MMSE-Based Lattice-Reduction for Near-ML Detection of MIMO Systems

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, Phone: +49 421 218 2545

***Abstract*— Recently 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 suboptimum and propose an alternative method based on an extended system model. In conjunction with simple successive interference cancellation this scheme almost reaches the performance of maximum-likelihood detection. Furthermore, we demonstrate that the application of Sorted QR Decomposition (SQRD) as a initialization step 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. Within the V-BLAST architecture parallel data streams are simultaneously transmitted over n_T different antennas in order to increase the spectral efficiency. 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]–[4]. 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 spent 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 [5]–[8]. 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 [9] and further investigated in [10], [11].

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. The computational complexity as well as the performance is analyzed in Section V. Concluding remarks can be found in Section VI and some additional comments on required shifting and scaling operations are given in the Appendix.

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

II. System description

We consider the multiple antenna system shown in **Fig. 1** with n_T transmit and $n_R \geq n_T$ receive antennas in a flat fading environment. According to the V-BLAST architecture, the data is demultiplexed into n_T data substreams of equal length (called layers). These substreams are mapped onto M-QAM symbols and transmitted over the n_T antennas simultaneously.

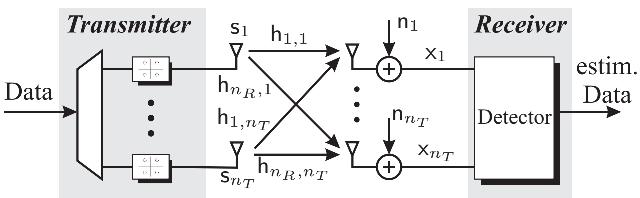

Fig. 1. Model of a MIMO system with n_T transmit and n_R receive antennas.

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

[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.

power of each antenna is normalized to one, i.e. $\mathrm{E}\left\{\mathsf{ss}^H\right\} = \mathbf{I}_{n_T}$ and $\mathrm{E}\left\{\mathsf{nn}^H\right\} = \sigma_n^2 \mathbf{I}_{n_R}$. The $n_R \times n_T$ channel matrix H contains uncorrelated complex gaussian fading gains $\mathsf{h}_{i,j}$ with unit variance. We assume a flat fading environment, where the channel matrix H is constant over a frame and changes independently from frame to frame (*block fading channel*). Within the derivation and discussion of the detector structures we assume perfect knowledge about the channel state information (CSI) by the receiver. However, within the performance evaluation in Section V we also consider imperfect channel knowledge due to channel estimation [12].

Treating real and imaginary parts 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 [13].

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 $\hat{\mathbf{s}}_{\mathrm{ML}}$ with minimum euclidian distance to the receive vector $\mathbf{x}$, i.e.

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

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]–[4], 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 (LR) in order to improve the performance of SIC and linear detection (LD). One advantage of this strategy is, that the computational overhead required by the lattice-reduction algorithm 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 $\mathbf{h}_k$ ($1 \le k \le m$) 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

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

and is called lattice $\mathcal{L}(\mathbf{H})$. However, the same lattice is also spanned by any matrix $\tilde{\mathbf{H}}$ emenated from $\mathbf{H}$ by the elementary column operations

- interchanging columns
- multiplying any column by -1
- Adding a multiple of one column to another.

It can be shown, that the product of the corresponding elementary matrices results in a *unimodular* transformation matrix $\mathbf{T}$, i.e. $\mathbf{T}$ contains only integer entries and the determinant is $\det(\mathbf{T}) = \pm 1$ [3]. Therefore, each matrix $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{T}$ generates the *same* lattice $\mathcal{L}(\tilde{\mathbf{H}}) = \mathcal{L}(\mathbf{H})$, if and only if the $m \times m$ matrix $\mathbf{T}$ is unimodular:

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

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. As both matrices describe the same receive signal space, we may choose that basis with nicer properties for signal detection in Section IV.

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 $\mathbf{H}$ and therefore leads to less noise enhancement for linear detection. In order to describe the impact of $\mathbf{H}$ on the noise enhancement, we introduce the condition number $\kappa(\mathbf{H}) = \sigma_{\max}/\sigma_{\min} \ge 1$, with $\sigma_{\max}$ and $\sigma_{\min}$ denoting the largest and the smallest singular value of $\mathbf{H}$, respectively [14]. With $\|\mathbf{H}\|_2$ defining the spectral norm of $\mathbf{H}$, the condition number corresponds to $\kappa(\mathbf{H}) = \|\mathbf{H}\|_2\|\mathbf{H}^{-1}\|_2 = \|\mathbf{R}\|_2\|\mathbf{R}^{-1}\|_2$. For ill-conditioned matrices $\kappa(\mathbf{H})$ is large, whereas for orthogonal matrices $\kappa(\mathbf{H}) = 1$, as no noise enhancement is caused. As the reduced basis of $\tilde{\mathbf{H}}$ is chosen to have roughly orthogonal basis vectors, the corresponding condition number $\kappa(\tilde{\mathbf{H}})$ is usually much smaller than $\kappa(\mathbf{H})$. Consequently, a linear detector with respect to this new basis may achieve better performance results, as the impact of noise enhancement is reduced.

With respect to the QR decomposition, the basis vector $\tilde{\mathbf{h}}_k$ is almost orthogonal to the space spanned by $\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_{k-1}$, if the elements $|\tilde{r}_{1,k}|, \ldots, |\tilde{r}_{k-1,k}|$ are close to zero. An efficient (though not optimal) way to determine a reduced basis was proposed by Lenstra, Lenstra and Lovàsz [15].

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 $(1/4 < \delta \leq 1)$, if [15]

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

and

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

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 [15]. 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 $\mathbf{T}$.

Tab. 1 LLL LATTICE-REDUCTION ALGORITHM [15]

INPUT: $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{P}$ (default: $\mathbf{P} = \mathbf{I}_m$)
OUTPUT: $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{R}}$, $\mathbf{T}$
(1) Initialization: $\tilde{\mathbf{Q}} := \mathbf{Q}$, $\tilde{\mathbf{R}} := \mathbf{R}$, $\mathbf{T} := \mathbf{P}$
(2) $k = 2$
(3) while $k \leq m$
(4) for $\ell = k-1, \ldots, 1$
(5) $\mu = \lceil \tilde{\mathbf{R}}(\ell, k)/\tilde{\mathbf{R}}(\ell, \ell) \rfloor$
(6) if $\mu \neq 0$
(7) $\tilde{\mathbf{R}}(1:\ell, k) := \tilde{\mathbf{R}}(1:\ell, k) - \mu\, \tilde{\mathbf{R}}(1:\ell, \ell)$
(8) $\mathbf{T}(:, k) := \mathbf{T}(:, k) - \mu\, \mathbf{T}(:, \ell)$
(9) end
(10) end
(11) if $\delta\, \tilde{\mathbf{R}}(k-1, k-1)^2 > \tilde{\mathbf{R}}(k, k)^2 + \tilde{\mathbf{R}}(k-1, k)^2$
(12) Swap columns $k-1$ and k in $\tilde{\mathbf{R}}$ and $\mathbf{T}$
(13) Calculate Givens rotation matrix $\mathbf{\Theta}$ such that element $\tilde{\mathbf{R}}(k, k-1)$ becomes zero:
$$\mathbf{\Theta} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \text{ with } \begin{array}{l} \alpha = \frac{\tilde{\mathbf{R}}(k-1,k-1)}{\|\tilde{\mathbf{R}}(k-1:k,k-1)\|} \\ \beta = \frac{\tilde{\mathbf{R}}(k,k-1)}{\|\tilde{\mathbf{R}}(k-1:k,k-1)\|} \end{array}$$
(14) $\tilde{\mathbf{R}}(k-1:k, k-1:m) := \mathbf{\Theta}\tilde{\mathbf{R}}(k-1:k, k-1:m)$
(15) $\tilde{\mathbf{Q}}(:, k-1:k) := \tilde{\mathbf{Q}}(:, k-1:k)\mathbf{\Theta}^T$
(16) $k := \max\{k-1, 2\}$
(17) else
(18) $k := k+1$
(19) end
(20) end

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 (see line 16 of Tab. 1). In the first step ($k = 2$), no exchange operation is necessary if $\delta\, \tilde{r}^2_{1,1} \leq \tilde{r}^2_{2,2} + \tilde{r}^2_{1,2}$ 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 [16] and extended in [10], [11] 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 [10]. In Section V we will see, 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

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

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 (see **Fig. 2**). 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. In fact, using a result from random matrix theory [17], it can be shown that in the large system limit for $n_T = n_R \to \infty$ the noise amplification tends to infinity almost surely.

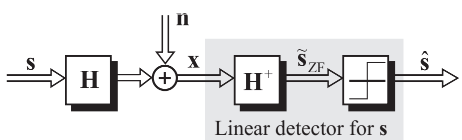

Fig. 2. Block diagram of linear zero-forcing detection.

The minimum mean square error (MMSE) detector takes the noise term into account and thereby leads to an improved performance. As shown in [9]–[11], 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

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

[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.

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

$$\begin{aligned} \tilde{\mathbf{s}}_{\text{MMSE}} &= \left(\mathbf{H}^T\mathbf{H} + \sigma_n^2 \mathbf{I}_m\right)^{-1} \mathbf{H}^T \mathbf{x} \qquad (12) \\ &= \left(\underline{\mathbf{H}}^T \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{H}}^T \underline{\mathbf{x}} = \underline{\mathbf{H}}^+ \underline{\mathbf{x}} \,, \qquad (13) \end{aligned}$$

which exactly matches the structure of (10), i.e. $\underline{\mathbf{x}}$ is filtered by the pseudo-inverse of $\underline{\mathbf{H}}$. Thus, the MMSE detector agrees to a ZF detector with respect to the extended system model and consequently the condition number $\kappa(\underline{\mathbf{H}})$ determines the effective noise amplification. This observation will be extremely important for incorporating the MMSE criterion in the lattice-based detection algorithms.

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

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

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. 3** for 16-QAM, one transmit antenna ($m = 2$) and a transformation matrix $\mathbf{T} = [1, -1; 0, 1]$.

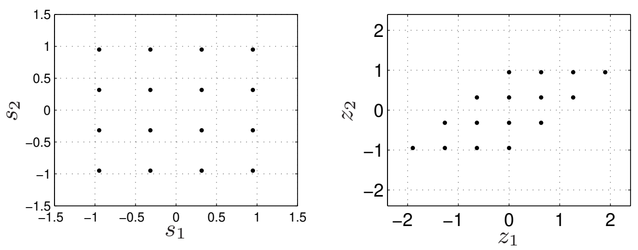

Fig. 3. Original 16-QAM symbols $\mathbf{s} \in \mathcal{A}^2$ with $\mathcal{A} = \{\pm\frac{1}{2}a, \pm\frac{3}{2}a\}$ (left) and transformed symbols $\mathbf{z} \in \mathbf{T}^{-1}\mathcal{A}^2$ (right).

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

$$\begin{aligned} \tilde{\mathbf{z}}_{\text{LR}-\text{ZF}} &= \tilde{\mathbf{H}}^+\mathbf{x} = \mathbf{z} + \tilde{\mathbf{H}}^+\mathbf{n} \qquad (15) \\ &= \mathbf{T}^{-1}\tilde{\mathbf{s}}_{\text{ZF}} \qquad (16) \end{aligned}$$

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 $\tilde{\mathbf{z}}_{\text{LR}-\text{ZF}}$ 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. 3** the range of possible values for z_1 depends on z_2. A straightforward (though suboptimal) solution is to perform an unconstrained element-wise quantization[3] $\hat{\mathbf{z}}_{\text{LR}-\text{ZF}} = \mathcal{Q}\{\tilde{\mathbf{z}}_{\text{LR}-\text{ZF}}\}$, calculate $\hat{\mathbf{s}}_{\text{LR}-\text{ZF}} = \mathbf{T}\hat{\mathbf{z}}_{\text{LR}-\text{ZF}}$, 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. The principle block diagram for a LR-aided detector is shown in **Fig. 3**

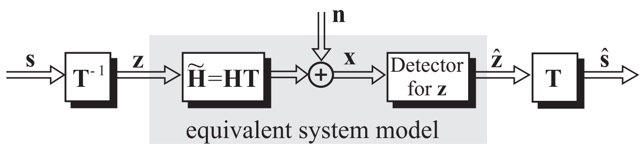

Fig. 4. Block diagram of a Lattice-Reduction based detector.

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 to do so is given by the MMSE-solution of the lattice-reduced system (14)

$$\begin{aligned} \tilde{\mathbf{z}}^{(\mathbf{H})}_{\text{LR}-\text{MMSE}} &= \left(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} + \sigma_n^2 \mathbf{T}^T\mathbf{T}\right)^{-1} \tilde{\mathbf{H}}^T\mathbf{x} \qquad (17) \\ &= \mathbf{T}^{-1}\tilde{\mathbf{s}}_{\text{MMSE}} \,. \end{aligned}$$

Again, this corresponds to simply replace $\tilde{\mathbf{s}}_{\text{ZF}}$ in (16) 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

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

because in this case the LR is executed with respect to the MMSE filter, i.e. 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 solution. 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. [10], [11], 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

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

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.

[3]Note that proper shifting and scaling is necessary in order to allow for simple rounding in the quantization step. The concerning details are discussed within the Appendix.

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 [9], [10], 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

$$\tilde{\mathbf{z}}_{\mathrm{LR-MMSE-SIC}} = \underline{\tilde{\mathbf{Q}}}^T \underline{\mathbf{x}} = \underline{\tilde{\mathbf{R}}}\mathbf{z} + \boldsymbol{\eta}\ . \tag{20}$$

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

D. Discussion of Decision Regions

In order to visualize the decision boundaries of the different detection schemes, we analyze a simple scheme with $m = n = 2$ and s_1 and s_2 being integers of large range.

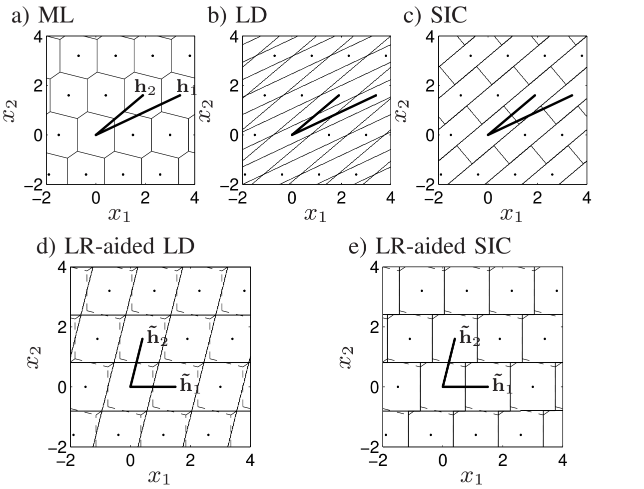

Fig. 5. Exemplary receive signal space and decision areas of a) Maximum-Likelihood detection, b) Linear Detection (LD), c) Successive Interference Cancellation (SIC), d) LR-aided LD and e) LR-aided SIC.

The undisturbed received signals $\mathbf{Hs}$ in **Fig. 5** a) can be viewed as a lattice spanned by the basis vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ where each lattice point corresponds to a linear combination of the basis vectors. Due to the ML condition (5) all points in the receive plane which are closer to a particular lattice point than to any other are associated with this particular lattice point forming the corresponding Voronoi region [4]. The boundary regions for LD in b) are parallelograms, where the sides are parallel to the basis vectors. An ill-conditioned basis results in a very narrow and stretched parallelogram, with highly probable decision errors. In contrast, SIC decides s_1 and s_2 one after another resulting in rectangular decision regions with sides parallel to $\mathbf{q}_1$ and $\mathbf{q}_2$ and side length equal to $r_{1,1}$ and $r_{2,2}$. In Fig. 5 d) the decision regions for LR-aided LD are shown. The sides of the corresponding parallelogram are now parallel to the almost orthogonal basis vectors $\tilde{\mathbf{h}}_1$ and $\tilde{\mathbf{h}}_2$. Consequently, this parallelogram is less stretched and the boundaries nearly correspond to the Voronoi regions. This difference becomes even smaller in case of LR-aided SIC, shown in e). Due to the almost orthogonal basis $\tilde{\mathbf{H}}$, the element $\tilde{r}_{2,2}$ (length of $\tilde{\mathbf{h}}_2$ perpendicular to $\tilde{\mathbf{h}}_1$) is larger than $r_{2,2}$ resulting in an almost quadratic decision region.

V. Performance Analysis

In the sequel, we investigate a MIMO system with n_T transmit and n_R receive antennas and M-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. At first, we discuss the influence of SQRD on the computational complexity of the LLL algorithm. Next, BER of simulation results for systems with perfect channel state information (CSI) at the receiver side and with imperfect CSI (pilot-aided channel estimation) are presented.

A. Hint on Computational Effort

As mentioned in Section III the overall complexity of LLL algorithm depends highly on the number of column exchanges, as this does not only require matrix multiplications, but also decreases the counter k. However, SQRD already exchanges the columns of the channel matrix within the (necessary) QR decomposition and achieves thereby already a pre-sorting for the LLL algorithm. In order to investigate the impact of this pre-sorting, **Fig. 6** shows the average number of required column exchanges $\bar{c}$ (lines 12-16 in Tab. 1) within the LLL algorithm in dependance of E_b/N_0 for a system with $n_T = n_R = 4$ antennas (or equivalently $m = n = 8$).

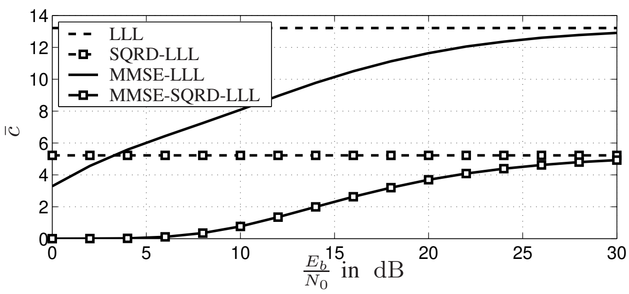

Fig. 6. Average number of column exchanges $\bar{c}$ within the LLL reduction of a system with $n_T = n_R = 4$ antennas ($m = n = 8$).

A common lattice-reduction (LLL) of $\mathbf{H}$ requires an average number of $\bar{c}_{\text{LLL}} = 13.2$ column exchanges. By applying SQRD, which requires only negligible overhead compared to standard QR decomposition, this number is significantly reduced by a factor of 2.5 to $\bar{c}_{\text{SQRD-LLL}} = 5.2$. The complexity decrease is even larger in the MMSE case, where the extended channel matrix $\underline{\mathbf{H}}$ is considered. In the low SNR region almost no additional column exchanges according to (9) are required by the SQRD-LLL combination, so only size-reduction to fulfill (8) is performed. For $E_b/N_0 = 10\,\text{dB}$ we obtain $\bar{c}_{\text{MMSE-LLL}} = 8.1$ and $\bar{c}_{\text{MMSE-SQRD-LLL}} = 0.8$ indicating a remarkable reduction of column exchanges by a factor of 10. Compared to the common LLL, this factor is even more impressive, i.e. $\bar{c}_{\text{LLL}}/\bar{c}_{\text{MMSE-SQRD-LLL}} = 17.3$. As the MMSE solution converges to the ZF solution in the large SNR region, the operated number of exchanges also approach the corresponding numbers for ZF. Obviously, using the QR decomposition achieved by SQRD reduces significantly the computational complexity of the LLL algorithm. Furthermore, performing a MMSE-reduction will lead not only to improved detection performance but requires also less column exchanges.

B. BER performance with perfect CSI

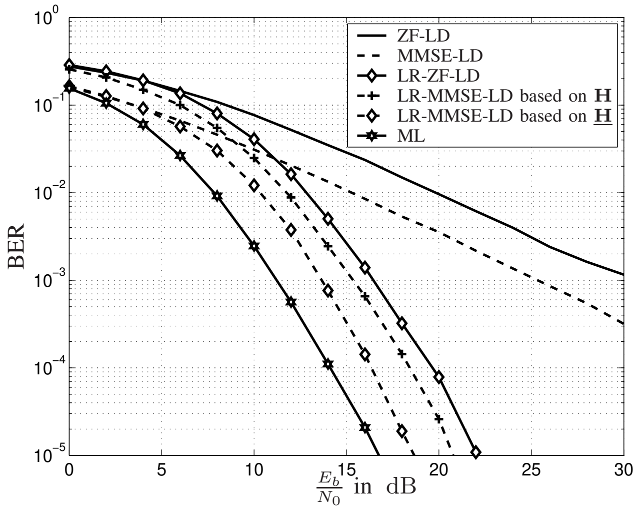

Fig. 7. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) linear detection (LD).

Fig. 7 shows the bit error rates (BER) of the standard linear ZF and MMSE detectors. Due to the noise enhancement, the 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 $\tilde{\mathbf{H}}$ in the following. The proposed LR-MMSE (based on $\underline{\mathbf{H}}$) achieves an improvement of approximately $3.3\,\text{dB}$ in comparison to LR-ZF for a BER of 10^{-5}.

The performances of the successive detection schemes with optimum ordering are illustrated in **Fig. 8**. As expected, they clearly outperform the linear detection methods from Fig. 7. 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

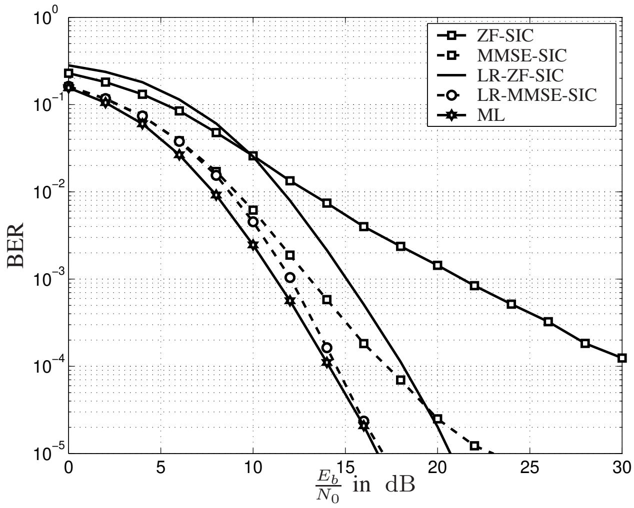

Fig. 8. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

the main computational effort is required only once per transmitted frame.

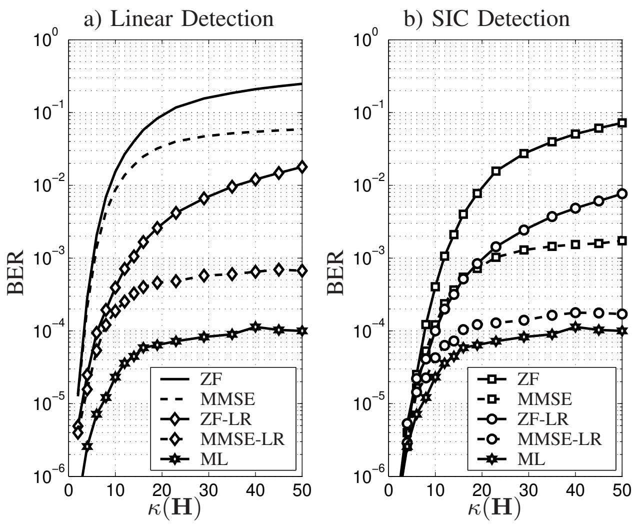

Fig. 9. Bit Error Rates versus condition number $\mathbf{H}$ for a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, $E_b/N_0 = 16\,\text{dB}$, ZF (continuous lines) and MMSE (dashed lines) a) linear detection and b) SIC detection.

Fig. 9 shows the achieved performance of several detection schemes versus the condition number $\kappa(\mathbf{H})$ of the channel matrix and $E_b/N_0 = 16\,\text{dB}$. Part a) contains the simulation results for the common and LR-aided linear detection schemes and ML detection. For $\kappa(\mathbf{H}) \approx 1$ all schemes achieve very good performance with BER $\leq 10^{-5}$, as the common ZF detector is equivalent to ML for $\kappa(\mathbf{H}) = 1$. However, if the matrix is ill-condition, the performances of the common ZF and MMSE linear detector are poor, whereas the LR-aided linear detectors achieve suitable results. The corresponding BERs for common SIC and LR-aided SIC are given in part part b) of Fig. 9. The successive schemes generally outperform the linear schemes. Furthermore, MMSE-LR-SIC almost achieves the BER of ML for the whole range of investigated condition numbers.

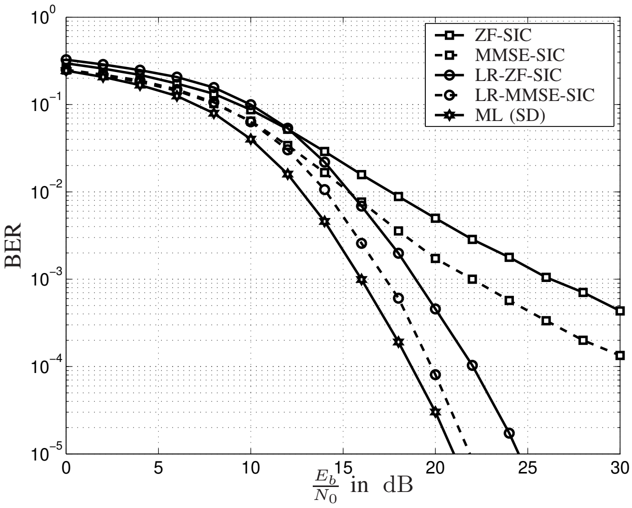

Fig. 10. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 16-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

Fig. 10 shows the performance of the SIC algorithms for the same antenna configuration ($n_T = n_R = 4$) and 16-QAM modulation. The performance of the standard SIC algorithms is far away of ML (realized by sphere detection). In contrast, the LR-aided SIC schemes achieve very good results. It is worth to note, that the effort for LLL does not depend on the modulation index M, so the computational complexity for lattice-reduction corresponds to the effort for the system with 4-QAM modulation.

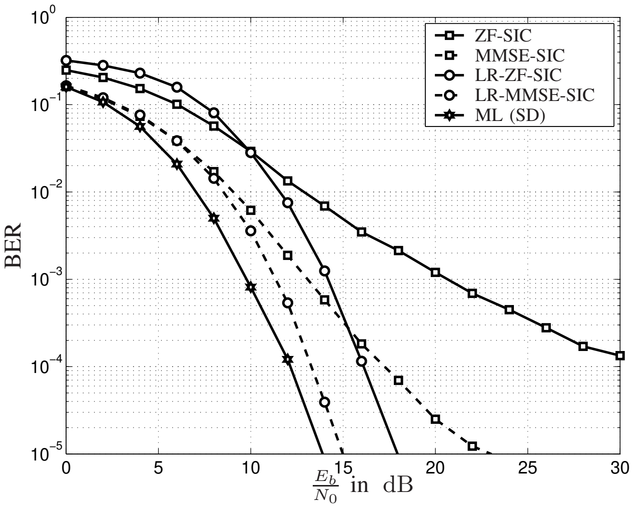

Fig. 11. Bit Error Rate of a system with $n_T = 6$ and $n_R = 6$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

The BERs of SIC-based detection for a system with $n_T = n_R = 6$ and 4-QAM modulation are shown in **Fig. 11**. 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}.

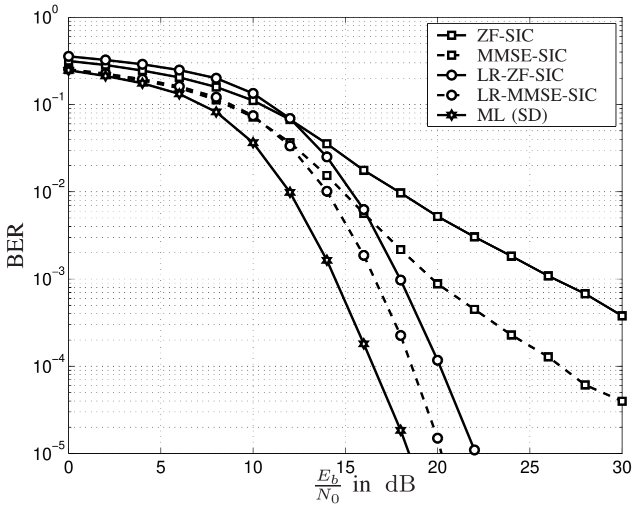

Fig. 12. Bit Error Rate of a system with $n_T = 6$ and $n_R = 6$ antennas, 16-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection.

In **Fig. 12** the BER performance is shown for the same antenna configuration, but 16-QAM modulation. Again, the LR-based schemes clearly outperform the standard SIC algorithms by several dB. The proposed LR-MMSE-SIC outperforms the LR-ZF-SIC by $2.2\,\mathrm{dB}$ and the loss to ML is only $1.7\,\mathrm{dB}$ for a BER of 10^{-5}.

C. BER performance with imperfect CSI

In the sequel, we investigate the performance of common and LR-aided detection schemes when only imperfect CSI information due to pilot-aided channel estimation is available at the receiver. For estimating the channel a complex $n_T \times L$ training matrix $\mathbf{S}_{\text{Pilot}} = [\mathbf{s}_1, \ldots, \mathbf{s}_L]$ with orthogonal rows is added to each transmit frame [12]. With $\mathbf{X}_{\text{Pilot}}$ denoting the corresponding $n_R \times L$ receive matrix, the ML channel estimation is given by $\hat{\mathbf{H}} = \mathbf{X}_{\text{Pilot}}\mathbf{S}^{+}_{\text{Pilot}}$.

Fig. 13 shows the corresponding BERs with imperfect CSI. Similar to the results with perfect CSI in Fig. 8, the LR-aided SIC schemes perform very well. Again, the LR-MMSE-SIC almost achieves ML performance, with a small gap of approximately $0.6\,\mathrm{dB}$.

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. The robustness of the

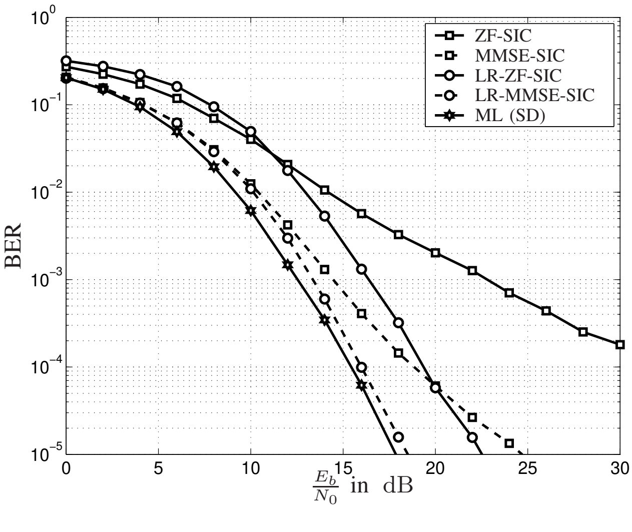

Fig. 13. Bit Error Rate of a system with $n_T = 4$ and $n_R = 4$ antennas, 4-QAM symbols, ZF (continuous lines) and MMSE (dashed lines) optimally sorted SIC detection with imperfect CSI due to channel estimation with $L = 12$ pilot symbols.

achieved detection algorithm with respect to channel estimation was proofed by simulation result. 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 low complexity.

Appendix

As shown in Section III, two basis $\mathbf{H}$ and $\tilde{\mathbf{H}}$ describe the same lattice $\mathcal{L}$ if relation (7) holds and if the input signals stem from the infinite integer space $\mathbb{Z}^m$. For a M-QAM constellation, the last condition is not fulfilled. However, we can interpret $\mathcal{A}^m$ as the shifted and scaled version of the integer subset $\mathcal{D}^m \subset \mathbb{Z}^m$, i.e. $\mathcal{A}^m = a\,(\mathcal{D}^m + \frac{1}{2}\mathbf{1}_m)$ [7]. For M-QAM this integer integer subset is given by $\mathcal{D} = \{-\frac{\sqrt{M}}{2}, \ldots, \frac{\sqrt{M}}{2} - 1\}$. As an example, the transmit signals for 16-QAM are from the set $\mathcal{A} = \{\pm 0.316, \pm 0.949\}$, whereas the integer subset is given by $\mathcal{D} = \{-2, -1, 0, 1\}$. With this definition, the transmit signal vector $\mathbf{s} \in \mathcal{A}^m$ can be rewritten as $\mathbf{s} = a\,(\bar{\mathbf{s}} + \frac{1}{2}\mathbf{1}_m)$ with $\bar{\mathbf{s}}$ stemming from the integer subset, i.e. $\bar{\mathbf{s}} \in \mathcal{D}^m$. Consequently, the transformed signal vector $\mathbf{z}$ can be represented by

$$\mathbf{z} = \mathbf{T}^{-1}\mathbf{s} = a\,\mathbf{T}^{-1}\left(\bar{\mathbf{s}} + \frac{1}{2}\mathbf{1}_m\right) = a\,\left(\bar{\mathbf{z}} + \frac{1}{2}\mathbf{T}^{-1}\mathbf{1}_m\right) \tag{21}$$

introducing the definition of the integer vector $\bar{\mathbf{z}} = \mathbf{T}^{-1}\bar{\mathbf{s}} \in \mathbf{T}^{-1}\mathcal{D}^m \subset \mathbb{Z}^m$.

The LR-aided detection schemes discussed in Section IV comprise the estimation of $\mathbf{z}$ with respect to $\mathbf{x} = \mathbf{H}\mathbf{z} + \mathbf{n}$ and mapping theses estimates onto the symbols $\mathbf{s} \in \mathcal{A}^m$. When performing LR-aided linear ZF detection, the filter output signal is given by $\tilde{\mathbf{z}} = \mathbf{z} + \tilde{\mathbf{H}}^{+}\mathbf{n} = a\,(\bar{\mathbf{z}} + \frac{1}{2}\mathbf{T}^{-1}\mathbf{1}_m) + \tilde{\mathbf{H}}^{+}\mathbf{n}$. In order to achieve an estimation for $\mathbf{z}$ the scaled and shifted filter output signal is component-wise quantized with respect to $\mathbb{Z}$ and re-scaled and re-shifted again

$$\hat{\mathbf{z}} = a\,\left(\mathcal{Q}_{\mathbb{Z}^m}\left\{\frac{1}{a}\,\tilde{\mathbf{z}} - \frac{1}{2}\mathbf{T}^{-1}\mathbf{1}_m\right\} + \frac{1}{2}\mathbf{T}^{-1}\mathbf{1}_m\right)\,. \tag{22}$$

The estimation for the transmit signal is $\hat{\mathbf{s}} = \mathbf{T}\hat{\mathbf{z}}$ and can obviously be rewritten as

$$\hat{\mathbf{s}} = \mathbf{T}\hat{\mathbf{z}} = a\,\mathbf{T}\mathcal{Q}_{\mathbb{Z}^m}\left\{\frac{1}{a}\,\tilde{\mathbf{z}} - \frac{1}{2}\,\mathbf{T}^{-1}\mathbf{1}_m\right\} + \frac{a}{2}\mathbf{1}_m\,. \tag{23}$$

As these scaling and shifting operations are straightforward, they have been omitted within the presentation of the LR-aided detection schemes.

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 *Proc. International Symposium on Signals, Systems, and Electronics (ISSSE)*, Pisa, Italy, September 1998.

[2] U. Fincke and M. Pohst, "Improved Methods for Calculating Vectors of Short Length in a Lattice, Including a Complexity Analysis," *Math. Comp*, vol. 44, pp. 463–471, 1985.

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

[4] 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.

[5] H. Yao and G. Wornell, "Lattice-Reduction-Aided Detectors for MIMO Communication Systems," in *IEEE Proc. Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, November 17-21 2002.

[6] H. Yao, "Effcient Signal, Code, and Receiver Designs for MIMO Communication Systems," Ph.D. dissertation, Massachusetts Institute of Technology, June 2003.

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

[8] ——, "Optimum and Sub-Optimum Lattice-Reduction-Aided Detection and Precoding for MIMO Communications," in *Proc. Canadian Workshop on Information Theory*, Waterloo, Ontario, Canada, May 2003, pp. 88–91.

[9] B. Hassibi, "An Efficient Square-Root Algorithm for BLAST," in *Proc. IEEE Intl. Conf. Acoustic, Speech, Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, pp. 5–9.

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

[11] R. Böhnke, D. Wübben, V. Kühn, and K. D. Kammeyer, "Reduced Complexity MMSE Detection for BLAST Architectures," in *IEEE Proc. Global Communications Conference (GLOBECOM)*, San Francisco, California, USA, December 2003.

[12] T. L. Marzetta, "BLAST training: Estimating Channel Characteristics for High Capacity Space-Time Wireless," in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, Monticell, Illinois, USA, September 1999, pp. 958–966.

[13] 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.

[14] G. Strang, *Linear Algebra and its Applications*, 3rd ed. Orlando, Florida: Harcout Brace Jovanovich College Publishers, 1988.

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

[16] 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.

[17] J. Silverstein and Z. Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.