the extended channel matrix (11), i.e. $\underline{\tilde{\mathbf{H}}} = \underline{\mathbf{H}}\,\underline{\mathbf{T}}$, and compute

$$\tilde{\mathbf{z}}^{(\underline{\mathbf{H}})}_{\text{LR-MMSE}} = \underline{\tilde{\mathbf{H}}}^{+}\underline{\mathbf{x}} \; , \qquad (18)$$

because in this case the LR is optimized with respect to the MMSE criterion. As the condition of $\underline{\mathbf{H}}$ determines the noise amplification of a common MMSE detector and not the condition of $\mathbf{H}$, this second solution will outperform the obvious one. We will see in Section V that this solution does not only yield a performance gain, but also reduces the computational complexity.

C. Lattice Reduction aided SIC

As $\tilde{\mathbf{H}}$ is only roughly orthogonal, the mutual influence of the transformed signals z_i is small, but still present. Thus, successive interference cancellation techniques like V-BLAST may result in additional improvements. As shown in several publications, e.g. [6], [7], SIC can be well described in terms of the QR decomposition of the channel matrix. Applying this strategy to the system model from (14) we get

$$\tilde{\mathbf{z}}_{\text{LR-ZF-SIC}} = \tilde{\mathbf{Q}}^T\mathbf{x} = \tilde{\mathbf{R}}\mathbf{z} + \tilde{\mathbf{Q}}^T\mathbf{n} \; , \qquad (19)$$

where $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$ have already been calculated by the LLL algorithm. Due to the upper triangular structure of $\tilde{\mathbf{R}}$, the m-th element of $\tilde{\mathbf{z}}$ is free of interference and can be used to estimate z_m. Proceeding with $\tilde{z}_{m-1}, \ldots, \tilde{z}_1$ and assuming correct previous decisions, the interference can be perfectly cancelled in each step.

It is well known, that because of error propagation the order of detection has a large influence on the performance of SIC. The optimum order can be calculated efficiently by the so-called Post-Sorting-Algorithm (PSA) proposed in [5], [6], which exploits the fact, that the mean error in each detection step is proportional to the diagonal elements of $\tilde{\mathbf{R}}^{-1}$.

Similar to linear detection, we can consider the lattice-reduced version of the extended system model with the equivalent channel matrix $\underline{\tilde{\mathbf{H}}} = \underline{\tilde{\mathbf{Q}}}\,\underline{\tilde{\mathbf{R}}}$. This leads to LR-aided MMSE-SIC with decision variables given by

$$\tilde{\mathbf{z}}_{\text{LR-MMSE-SIC}} = \underline{\tilde{\mathbf{Q}}}^T\underline{\mathbf{x}} = \underline{\tilde{\mathbf{R}}}\mathbf{z} + \boldsymbol{\eta} \; . \qquad (20)$$

[3]Note that again proper shifting and scaling is necessary in order to allow for simple rounding in the quantization step [4].

where the newly defined noise term $\boldsymbol{\eta}$ also incorporates residual interference. The detection procedure equals that of LR-aided ZF-SIC.

V. Performance Analysis

In the sequel, we investigate a MIMO system with $n_T = 4$ transmit and $n_R = 4$ receive antennas and 4-QAM modulation. E_b denotes the average energy per information bit arriving at the receiver, thus $E_b/N_0 = n_R/(\log_2(M)\,\sigma_n^2)$ holds. **Tab. 2** shows the average number of required outer loops (lines 3-20 in Tab. 1) and column exchanges (lines 12-16 in Tab. 1) in the LLL algorithm for[4] $E_b/N_0 = 10\,\text{dB}$ in order to investigate the influence of using SQRD instead of a common unsorted QR decomposition as initialization.

Tab. 2 Average number of loops and column exchanges in the LLL algorithm

LR		Outer Loops (3-20)	Exchanges (12-16)
ZF	QR	32.3	13.2
	SQRD	17.1	5.2
MMSE	QR	22.4	8.1
	SQRD	8.5	0.8

Obviously, SQRD, which requires only negligible overhead, significantly reduces the number of loops and thereby the overall computational complexity of LLL. This is especially true in the MMSE case, where the extended channel matrix $\underline{\mathbf{H}}$ is considered. As an example, the number of column exchanges is reduced by a factor of 2.5 for ZF and 10 for MMSE detection, respectively.

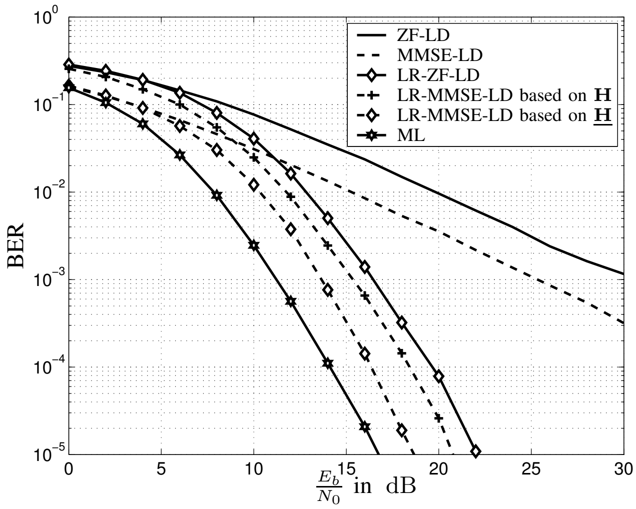

Fig. 3. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) linear detection.

Fig. 3 shows the bit error rates (BER) of the standard linear ZF and MMSE detectors. Due to the noise enhancement, the

[4]Of course, the complexity of LLL does only depend on E_b/N_0 for MMSE, not for ZF.

performance is poor in comparison to ML and both schemes achieve only a diversity degree of $d = n_R - n_T + 1 = 1$. In contrast, linear equalization of the lattice-reduced system reaches the full diversity degree of $d = 4$ and leads to a significant performance improvement. As indicated in Section IV-B, for MMSE detection it is much better to apply LR to the extended channel matrix $\underline{\mathbf{H}}$ instead of using $\tilde{\mathbf{H}}$ for filtering. Therefore, we will disregard the version based on $\mathbf{H}$ in the following. The proposed LR-MMSE (based on $\underline{\mathbf{H}}$) achieves an improvement of approximately $3.3\,\mathrm{dB}$ in comparison to LR-ZF for a BER of 10^{-5}.

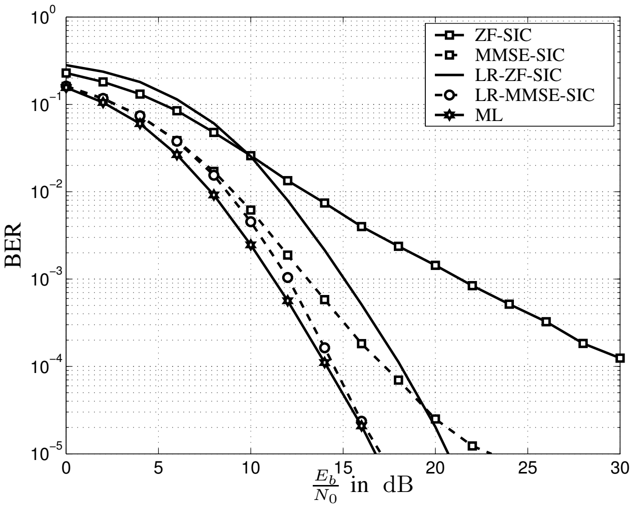

Fig. 4. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

The performances of the successive detection schemes with optimum ordering are illustrated in **Fig. 4**. As expected, they clearly outperform the linear detection methods from Fig. 3. Note that this improvement comes at almost no cost, because the complexity of SIC (after having calculated the QR decomposition) is comparable to that of linear detection. Again, detection with respect to the lattice-reduced system significantly reduces the bit error rate. The proposed LR-MMSE-SIC scheme achieves almost ML performance, while the main computational effort is required only once per transmitted frame.

The BERs of SIC-based detection for a system with $n_T = 6$ and $n_R = 6$ are shown in **Fig. 5**. We observe the same performance improvement for the LR-aided schemes and the benefit of the MMSE extension. The gap between LR-MMSE-SIC and ML is only $1\,\mathrm{dB}$ for a BER of 10^{-5}.

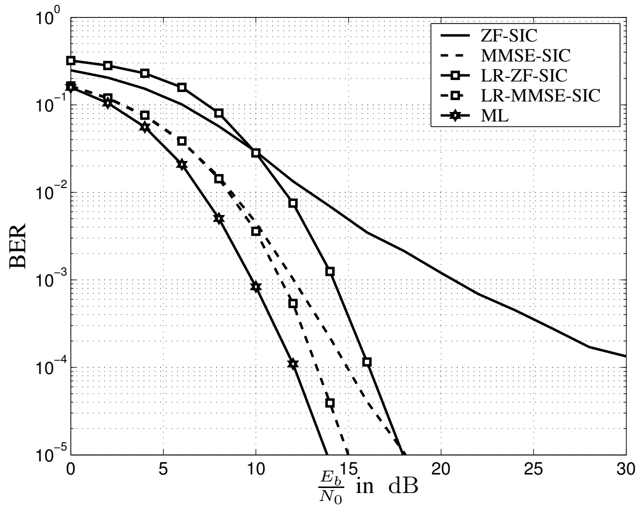

Fig. 5. Bit Error Rate of a system with $n_T = 6$ and $n_R = 6$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

VI. Summary and Conclusions

In this paper, we investigated several detection schemes for multiple antenna systems making use of the lattice-reduction algorithm proposed by Lenstra, Lenstra and Lovász. We showed that the straightforward way to perform MMSE detection after lattice-reduction does not yield satisfying results. Instead, we proposed a new method, where LR is applied to an extended system model. In conjunction with successive interference cancellation, this strategy nearly leads to maximum-likelihood performance. Furthermore, we analyzed the impact of a sorted QR decomposition on the LLL algorithm and demonstrated that SQRD can dramatically decrease the computational effort. Thus, we arrived at a near-optimum detector with very low complexity.

References

[1] 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 *IEEE Proc. ISSSE*, Pisa, Italy, September 1998.

[2] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest Point Search in Lattices," *IEEE Trans. on Information Theory*, vol. 48, no. 8, pp. 2201–2214, August 2002.

[3] H. Yao and G. Wornell, "Lattice-Reduction-Aided Detectors for MIMO Communication Systems," in *IEEE Proc. Globecom*, Taipei, Taiwan, November 17-21 2002.

[4] C. Windpassinger and R. F. H. Fischer, "Low-Complexity Near-Maximum-Likelihood Detection and Precoding for MIMO Systems using Lattice Reduction," in *IEEE Proc. ITW*, Paris, France, March 2003.

[5] B. Hassibi, "An Efficient Square-Root Algorithm for BLAST," in *IEEE Proc. ICASSP*, Istanbul, Turkey, June 2000, pp. 5–9.

[6] D. Wübben, R. Böhnke, V. Kühn, and K. D. Kammeyer, "MMSE Extension of V-BLAST based on Sorted QR Decomposition," in *IEEE Proc. VTC-Fall*, Orlando, Florida, USA, Ocober 2003.

[7] R. Böhnke, D. Wübben, V. Kühn, and K. D. Kammeyer, "Reduced Complexity MMSE Detection for BLAST Architectures," in *IEEE Proc. Globecom*, San Francisco, California, USA, December 2003.

[8] R. F. H. Fischer and C. Windpassinger, "Real- vs. Complex-Valued Equalization in V-BLAST Systems," *IEE Electronics Letters*, vol. 39, no. 5, pp. 470–471, March 2003.

[9] C. P. Schnorr and M. Euchner, "Lattice Basis Reduction: Improved Practical Alorithms and Solving Subset Sum Problems," *Mathematical Programming*, vol. 66, pp. 181–191, 1994.

[10] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring Polynomials with Rational Coefficients," *Math. Ann*, vol. 261, pp. 515–534, 1982.

[11] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K. D. Kammeyer, "Efficient Algorithm for Decoding Layered Space-Time Codes," *IEE Electronics Letters*, vol. 37, no. 22, pp. 1348–1350, October 2001.