5

From Linear Equalization to Lattice-Reduction-Aided Sphere-Detector as an Answer to the MIMO Detection Problematic in Spatial Multiplexing Systems

Sébastien Aubert[1] and Manar Mohaisen[2]
[1]*ST-Ericsson/INSA IETR*
[2]*Korea University of Technology and Education (KUT)*
[1]*France*
[2]*Republic of Korea*

1. Introduction

Employing multiple antennas at both the transmitter and the receiver linearly boosts the channel capacity by $\min(n_T, n_R)$, where n_T and n_R are the number of transmit and receive antennas, respectively A. Telatar (1999). Multiple-Input Multiple-Output (MIMO) technologies are classified into three categories: (*i*) MIMO diversity, (*ii*) MIMO Spatial Multiplexing (MIMO-SM) and (*iii*) beamforming that will not be addressed here since it particularly deals with transmitter algorithms. MIMO diversity techniques are deployed to increase the reliability of communications by transmitting or receiving multiple copies of the same signal at different resource entities of the permissible dimensions, *i.e.*, time, frequency, or space. In contrast, the target of MIMO-SM is to increase the capacity of the communication channel. To this end, independent symbols are transmitted simultaneously from the different transmit antennas. Due to its attracting implementation advantages, Vertical Bell Laboratories Layered Space-Time (V-BLAST) transmitter structure is often used in the practical communication systems P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela (1998).

In 3GPP Long Term Evolution-Advanced (3GPP LTE-A) 3GPP (2009), the challenge of de-multiplexing the transmitted symbols via SM techniques, *i.e.* detection techniques, stands as one of the main limiting factors in linearly increasing system's throughput without requiring additional spectral resources. The design of detection schemes with high performance, low latency, and applicable computational complexity is being a challenging research topic due to the power and latency limitations of the mobile communication systems M. Mohaisen, H.S. An, and K.H.Chang (2009).

1.1 System model and problem statement

We consider a MIMO-SM system employing n_T transmit antennas and n_R receive antennas, where $n_T \leq n_R$ P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela (1998). The

This work was supported by ST-Ericsson and by research subsidy for newly-appointed professor of Korea University of Technology and Education for year 2011

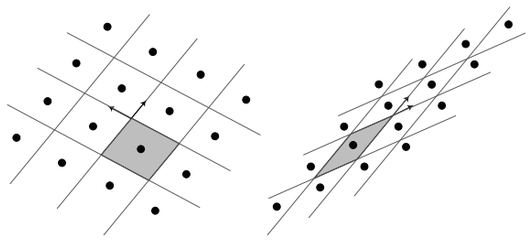
simultaneously transmitted symbols given by the vector $\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}$ are drawn independently from a Quadrature Amplitude Modulation (QAM) constellation, where $\Omega_{\mathbb{C}}$ indicates the constellation set with size $|\Omega_{\mathbb{C}}|$.
Under the assumption of narrow-band flat-fading channel, the received vector $\mathbf{r} \in \mathbb{C}^{n_R}$ is given by:

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

where $\mathbf{n} \in \mathbb{C}^{n_R}$ is the Additive White Gaussian Noise (AWGN) vector whose elements are drawn from i.i.d. circularly symmetric complex Gaussian processes with mean and variance of zero and σ_n^2, respectively. $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ denotes the complex channel matrix whose element $\mathbf{H}_{i,j} \sim \mathcal{CN}(0,1)$ is the channel coefficient between the j-th transmit antenna and the i-th receive antenna.
Working on the transmitted vector $\mathbf{x}$, $\mathbf{H}$ generates the complex lattice

$$\mathcal{L}(\mathbf{H}) = \left\{\mathbf{z} = \mathbf{H}\mathbf{x} | \mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}\right\} = \left\{\mathbf{x}_1\mathbf{H}_1 + \mathbf{x}_2\mathbf{H}_2 + \cdots + \mathbf{x}_{n_T}\mathbf{H}_{n_T} | \mathbf{x}_i \in \Omega_{\mathbb{C}}^{n_T}\right\}, \tag{2}$$

where the columns of $\mathbf{H}$, $\{\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_{n_T}\}$, are known as the basis vectors of the lattice $\mathcal{L} \in \mathbb{C}^{n_R}$ L. Lovász (1986). Also, n_T and n_R refer to the *rank* and *dimension* of the lattice $\mathcal{L}$, respectively, where the lattice is said to be *full-rank* if $n_T = n_R$.



(a) (b)

Fig. 1. Examples of 2-dimensional real lattices with orthogonal bases (a) and correlated bases (b).

Figure 1(a) shows an example of a 2-dimensional real lattice whose basis vectors are $\mathbf{H}_1 = [0.39 \;\; 0.59]^T$ and $\mathbf{H}_2 = [-0.59 \;\; 0.39]^T$, and Figure 1(b) shows another example of a lattice with basis vectors $\mathbf{H}_1 = [0.39 \;\; 0.60]^T$ and $\mathbf{H}_2 = [0.50 \;\; 0.30]^T$. The elements of the transmitted vector $\mathbf{x}$ are withdrawn independently from the real constellation set $\{-3, -1, 1-, 3\}$. Herein we introduce the *orthogonality defect od* which is usually used as a measure of the orthogonality of the lattice basis:

$$od = \frac{\prod_{i=1}^{n_T} \|\mathbf{H}_i\|}{|\det\{\mathbf{H}\}|}, \tag{3}$$

where $\det\{\cdot\}$ refers to determinant and $od \geq 1$. The *od* of the lattices depicted in Figure 1(a) and Figure 1(b) are 1 and 2.28, respectively. This indicates that the first set of basis vectors is perfectly orthogonal while the second set is correlated, which implies inter-layer interferences and induces the advantage of joint detectors among others. The form of the resulting Voronoi regions of the different lattice points, an example is indicated in gray in Figure 1, also indicates the orthogonality of the basis; when the basis vectors are orthogonal with equal norms, the resulting Voronoi regions are squares, otherwise different shapes are obtained.
In light of the above and from a geometrical point of view, the signal detection problem is

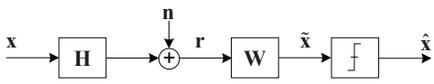


Fig. 2. Block diagram of the linear detection algorithms.

defined as finding the lattice point $\hat{\mathbf{z}} = \mathbf{H}\hat{\mathbf{x}}$, such that $\|\mathbf{r} - \hat{\mathbf{z}}\|^2$ is minimized, where $\|\cdot\|$ is the Euclidean norm and $\hat{\mathbf{x}}$ is the estimate of the transmitted vector $\mathbf{x}$.

1.2 Maximum-likelihood detection

The optimum detector for the transmit vector estimation is the well-known Maximum-Likelihood Detector (MLD) W. Van Etten (1976). MLD employs a brute-force search to find the vector $\mathbf{x}_k$ such that the a-posteriori probability $P\{\mathbf{x}_k \mid \mathbf{r}\}$, $k = 1, 2, \cdots, |\Omega_{\mathbb{C}}|^{n_T}$, is maximized; that is,

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\arg\max} \left(P\{\mathbf{x}_k | \mathbf{r}\} \right). \tag{4}$$

After some basic probability manipulations, the optimization problem in (4) is reduced to:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\arg\max} \left(p\left(\mathbf{r} | \mathbf{x}_k\right) \right), \tag{5}$$

where $p\left(\mathbf{r}|\mathbf{x}_k\right)$ is the probability density function of $\mathbf{r}$ given $\mathbf{x}_k$. By assuming that the elements of the noise vector $\mathbf{n}$ are i.i.d. and follow Gaussian distribution, the noise covariance matrix becomes $\Sigma_n = \sigma_n^2 \mathbf{I}_{n_R}$. As a consequence, the received vector is modelled as a multivariate Gaussian random variable whose mean is $(\mathbf{H}\mathbf{x}_k)$ and covariance matrix is Σ_n. The optimization problem in (5) is rewritten as follows:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\arg\max} \left(\frac{1}{\pi^{n_R} \det\left(\Sigma\right)} \exp^{-(\mathbf{r}-\mathbf{H}\mathbf{x}_k)^H \Sigma_n^{-1} (\mathbf{r}-\mathbf{H}\mathbf{x}_k)} \right) = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\arg\min} \left(\|\mathbf{r} - \mathbf{H}\mathbf{x}_k\|^2 \right). \tag{6}$$

This result coincides with the conjuncture based on the lattice theory given in section 1.1.
The computational complexity of the MLD is known to be exponential in the modulation set size $|\Omega_{\mathbb{C}}|$ and the number of transmit antennas n_T. For mobile communications systems, which are computational complexity and latency limited, MLD becomes infeasible. In the following Sections, we review the conventional sub-optimal detection algorithms J. Wang, and B. Daneshrad (2005), and analyse their advantages and inconveniences.

2. Linear detection algorithms

The idea behind linear detection schemes is to treat the received vector by a filtering matrix $\mathbf{W}$, constructed using a performance-based criterion, as depicted in Figure 2 A. Paulraj, R. Nabar, and D. Gore (2003), C. Windpassenger (2004), B. Schubert (2006). The well known Zero-Forcing (ZF) and Minimum-Mean Square Error (MMSE) performance criteria are used in the Linear ZF (LZF) and MMSE (LMMSE) detectors.

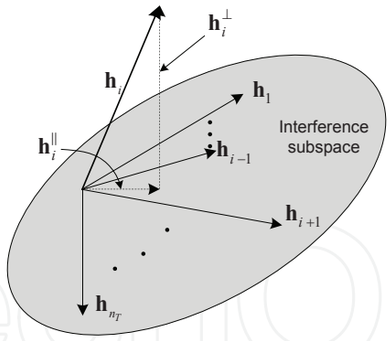


Fig. 3. Geometrical representation of the linear zero-forcing detection algorithm.

2.1 Linear Zero-Forcing detector

LZF detector treats the received vector by the pseudo-inverse of the channel matrix, resulting in full cancellation of the interference with colored noise. The detector in matrix form is given by:

$$\mathbf{W}_{\mathrm{ZF}} = \left(\mathbf{H}\mathbf{H}^H\right)^{-1}\mathbf{H}^H, \tag{7}$$

where $(\cdot)^H$ is the Hermitian transpose.

Figure 3 depicts a geometrical representation of the LZF detector. Geometrically, to obtain the k-th detector's output, the received vector is processed as follows:

$$\tilde{\mathbf{x}}_k = \underline{\mathbf{w}}_k\mathbf{r} = \frac{\left(\mathbf{H}_k^{\perp}\right)^H}{\left\|\mathbf{H}_k^{\perp}\right\|^2}\mathbf{r} = \mathbf{x}_k + \mu_k, \tag{8}$$

where $\underline{\mathbf{w}}_k$ is the k-th row of $\mathbf{W}$, $\mathbf{H}_k^{\perp}$ is the perpendicular component of $\mathbf{H}_i$ on the interference space, and μ_k equals $\underline{\mathbf{w}}_k\mathbf{n}$. Note that $\mathbf{H}_k^{\perp}$ equals $\left(\mathbf{H}_k - \mathbf{H}_k^{||}\right)$, where $\mathbf{H}_k^{||}$ is the parallel component of $\mathbf{H}_k$ to the interference subspace. Then, the mean and variance of the noise μ_k at the output of the LZF detector are 0 and $\|\underline{\mathbf{w}}_k\|^2\sigma_n^2$, respectively. When the channel matrix is ill-conditioned, e.g., if a couple or more of columns of the channel matrix are correlated, $\mathbf{H}_k^{||}$ becomes large, and the noise is consequently amplified.

2.2 Linear minimum-mean square error detector

To alleviate the noise enhancement problem induced by the ZF equalization, the LMMSE can be used. The LMMSE algorithm optimally balances the residual interference and noise enhancement at the output of the detector. To accomplish that, the filtering matrix $\mathbf{W}_{\mathrm{MMSE}}$ is given by:

$$\mathbf{W}_{\mathrm{MMSE}} = \arg\min_{\mathbf{G}} \left(\mathrm{E}\left[\|\mathbf{G}\mathbf{r} - \mathbf{x}\|^2\right]\right), \tag{9}$$

where $\mathrm{E}[\cdot]$ denotes the expectation. Due to the orthogonality between the received vector and the error vector given in (9), we have:

$$\mathrm{E}\left[\left(\mathbf{W}_{\mathrm{MMSE}}\mathbf{r} - \mathbf{x}\right)\mathbf{r}^H\right] = \mathbf{0}, \tag{10}$$

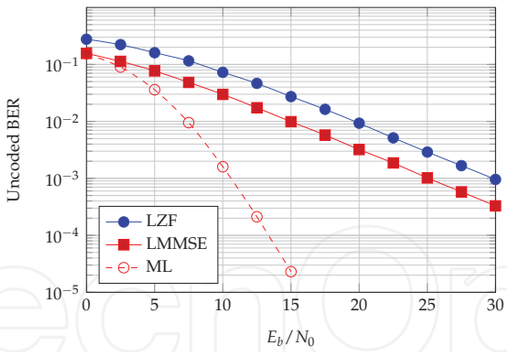


Fig. 4. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, LZF, LMMSE and ML detectors, QPSK modulations at each layer.

and by extending the left side of (10), it directly follows that:

$$\mathbf{W}_{\text{MMSE}} = \left(\Phi_{\mathbf{xx}}^{-1} + \mathbf{H}^H \Phi_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \Phi_{\mathbf{nn}}^{-1} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma_n^2}{\sigma_x^2} \mathbf{I}_{n_T}\right)^{-1} \mathbf{H}^H, \tag{11}$$

where $\Phi_{\mathbf{nn}}$ equals $\sigma_n^2 \mathbf{I}$ and $\Phi_{\mathbf{xx}}$ equals $\sigma_x^2 \mathbf{I}$ are the covariance matrices of the noise and the transmitted vectors, respectively. Theoretically, at high Signal-to-noise Ratio (SNR), the LMMSE optimum filtering converges to the LZF solution. However, we show in M. Mohaisen, and K.H. Chang (2009b) that the improvement by the LMMSE detector over the LZF detector is not only dependent on the plain value of the noise variance, but also on how close σ_n^2 is to the singualr values of the channel matrix. Mathematically, we showed that the ratio between the condition number of the filtering matrices of the linear MMSE and ZF detectors is approximated as follows:

$$\frac{cond(\mathbf{W}_{\text{MMSE}})}{cond(\mathbf{W}_{\text{ZF}})} \approx \frac{1 + \sigma_n^2/\sigma_1^2(\mathbf{H})}{1 + \sigma_n^2/\sigma_N^2(\mathbf{H})}, \tag{12}$$

where σ_1 and σ_N are the maximal and minimal singular values of the channel matrix $\mathbf{H}$, and σ_n^2 is the noise variance. Also, $cond(\mathbf{A}) = (\sigma_1(\mathbf{A})/\sigma_N(\mathbf{A}))$ is the condition number that attains a minimum value of one for orthogonal $\mathbf{A}$.
Figure 4 shows the Bit Error Rate (BER) of the linear detection algorithms in 4×4 MIMO multiplexing system, using 4-QAM signalling. Although the BER performance of LMMSE is close to that of MLD for low E_b/N_0 values, the error rate curves of the two linear detection algorithms have a slope of -1, viz., diversity order equals one, whereas the diversity order of the MLD equals $n_R = 4$.

3. Decision-feedback detection

3.1 Introduction

Although linear detection approaches are attractive in terms of computational complexity, they lead to degradation in the BER performance, due to independent detection of $\mathbf{x}$ components. Superior performance can be obtained if non-linear approaches are employed, as in the Decision-Feedback Detection (DFD) algorithms. In DFD approach, symbols are detected successively, where already-detected components of $\mathbf{x}$ are subtracted out from the

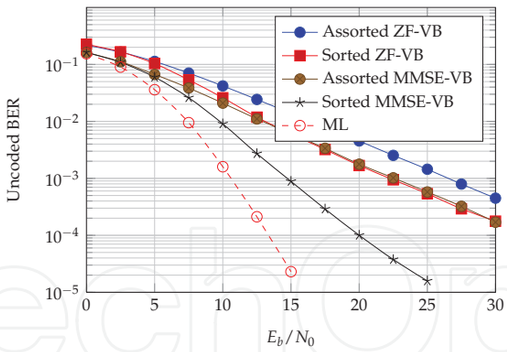


Fig. 5. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, Assorted ZF-VB, Sorted ZF-VB, Assorted MMSE-VB, Sorted MMSE-VB and ML detectors, QPSK modulations at each layer.

received vector. This leads to a system with fewer interferers. In the following two sections, we introduce two categories of DFD algorithms.

3.2 The V-BLAST detection algorithm

In V-BLAST, symbols are detected successively using the aforementioned linear detection approaches. At the end of each iteration, the already-detected component of $\mathbf{x}$ is subtracted out from the received vector P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela (1998), S. Haykin, and M. Moher (2005). Also, the corresponding column of the matrix $\mathbf{H}$ is removed. When decision-feedback approach is used, error propagation becomes a challenging issue. Therefore, the order in which symbols are detected has a great impact on the system performance.

The idea behind the ZF-based V-BLAST (ZF-VB) algorithm is to detect the components of $\mathbf{x}$ that suffer the least noise amplification at first. For the first decision, the pseudo-inverse, *i.e.*, $\mathbf{W}$ equals $\mathbf{H}^{\dagger}$, of the matrix $\mathbf{H}$ is obtained. By assuming that the noise components are i.i.d and that noise is independent of $\mathbf{x}$, then the row of $\mathbf{W}$ with the least Euclidean norm corresponds to the required component of $\mathbf{x}$. That is $k_1 = \arg\min_j \left(\|\underline{\mathbf{w}}_j\|^2\right)$ and $\tilde{\mathbf{x}}_{k_1} = \underline{\mathbf{w}}_{k_1}\mathbf{r}^{(1)}$, where $\mathbf{r}^{(1)} = \mathbf{r}$, the superscript indicates the iteration number, and $\hat{\mathbf{x}}_{k_1} = \mathcal{Q}_{\Omega_c}(\tilde{\mathbf{x}}_{k_1})$ is the decision for the k_1-th component of $\mathbf{x}$. The interference due to the k_1-th symbol is then cancelled out as follows: $\mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \hat{\mathbf{x}}_{k_1}\mathbf{H}_{k_1}$ and $\mathbf{H}^{(2)} = \left[\cdots, \mathbf{H}_{k_1-1}, \mathbf{H}_{k_1+1}, \cdots\right]$. This strategy is repeated up to the last component of $\mathbf{x}$.

In analogy with the linear detection approaches, MMSE-based V-BLAST (MMSE-VB) improves the BER performance, by alleviating the noise enhancement problem. Therefore, the filtering matrix $\mathbf{G}^{(i)}$ at the i-th iteration is given by:

$$\mathbf{G}^{(i)} = \left(\mathbf{H}^{H(i)}\mathbf{H}^{(i)} + \frac{\sigma_n^2}{\sigma_x^2}\mathbf{I}\right)^{-1}\mathbf{H}^{H(i)}, \tag{13}$$

rather than $\mathbf{H}^{(i)\dagger}$ in the case of ZF-VB.

Figure 5 shows the BER performance of the V-BLAST for several detection algorithms. In the assorted V-BLAST schemes, the symbols are detected in an ascending order, *i.e.*, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_T}$, without considering their noise conditions. Obviously, signal ordering leads to improvements for both the ZF-VB and MMSE-VB algorithms, and the improvement is larger in the case of the MMSE-VB algorithm. We note also that the diversity order of the MMSE-VB

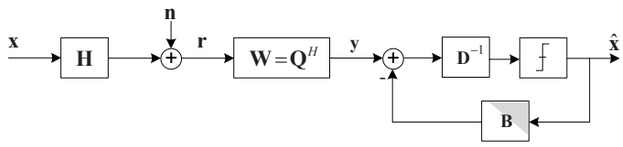


Fig. 6. QRD-DFD.

is larger than 1, unlike the other V-BLAST algorithms which have a diversity order close to 1. This indicates less error propagation in the case of MMSE-VB, compared to other V-BLAST detection algorithms.
The computational complexity of the BLAST detection algorithm, both ZF-VB and MMSE-VB, is $\mathcal{O}(n^4)$, which is infeasible due to power and latency limitations of the mobile communication systems. Although several techniques were proposed to reduce the complexity of the BLAST detection algorithm, it is still complex B. Hassibi (2000); H. Zhu, Z. Lei, and F. Chin (2004); J. Benesty, Y. Huang, and J. Chen (2003).

3.3 QR Decomposition-based detection-feedback detection

The DFD scheme, that lies in the QR Decomposition (QRD) of the channel matrix S. Aubert, M. Mohaisen, F. Nouvel, and K.H. Chang (2010), requires only a fraction of the computational efforts required by the V-BLAST detection algorithm D. Shiu, and J. Kahn (1999). This is why QRD-based DFD (QRD-DFD) is preferable for power and latency limited wireless communication systems.
In QRD-DFD, the channel matrix is decomposed into the multiplication of a unitary matrix $\mathbf{Q} \in \mathbb{C}^{n_R \times n_T}$, *i.e.*, $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_{n_T}$, and an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{n_T \times n_T}$; that is $\mathbf{H} = \mathbf{QR}$.
Then, the received vector is multiplied by the Hermitian transpose of $\mathbf{Q}$, we have

$$\mathbf{y} = \mathbf{Rx} + \mathbf{v} = (\mathbf{D} + \mathbf{B})\,\mathbf{x} + \mathbf{v}, \tag{14}$$

where $\mathbf{y} = \mathbf{Q}^H\mathbf{r}$ and $\mathbf{v} = \mathbf{Q}^H\mathbf{n}$. Note that the noise statistics does not changeable due to the orthogonality of $\mathbf{Q}$. The matrix $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the diagonal elements of $\mathbf{R}$, and $\mathbf{B}$ is strictly upper triangular matrix. As a consequence, the MIMO system becomes spatially causal which implies that:

$$\mathbf{y}_k = \mathbf{R}_{k,k}\tilde{\mathbf{x}}_k + \sum_{i=k+1}^{n_T} \mathbf{R}_{k,j}\hat{\mathbf{x}}_i, \tag{15}$$

where $\tilde{\mathbf{x}}_k$ is a candidate symbol and $\hat{\mathbf{x}}_k$ is the estimate, both for the $k-$th component of $\mathbf{x}$. Therefore,

$$\hat{\mathbf{x}}_k = \mathcal{Q}_{\Omega_C}\left(\frac{\mathbf{y}_k - \sum_{i=k+1}^{n_T} \mathbf{R}_{k,j}\hat{\mathbf{x}}_i}{\mathbf{R}_{k,k}}\right). \tag{16}$$

Note that due to the structure of the matrix $\mathbf{R}$, the last component of $\mathbf{x}$, *i.e.*, x_{n_T}, is interference-free, hence, it can be detected first. Already-detected component of $\mathbf{x}$ is cancelled out to detect the following component. This technique is repeated up to the first component of $\mathbf{x}$, *i.e.*, $\mathbf{x}_1$ D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K.-D. Kammeyer (2001); D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2003); M. Mohaisen, and K.H. Chang (2009a).
Fig. 6 depicts the detailed QRD-based detection algorithm (ZF-QRD). Note that the feedback loop is equivalent to $(\mathbf{D} + \mathbf{B})^{-1} = \mathbf{R}^{-1}$. Figure 7 depicts the BER performance of the the QRD-based DFD algorithms. The MMSE-SQRD algorithm has the best performance but its diversity order converges to unity for high E_b/N_0 values.

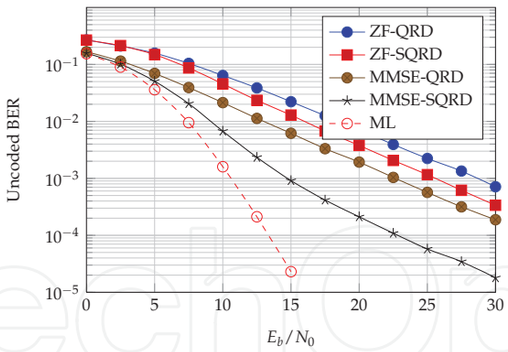


Fig. 7. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, ZF-QRD, ZF-SQRD, MMSE-QRD, MMSE-SQRD and ML detectors, QPSK modulations at each layer.

4. Tree-search detection

Several tree-search detection algorithms have been proposed in the literature that achieve quasi-ML performance while requiring lower computational complexity. In these techniques, the lattice search problem is presented as a tree where nodes represent the symbols' candidates. In the following, we introduce three tree-search algorithms and discuss their advantages and drawbacks.

4.1 Sphere Decoder

The Sphere Decoder (SD) was proposed in the literature to solve several lattice search problems J. Boutros, N. Gresset, L. Brunel, and M. Fossorier (2003). Based on Hassibi and Vikalo analysis, SD achieves quasi-ML performance with polynomial average computational complexity for large range of SNR B. Hassibi, and H. Vikalo (2001). Hence, instead of testing all the hypotheses of the transmitted vector, SD restricts the search in (6) to the lattice points that reside in the hypersphere of radius d and implicitly centred at the unconstrained ZF estimate. Therefore,

$$\hat{\mathbf{x}}_{\mathrm{SD}} = \underset{\mathbf{x}\in\Omega_C^{n_T}}{\arg\min} \left(\|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \leq d^2 \right). \tag{17}$$

The order in which hypotheses are tested at each detection level is defined by the employed search strategy, namely Fincke-Pohst (FP) U. Fincke, and M. Pohst (1985) or Schnorr-Euchner (SE) C. Schnorr, and M. Euchner (1994) strategies. SE strategy orders, and then examines, the hypotheses based on their Euclidean distance from the unconstrained ZF solution; closer hypothesis is tested first. On the other hand, FP strategy tests the hypotheses at each layer without considering the distance from the unconstrained ZF solution. That is why FP strategy leads to higher computational complexity. Figure 8 shows the details of these two strategies for 6-Pulse Amplitude Modulation (PAM) constellation where $\hat{\mathbf{z}}_i$ is the unconstrained ZF solution of the $i-$th transmitted symbol and the numbers under the constellation points represent the order in which hypotheses are tested. It is clear that employing SE strategy leads to reduction in the complexity since the most probable hypothesis for any considered layer, independently of others, is tested first.

It could be shown that the BER performance of the SD coincides with that of the optimum detector. However, some drawbacks remain. In particular, the complexity of SD is variable

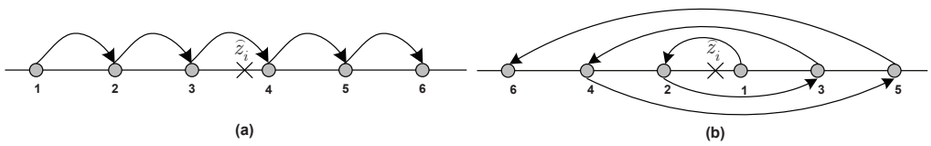


Fig. 8. Search strategies employed by the sphere decoder FP strategy (a) and SE strategy (b).

and depends on the conditionality of the channel matrix and the noise variance, where the worst-case complexity of SD is consequently comparable with that of MLD. That is, the worst-case complexity of SD is exponential. In fact, Jalden and Otterson have shown in J. Jaldén, and B. Ottersten (2005) that even the average complexity of SD is exponential for a fixed SNR value. Also, the SD has a sequential nature because it requires the update of the search radius at every time a new lattice point with smaller accumulative metric is found. This limits the possibility of parallel processing and hence reduces the detection throughput, *i.e.*, increases the detection latency.

To fix the complexity of the detection stage and allow a parallel processing, the QRD with M-algorithm (QRD-M) has been proposed. The QRD-M algorithm and several related complexity reduction techniques are introduced in the following sections. Also, it is equivalently denoted as the K-Best.

4.2 QRD-M detection algorithm

In the QRD-M detection algorithm, only a fixed number, M, of symbol candidates is retained at each detection level J. Anderson, and S. Mohan (1984); K.J. Kim, J. Yue, R.A. Iltis, and J.D. Gibson (2005). At the first detection level, the root node is extended to all the possible candidates of $\mathbf{x}_{n_T}$, the accumulative metrics of the resulting branches are calculated and the best M candidates with the smallest metrics are retained for the next detection level. At the second detection level, the retained M candidates at the previous level are extended to all possible candidates. The resulting $(|\Omega_{\mathbf{C}}| \times M)$ branches are sorted according to their accumulative metrics where the M branches with the smallest accumulative metrics are retained for the next detection level. This strategy is repeated down to the last detection level. By employing the QRD-M strategy, near-ML BER performance are reached for a sufficient M value, as depicted in Figure 9, while the computational complexity of the detection algorithm becomes fixed and only dependent on the size of the modulation set $|\Omega_{\mathbf{C}}|$ and the number of transmit antennas n_T. It also makes a parallel implementation possible. Note that the overall number of visited nodes by the QRD-M algorithm $= (|\Omega_{\mathbf{C}}| + (n_T - 1) \times |\Omega_{\mathbf{C}}| \times M)$.

Although the conventional QRD-M has a fixed complexity which is an advantage, it does not take into consideration the noise and channel conditions. Thus, unnecessary computations are usually done when the channel is well-conditioned and the signal to noise ratio is high. Also, for high $|\Omega_{\mathbf{C}}|$ and n_T, the computational complexity of the QRD-M algorithm becomes high, where as a consequence the detection latency increases.

In order to solve this point, several algorithms have been proposed in the literature. A solution that lies on applying a variable M has been widely studied and is denoted as the dynamic QRD-M algorithm. It offers promising performance results in the case of a SQRD pre-processing step. These algorithms reduce the number of retained candidates at each detection level with tolerable degradation in the performance H. Kawai, K. Higuchi, N. Maeda, and M. Sawahashi (2006); K. Jeon, H. Kim, and H. Park (2006); M. Mohaisen, and K.H. Chang (2010) and references therein. In M. Mohaisen, K.H. Chang, and B.T. Koo (2009), two

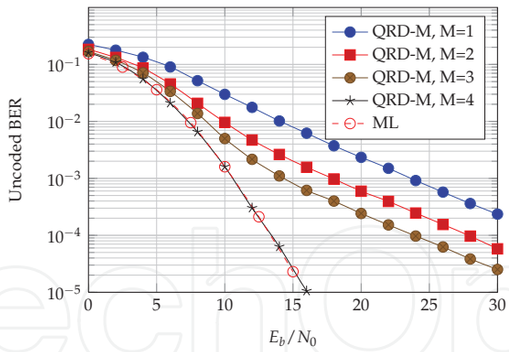


Fig. 9. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, QRD-M algorithm for multiple M values and ML detectors, QPSK modulations at each layer.

algorithms have been proposed, that in addition of reducing the computational complexity of the QRD-M algorithm, they either reduce the processing delay by parallelizing the detection stage or reduce the hardware requirements by iteratively processing the QRD-M algorithm. While a variable M also implies a variable computational complexity, a particular case has to be introduced.

4.3 Fixed-complexity sphere decoder

Fixed-complexity Sphere Decoder (FSD) was proposed by Barbero *et al.* to overcome the aforementioned drawbacks of SD. FSD achieves a quasi-ML performance by performing the following two-stage tree search L. Barbero, and J. Thompson (2008a;b):

- Full expansion: In the first p levels, a full expansion is performed, where all symbols replicas candidates are retained to the following levels;
- Single expansion: A single expansion of each retained branch is done in the remaining $(n_T - p)$ levels, where only the symbol replica candidate with the lowest accumulative metric is considered for next levels.

Because all possible symbols candidates are retained in the first p levels, the reliability of signals detected in these levels does not affect the final detection performance compared to MLD. Therefore, signals with the least robustness are detected in the full expansion stage. On the other hand, in the remaining $(n_T - p)$ levels, signals are sorted based on their reliability, where signals with the least noise amplification are detected first. In the conventional FSD, the V-BLAST algorithm is employed to obtain the required signal ordering by the FSD.
Figure 10 depicts the BER performance of the FSD for $p = 1$ in 4×4 MIMO-SM system using 4-QAM. Results show that the ordering has a crucial effect on the the performance of the FSD. For instance, both the performance and the attained diversity order are degraded when the ordering stage is skipped or when a non-optimal signal ordering is used. A low complexity FSD ordering scheme that requires a fraction of the computations of the V-BLAST scheme was proposed in M. Mohaisen, and K.H. Chang (2009a), where a close to optimum performance was achieved by embedding the sorting stage in the QR factorization of the channel matrix.

4.4 MMSE-centred sphere decoder

The SD principle may be extended. By defining the Babai point - only in the case of a depth-first search algorithm - as the first obtained solution by the algorithm, the induced Babai point in this case is implicitly a ZF-QRD. In the case of a QRD-M algorithm, this

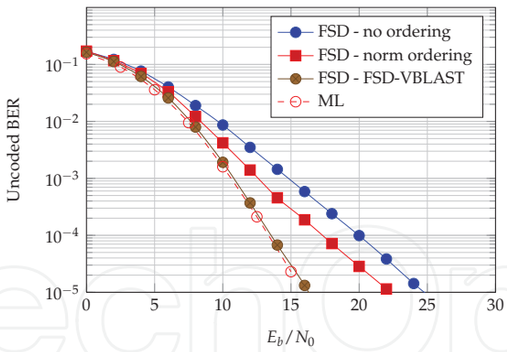


Fig. 10. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, FSD algorithms with $p = 1$ and ML detectors, QPSK modulations at each layer.

definition is extended and is considered as the solution that would be directly reached, without neighborhood study. Another useful notation that has to be introduced is the sphere search around the center search $\mathbf{x}_C$, namely the signal in any equation of the form $\|\mathbf{x}_C - \mathbf{x}\|^2 \leq d^2$, where $\mathbf{x}$ is any possible hypothesis of the transmitted vector $\mathbf{x}$, which is consistent with the equation of an $(n_T - 1)-$ sphere.

Classically, the SD formula is centred at the unconstrained ZF solution and the corresponding detector is denoted in the sequel as the naïve SD. Consequently, a fundamental optimization may be considered by introducing an efficient search center that results in an already close-to-optimal Babai point. In other words, to obtain a solution that is already close to the ML solution. This way, it is clear that the neighborhood study size can be decreased without affecting the outcome of the search process. In the case of the QRD-M algorithm, since the neighborhood size is fixed, it will induce a performance improvement for a given M or a reduction of M for a given target BER.

The classical SD expression may be re-arranged, leading to an exact formula that has been firstly proposed by Wong *et al.*, aiming at optimizations for a VLSI implementation through an efficient Partial Euclidean distance (PED) expression and early pruned nodes K.-W. Wong, C.-Y. Tsui, S.-K. Cheng, and W.-H. Mow (2002):

$$\mathbf{x}_{ZF-DFD} = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \|\mathbf{R}\mathbf{e}_{ZF}\|^2, \tag{18}$$

where $\mathbf{e}_{ZF} = \mathbf{x}_{ZF} - \mathbf{x}$ and $\mathbf{x}_{ZF} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{r}$. Equation (18) clearly exhibits the point that the naïve SD is unconstrained ZF-centred and implicitly corresponds to a ZF-QRD procedure with a neighborhood study at each layer.

The main idea proposed by B.M. Hochwald, and S. ten Brink (2003); L. Wang, L. Xu, S. Chen, and L. Hanzo (2008); T. Cui, and C. Tellambura (2005) is to choose a closer-to-ML Babai point than the ZF-QRD, which is the case of the MMSE-QRD solution. For sake of clearness with definitions, we say that two ML equations are equivalent if the lattice points argument outputs of the minimum distance are the same, even in the case of different metrics. Two ML equations are equivalent iff:

$$\underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \{\|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2\} = \underset{\mathbf{x} \in \Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \{\|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 + c\}, \tag{19}$$

where c is a constant.

In particular, Cui *et al.* T. Cui, and C. Tellambura (2005) proposed a general equivalent

minimization problem: $\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}\in\Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \{\|\mathbf{r}-\mathbf{H}\mathbf{x}\|^2 + \alpha\mathbf{x}^H\mathbf{x}\}$, by noticing that signals $\mathbf{x}$ have to be of constant modulus. This assumption is obeyed in the case of QPSK modulation and is not directly applicable to 16-QAM and 64-QAM modulations, even if this assumption is not limiting since a QAM constellation can be considered as a linear sum of QPSK points T. Cui, and C. Tellambura (2005).

This expression has been applied to the QRD-M algorithm by Wang *et al.* in the case of the unconstrained MMSE-center which leads to an MMSE-QRD procedure with a neighborhood study at each layer L. Wang, L. Xu, S. Chen, and L. Hanzo (2008). In this case, the equivalent ML equation is rewritten as:

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}\in\Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \left(\mathbf{x}_C - \mathbf{x}\right)^H \left(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}\right) \left(\mathbf{x}_C - \mathbf{x}\right). \tag{20}$$

Through the use of the Cholesky Factorization (CF) of $\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I} = \mathbf{U}^H\mathbf{U}$ in the MMSE case ($\mathbf{H}^H\mathbf{H} = \mathbf{U}^H\mathbf{U}$ in the ZF case), the ML expression equivalently rewrites:

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}\in\Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \left(\tilde{\mathbf{x}} - \mathbf{x}\right)^H \mathbf{U}^H\mathbf{U} \left(\tilde{\mathbf{x}} - \mathbf{x}\right), \tag{21}$$

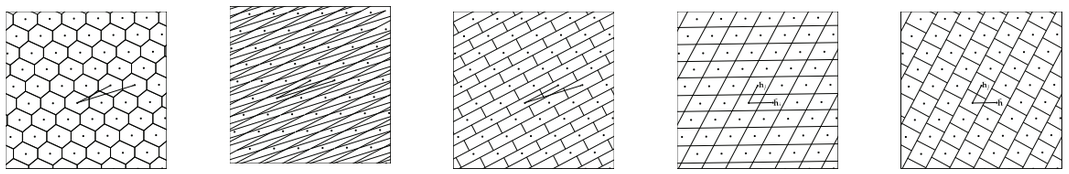
where $\mathbf{U}$ is upper triangular with real elements on diagonal and $\tilde{\mathbf{x}}$ is any (ZF or MMSE) unconstrained linear estimate.

5. Lattice reduction

For higher dimensions, the ML estimate can be provided correctly with a reasonable complexity using a Lattice Reduction (LR)-aided detection technique.

5.1 Lattice reduction-aided detectors interest

As proposed in H. Yao, and G.W. Wornell (2002), LR-Aided (LRA) techniques are used to transform any MIMO channel into a better-conditioned (short basis vectors norms and roughly orthogonal) equivalent MIMO channel, namely generating the same lattice points. Although classical low-complexity linear, and even (O)DFD detectors, fail to achieve full diversity as depicted in D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004), they can be applied to this equivalent (the exact definition will be introduced in the sequel) channel and *significantly improve performance* C. Windpassinger, and R.F.H. Fischer (2003). In particular, it has been shown that LRA detectors achieve the full diversity C. Ling (2006); M. Taherzadeh, A. Mobasher, and A.K. Khandani (2005); Y.H. Gan, C. Ling, and W.H. Mow (2009). By assuming $i < j$, Figure 11 depicts the decision regions in a trivial two-dimensional case and demonstrates to the reader the reason why LRA detection algorithms offer better performance by approaching the optimal ML decision areas D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004). From a singular value theory point of view, when the lattice basis is reduced, its singular values becomes much closer to each other with equal singular values for orthogonal basis. Therefore, the power of the system will be distributed almost equally on the singular values and the system become more immune against the noise enhancement problem when the singular values are inverted during the equalization process.



(a) ML (b) LD (c) DFD (d) LRA-LD (e) LRA-DFD

Fig. 11. Undisturbed received signals and decision areas of (a) ML, (b) LD, (c) DFD, (d) LRA-LD and (e) LRA-DFD D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004).

5.2 Summary of the lattice reduction algorithms

To this end, various reduction algorithms, namely the optimal (the orthogonality is maximized) but NP-hard Minkowski B.A. Lamacchia (1991), Korkine-Zolotareff B.A. Lamacchia (1991) algorithms E. Agrell , T. Eriksson, A. Vardy, and K. Zeger (2002), the well-known LLL reduction A.K. Lenstra, H.W. Lenstra, and L. Lovász (1982), and Seysen's B.A. Lamacchia (1991); M. Seysen (1993) LR algorithm have been proposed.

5.3 Lattice definition

By interpreting the columns $\mathbf{H}_i$ of $\mathbf{H}$ as a *generator basis* , note that $\mathbf{H}$ is also referred to as the lattice basis whose columns are referred to as "basis vectors", the lattice $\Lambda(\mathbf{H})$ is defined as all the complex integer combinations of $\mathbf{H}_i$, *i.e.*,

$$\Lambda(\mathbf{H}) \triangleq \left\{ \sum_{i=1}^{n_T} a_i \mathbf{H}_i \mid a_i \in \mathbb{Z}_{\mathbb{C}} \right\}, \tag{22}$$

where $\mathbb{Z}_{\mathbb{C}}$ is the set of complex integers which reads: $\mathbb{Z}_{\mathbb{C}} = \mathbb{Z} + j\mathbb{Z}$, $j^2 = -1$.
The lattice $\Lambda(\tilde{\mathbf{H}})$ generated by the matrix $\tilde{\mathbf{H}}$ and the lattice generated by the matrix $\mathbf{H}$ are *identical* iff all the lattice points are the same. The two aforementioned bases generate an identical lattices iff $\tilde{\mathbf{H}} = \mathbf{HT}$, where the $n_T \times n_T$ transformation matrix is *unimodular* E. Agrell , T. Eriksson, A. Vardy, and K. Zeger (2002), *i.e.*, $\mathbf{T} \in \mathbb{Z}_{\mathbb{C}}^{n_T \times n_T}$ and such that $|\det(\mathbf{T})| = 1$.
Using the reduced channel basis $\tilde{\mathbf{H}} = \mathbf{HT}$ and introducing $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$, the system model given in (1) can be rewritten D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004):

$$\mathbf{r} = \tilde{\mathbf{H}}\mathbf{z} + \mathbf{n}. \tag{23}$$

The idea behind LRA equalizers or detectors is to consider the identical system model above. The detection is then performed with respect to the reduced channel matrix ($\tilde{\mathbf{H}}$), which is now roughly orthogonal by definition, and to the equivalent transmitted signal that still belongs to an integer lattice since $\mathbf{T}$ is unimodular D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004). Finally, the estimated $\hat{\mathbf{x}}$ in the original problem is computed with the relationship $\tilde{\mathbf{x}} = \mathbf{T}\hat{\mathbf{z}}$ and by hard-limiting $\tilde{\mathbf{x}}$ to a valid symbol vector. These steps are summarized in the block scheme in Figure 12.

The following Subsections briefly describe the main aspects of the LLL Algorithm (LA) and the Seysen's Algorithm (SA).

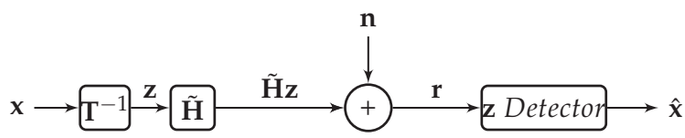


Fig. 12. LRA detector bloc scheme.

5.4 LLL algorithm

The LA is a local approach that transforms an input basis $\mathbf{H}$ into an LLL-reduced basis $\tilde{\mathbf{H}}$ that satisfies both of the orthogonality and norm reduction conditions, respectively:

$$|\Re\{\mu_{i,j}\}|,\ |\Im\{\mu_{i,j}\}| \leq \tfrac{1}{2},\ \forall\ 1 \leq j < i \leq n_T, \tag{24}$$

where $\mu_{i,j} \triangleq \frac{<\mathbf{H}_i,\ \tilde{\mathbf{H}}_j>}{\|\tilde{\mathbf{H}}_j\|^2}$, and:

$$\|\tilde{\mathbf{H}}_i\|^2 = (\delta - |\mu_{i,i-1}|^2)\|\tilde{\mathbf{H}}_{i-1}\|^2,\ \forall\ 1 < i \leq n_T, \tag{25}$$

where δ, with $\frac{1}{2} < \delta < 1$, is a factor selected to achieve a good quality-complexity trade-off A.K. Lenstra, H.W. Lenstra, and L. Lovász (1982). In this book chapter, δ is assumed to be $\delta = \frac{3}{4}$, as commonly suggested, and $\tilde{\mathbf{H}}_i = \tilde{\mathbf{H}}_i - \sum_{j=1}^{i-2}\{\lceil \mu_{i,j} \rfloor \mathbf{H}_j\}$. Another classical result consists of directly considering the Complex LA (CLA) that offers a saving in the average complexity of nearly 50% compared to the straightforward real model system extension with negligible performance degradation Y.H. Gan, C. Ling, and W.H. Mow (2009).
Let us introduce the QR Decomposition (QRD) of $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ that reads $\mathbf{H} = \mathbf{QR}$, where the matrix $\mathbf{Q} \in \mathbb{C}^{n_R \times n_T}$ has orthonormal columns and $\mathbf{R} \in \mathbb{C}^{n_T \times n_T}$ is an upper-triangular matrix. It has been shown D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004) the QRD of $\mathbf{H} = \mathbf{QR}$ is a possible starting point for the LA, and it has been introduced L.G. Barbero, T. Ratnarajah, and C. Cowan (2008) that the Sorted QRD (SQRD) provides a better starting point since it finally leads to a significant reduction in the expected computational complexity D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004) and in the corresponding variance B. Gestner, W. Zhang, X. Ma, and D.V. Anderson (2008).
By denoting the latter algorithm as the SQRD-based LA (SLA), these two points are depicted in Figure 13 (a-c) under DSP implementation-oriented assumptions on computational complexities (see S. Aubert, M. Mohaisen, F. Nouvel, and K.H. Chang (2010) for details).
Instead of applying the LA to the only basis $\mathbf{H}$, a simultaneous reduction of the basis $\mathbf{H}$ and the dual basis $\mathbf{H}^{\#} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$ D. Wübben, and D. Seethaler (2007) may be processed.

5.5 Seysen's algorithm

At the beginning, let us introduce the *Seysen's orthogonality measure* M. Seysen (1993)

$$\mathcal{S}(\tilde{\mathbf{H}}) \triangleq \sum_{i=1}^{n_T} \left\|\tilde{\mathbf{H}}_i\right\|^2 \left\|\tilde{\mathbf{H}}_i^{\#}\right\|^2, \tag{26}$$

where $\tilde{\mathbf{H}}_i^{\#}$ is the i-th basis vector of the dual lattice, *i.e.*, $\tilde{\mathbf{H}}^{\#H}\tilde{\mathbf{H}} = \mathbf{I}$.
The SA is a global approach that transforms an input basis $\mathbf{H}$ (and its dual basis $\mathbf{H}^{\#}$) into a Seysen-reduced basis $\tilde{\mathbf{H}}$ that (locally) minimizes $\mathcal{S}$ and that satisfies, $\forall\ 1 \leq i \neq j \leq n_T$ D. Seethaler, G. Matz, and F. Hlawatsch (2007)

$$\lambda_{i,\,j} \triangleq \left\lfloor \frac{1}{2}\left(\frac{\tilde{\mathbf{H}}_j^{\#H}\tilde{\mathbf{H}}_i^{\#}}{\|\tilde{\mathbf{H}}_i^{\#}\|^2} - \frac{\tilde{\mathbf{H}}_j^{H}\tilde{\mathbf{H}}_i}{\|\tilde{\mathbf{H}}_j^{\#}\|^2}\right)\right\rceil = 0. \tag{27}$$

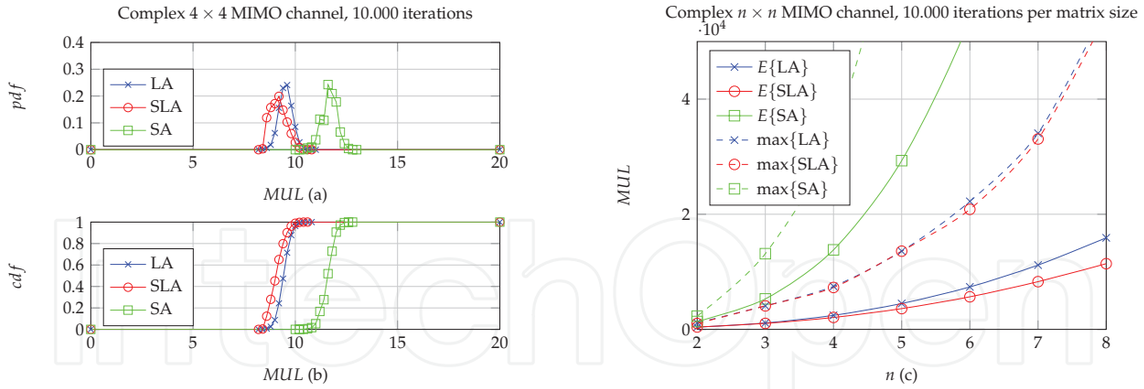


Fig. 13. PDF (a) and CDF (b) of the number of equivalent MUL of the LA, SLA and SA, and average and maximum total number of equivalent MUL of the LA, SLA and SA as a function of the number of antennas n (c).

SA computational complexity is depicted in Figure 13 (a-c) as a function of the number of equivalent *real multiplication* MUL, which allow for some discussion.

5.5.1 Concluding remarks

The aforementioned LR techniques have been presented and both their performances (orthogonality of the obtained lattice) D. Wübben, and D. Seethaler (2007) and computational complexities L.G. Barbero, T. Ratnarajah, and C. Cowan (2008) have been compared when applied to MIMO detection in the Open Loop (OL) case. In Figure 14 (a-f), the od, $cond$, and S of the reduced basis provided by the SA compared to the LA and SLA are depicted. These measurements are known to be popular measures of the quality of a basis for data detection C. Windpassinger, and R.F.H. Fischer (2003). However, this orthogonality gain is obtained at the expense of a higher computational complexity, in particular compared to the SLA. Moreover, it has been shown that a very tiny uncoded BER performance improvement is offered in the case of LRA-LD only D. Wübben, and D. Seethaler (2007). In particular, in the case of LRA-DFD detectors, both LA and SA yield almost the same performance L.G. Barbero, T. Ratnarajah, and C. Cowan (2008).

According to the curves depicted in Figure 13 (a), the mean computational complexities of LA, SLA and SA are $1,6.10^4$, $1,1.10^4$ and $1,4.10^5$ respectively in the case of a 4×4 complex matrix. The variance of the computational complexities of LA, SLA and SA are 3.10^7, $2,3.10^7$ and $2,4.10^9$ respectively, which illustrates the aforementioned reduction in the mean computational complexity and in the corresponding variance and consequently highlights the SLA advantage over other LR techniques.

In Figure 14, the Probability Density Function (PDF) and Cumulative Density Function (CDF) of $ln(cond)$, $ln(od)$ and $ln(S)$ for LA, SLA and SA are depicted and compared to the performance without lattice reduction. It can be observed that both LA and SLA offer exactly the same performance, with the only difference in their computational complexities. Also, there is a tiny improvement in the od when SA is used as compared to (S)LA. This point will be discussed in the sequel.

The LRA algorithm preprocessing step has been exposed and implies some minor modifications in the detection step.

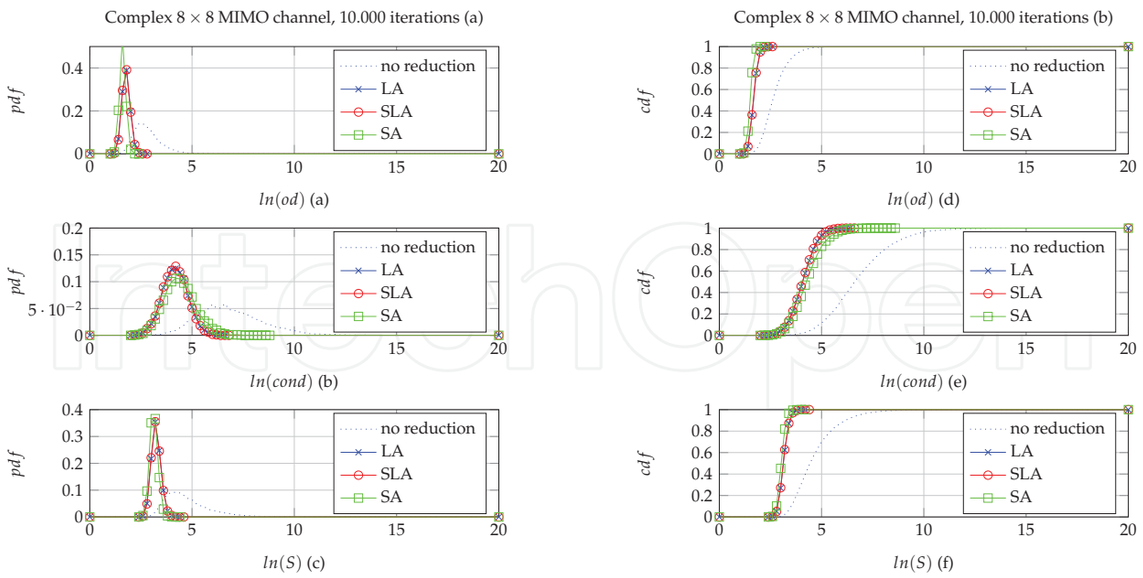


Fig. 14. PDF (a-c) and CDF (d-f) of $ln(cond)$ (a, d), $ln(od)$ (b, e) and $ln(S)$ (c, f) by application of the LA, SLA and SA and compared to the original basis.

5.6 Lattice reduction-aided detection principle

The key idea of the LR-aided detection schemes is to understand that the finite set of transmitted symbols $\Omega_{\mathbb{C}}^{n_T}$ can be interpreted as the De-normalized, Shifted then Scaled (DSS) version of the infinite integer subset $\mathcal{Z}_{\mathbb{C}}^{n_T} \subset \mathbb{Z}_{\mathbb{C}}^{n_T}$ C. Windpassinger, and R.F.H. Fischer (2003), where $\mathbb{Z}_{\mathbb{C}}^{n_T}$ is the infinite set of complex integers, *i.e.*,:

$$\Omega_{\mathbb{C}}^{n_T} = 2a\mathcal{Z}_{\mathbb{C}}^{n_T} + \tfrac{1}{2}\mathbf{T}^{-1}\mathbf{1}_{\mathbb{C}}^{n_T}), \tag{28}$$

and reciprocally

$$\mathcal{Z}_{\mathbb{C}}^{n_T} = \tfrac{1}{2a}\Omega_{\mathbb{C}}^{n_T} - \tfrac{1}{2}\mathbf{T}^{-1}\mathbf{1}_{\mathbb{C}}^{n_T}, \tag{29}$$

where a is a power normalization coefficient (*i.e.*, $1/\sqrt{2}$, $1/\sqrt{10}$ and $1/\sqrt{42}$ for QPSK, 16QAM and 64QAM constellations, respectively) and $\mathbf{1}_{\mathbb{C}}^{n_T} \in \mathbb{Z}_{\mathbb{C}}^{n_T}$ is a complex displacement vector (*i.e.*, $\mathbf{1}_{\mathbb{C}}^{n_T} = [1+j, \cdots, 1+j]^T$ in the complex case).

At this step, a general notation is introduced. Starting from the system equation, it can be rewritten equivalently in the following form, by de-normalizing, by dividing by two and subtracting $\mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}/2$ from both sides:

$$\frac{\mathbf{r}}{2a} - \frac{\mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}}{2} = \frac{\mathbf{H}\mathbf{x}}{2a} + \frac{\mathbf{n}}{2a} - \frac{\mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}}{2} \Leftrightarrow \frac{1}{2}\left(\frac{\mathbf{r}}{a} - \mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}\right) = \mathbf{H}\frac{1}{2}\left(\frac{\mathbf{x}}{a} - \mathbf{1}_{\mathbb{C}}^{n_T}\right) + \frac{1}{2a}\mathbf{n}, \tag{30}$$

where $\mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}$ is a simple matrix-vector product to be done at each channel realization.

By introducing the DSS signal $\mathbf{r}_{\mathbb{Z}} = \frac{1}{2}\left(\frac{\mathbf{r}}{a} - \mathbf{H}\mathbf{1}_{\mathbb{C}}^{n_T}\right) = dss\{\mathbf{r}\}$ and the re-Scaled, re-Shifted then Normalized (SSN) signal $\mathbf{x}_{\mathbb{Z}} = \frac{1}{2}\left(\frac{\mathbf{x}}{a} - \mathbf{1}_{\mathbb{C}}^{n_T}\right) = ssn\{\mathbf{x}\}$, which makes both belonging to $\mathbf{H}\mathcal{Z}_{\mathbb{C}}^{n_T}$ and $\mathcal{Z}_{\mathbb{C}}^{n_T}$, respectively, the expression reads:

$$\mathbf{r}_{\mathbb{Z}} = \mathbf{H}\mathbf{x}_{\mathbb{Z}} + \frac{\mathbf{n}}{2a}. \tag{31}$$

This intermediate step allows to define the symbols vector in the reduced transformed constellation through the relation $\mathbf{z}_{\mathbb{Z}} = \mathbf{T}^{-1}\mathbf{x}_{\mathbb{Z}} \in \mathbf{T}^{-1}\mathcal{Z}^{n_T} \subset \mathbb{Z}^{n_T}$. Finally, the lattice-reduced channel and reduced constellation expression can be introduced:

$$\mathbf{r}_{\mathbb{Z}} = \tilde{\mathbf{H}}\mathbf{z}_{\mathbb{Z}} + \frac{\mathbf{n}}{2a}. \tag{32}$$

The LRA detection steps comprise the $\hat{\mathbf{z}}_{\mathbb{Z}}$ estimation of $\mathbf{z}_{\mathbb{Z}}$ with respect to $\mathbf{r}_{\mathbb{Z}}$ and the mapping of these estimates onto the corresponding symbols belonging to the $\Omega_{\mathbb{C}}^{n_T}$ constellation through the $\mathbf{T}$ matrix. In order to finally obtain the $\hat{\mathbf{x}}$ estimation of $\mathbf{x}$, the DSS $\tilde{\mathbf{x}}_{\mathbb{Z}}$ signal is obtained following the $\tilde{\mathbf{z}}_{\mathbb{Z}}$ quantization with respect to $\mathbb{Z}_{\mathbb{C}}^{n_T}$ and re-scaled, re-shifted, then normalized again.
The estimation for the transmit signal is $\hat{\mathbf{x}} = \mathcal{Q}_{\Omega_{\mathbb{C}}^{n_T}}\{\tilde{\mathbf{x}}\}$, as described in the block scheme in Figure 15 in the case of the LRA-ZF solution, and can be globally rewritten as

$$\hat{\mathbf{x}} = \mathcal{Q}_{\Omega_{\mathbb{C}}^{n_T}}\left\{a\left(2\mathbf{T}\mathcal{Q}_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{\mathbf{z}}_{\mathbb{Z}}\} + \mathbf{1}_{\mathbb{C}}^{n_T}\right)\right\}, \tag{33}$$

where $\mathcal{Q}_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\cdot\}$ denotes the quantization operation of the n_T-th dimensional integer lattice, for which per-component quantization is such as $\mathcal{Q}_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\mathbf{x}\} = [\lfloor \mathbf{x}_1 \rceil, \cdots, \lfloor \mathbf{x}_{n_T} \rceil]^T$, where $\lfloor \cdot \rceil$ denotes the rounding to the nearest integer.

Due to its performance versus complexity, the LA is a widely used reduction algorithm.

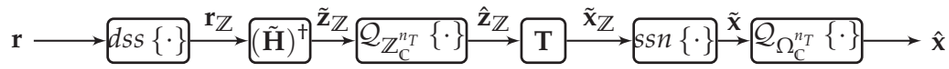


Fig. 15. LRA-ZF detector block scheme.

This is because SA requires a high additional computations compared to the LA to achieve a small, even negligible, gain in the BER performance L.G. Barbero, T. Ratnarajah, and C. Cowan (2008), as depicted in Figure 14. Based on this conjecture, LA will be considered as the LR technique in the remaining part of the chapter.
Subsequently to the aforementioned points, the SLA computational complexity has been shown J. Jaldén, D. Seethaler, and G. Matz (2008) to be unbounded through distinguishing the SQRD pre-processing step and the LA related two conditions. In particular, while the SQRD offers a polynomial complexity, the key point of the SLA computational complexity estimation lies in the knowledge of the number of iterations of both conditions. Since the number of iterations depends on the condition number of the channel matrix, it is consequently unbounded J. Jaldén, D. Seethaler, and G. Matz (2008), which leads to the conclusion that the worst-case computational complexity of the LA in the Open Loop (OL) case is exponential in the number of antennas. Nevertheless, the mean number of iterations (and consequently the mean total computational complexity) has been shown to be polynomial J. Jaldén, D. Seethaler, and G. Matz (2008) and, therefore, a thresholded-based version of the algorithm offers convenient results. That is, the algorithm is terminated when the number of iterations exceeds a pre-defined number of iterations.

5.7 Simulation results

In the case of LRA-LD, the quantization is performed on $\mathbf{z}$ instead of $\mathbf{x}$. The unconstrained LRA-ZF equalized signal $\tilde{\mathbf{z}}_{LRA-ZF}$ are denoted $(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^H\mathbf{r}$ and $\mathbf{T}^{-1}\tilde{\mathbf{x}}_{ZF}$, simultaneously D.

Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004). Consequently, the LRA-ZF estimate is $\hat{\mathbf{x}} = \mathcal{Q}_{\Omega_c^{n_T}}\{\mathbf{T}\mathcal{Q}_{\mathbb{Z}^{n_T}}\{\tilde{\mathbf{z}}_{LRA-ZF}\}\}$. Identically, the LRA-MMSE estimate is given as $\hat{\mathbf{x}} = \mathcal{Q}_{\Omega_c^{n_T}}\{\mathbf{T}\mathcal{Q}_{\mathbb{Z}^{n_T}}\{\tilde{\mathbf{z}}_{LRA-MMSE}\}\}$, considering the unconstrained LRA-MMSE equalized signal $\tilde{\mathbf{z}}_{LRA-MMSE} = (\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + \sigma^2\mathbf{T}^H\mathbf{T})^{-1}\tilde{\mathbf{H}}^H\mathbf{r}$.
It has been shown D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004) that the consideration of the MMSE criterion by reducing the extended channel matrix $\mathbf{H}_{ext} = [\mathbf{H}; \sigma\mathbf{I}_{n_T}]$, leading to $\tilde{\mathbf{H}}_{ext}$, and the corresponding extended receive vector $\mathbf{r}_{ext}$ leads to both an important performance improvement and while reducing the computational complexity compared to the straightforward solution. In this case, not only the $\tilde{\mathbf{H}}$ conditioning is considered but also the noise amplification, which is particularly of interest in the case of the LRA-MMSE linear detector. In the sequel, this LR-Aided linear detector is denoted as LRA-MMSE Extended (LRA-MMSE-Ext) detector.
The imperfect orthogonality of the reduced channel matrix induces the advantageous use of DFD techniques D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004). By considering the QRD outputs of the SLA, namely $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$, the system model rewrites $\tilde{\mathbf{z}}_{LR-ZF-QRD} = \tilde{\mathbf{Q}}^H\mathbf{r}$ and reads simultaneously $\tilde{\mathbf{R}}\mathbf{z} + \tilde{\mathbf{Q}}^H\mathbf{n}$. The DFD procedure can than then be performed in order to iteratively obtain the $\hat{\mathbf{z}}$ estimate. In analogy with the LRA-LD, the extended system model can be considered. As a consequence, it leads to the LRA-MMSE-QRD estimate that can be obtained via rewriting the system model as $\tilde{\mathbf{z}}_{LR-MMSE-QRD} = \tilde{\mathbf{Q}}_{ext}^H\mathbf{r}_{ext}$ and reads simultaneously $\tilde{\mathbf{R}}_{ext}\mathbf{z} + \tilde{\mathbf{n}}$, where $\tilde{\mathbf{n}}$ is a noise term that also includes residual interferences.
Figure 16 shows the uncoded BER performance versus E_b/N_0 (in dB) of some well-established LRA-(pseudo) LDs, for a 4×4 complex MIMO Rayleigh system, using QPSK modulation (a, c) and 16QAM (b, d) at each layer. The aforementioned results are compared to some reference results; namely, ZF, MMSE, ZF-QRD, MMSE-QRD and ML detectors. It has been shown that the (S)LA-based LRA-LDs achieve the full diversity M. Taherzadeh, A. Mobasher, and A.K. Khandani (2005) and consequently offer a strong improvement compared to the common LDs. The advantages in the LRA-(Pseudo)LDs are numerous. First, they constitute efficient detectors in the sense of the high quality of their hard outputs, namely the ML diversity is reached within a constant offset, while offering a low overall computational complexity.Also, by noticing that the LR preprocessing step is independent of the SNR, a promising aspect concerns the Orthogonal Frequency-Division Multiplexing (OFDM) extension that would offer a significant computational complexity reduction over a whole OFDM symbol, due to both the time and coherence band. However, there remains some important drawbacks. In particular, the aforementioned SNR offset is important in the case of high order modulations, namely 16-QAM and 64-QAM, despite some aforementioned optimizations. Another point is the LR algorithm's sequential nature because of its iterative running, which consequently limits the possibility of parallel processing. The association of both LR and a neighborhood study is a promising, although intricate, solution for solving this issue. For a reasonable K, a dramatic performance loss is observed with classical K-Best detectors in Figure 9. For a low complexity solution such as LRA-(Pseudo) LDs, a SNR offset is observed in Figure 16. Consequently, the idea that consists in reducing the SNR offset by exploring a neighborhood around a correct although suboptimal solution becomes obvious.

6. Lattice reduction-aided sphere decoding

While it seems to be computationally expensive to cascade two NP-hard algorithms, the promising perspective of combining both the algorithms relies on achieving the ML diversity

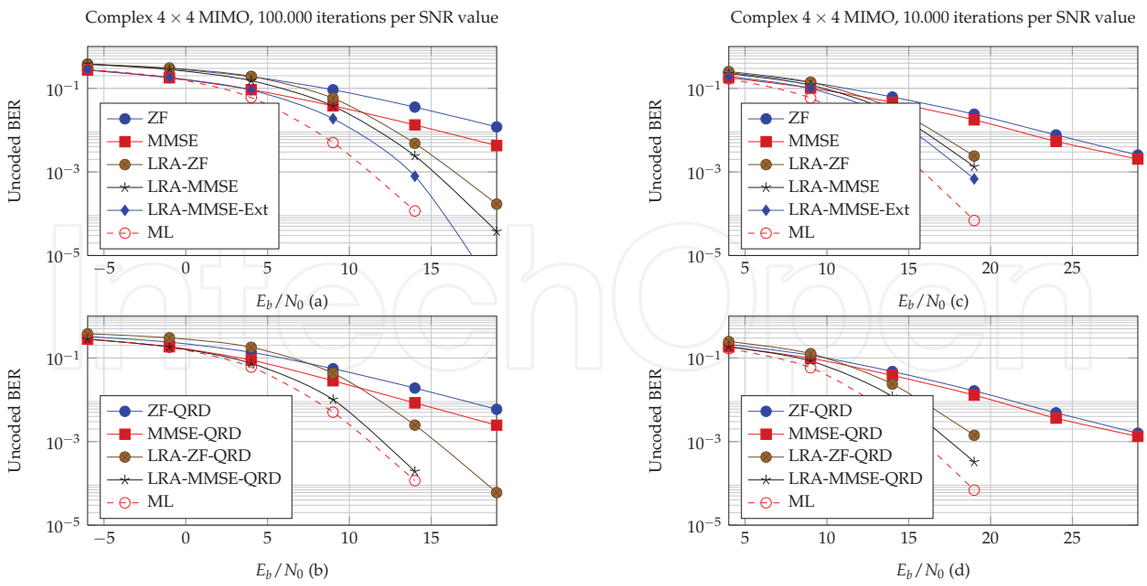


Fig. 16. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, ZF, MMSE, LRA-ZF, LRA-MMSE, LRA-MMSE-Ext and ML detectors (a, c), ZF-QRD, MMSE-QRD, LRA-ZF-QRD, LRA-MMSE-QRD and ML detectors (b, d), QPSK modulations at each layer (a-b) and 16QAM modulations at each layer (c-d).

through a LRA-(Pseudo)LD and on reducing the observed SNR offset thanks to an additional neighborhood study. This idea senses the neighborhood size would be significantly reduced while near-ML results would still be reached.

6.1 Lattice reduction-aided neighborhood study interest

Contrary to LRA-(O)DFD receivers, the application of the LR technique followed by the K-Best detector is not straightforward. The main problematic lies in the consideration of the possibly transmit symbols vector in the reduced constellation, namely $\mathbf{z}$. Unfortunately, the set of all possibly transmit symbols vectors can not be predetermined since it does not only depend on the employed constellation, but also on the $\mathbf{T}^{-1}$ matrix. Consequently, the number of children in the tree search and their values are not known in advance. A brute-force solution to determine the set of all possibly transmit vectors in the reduced constellation, $\mathbf{Z}_{all}$, is to get first the set of all possibly transmit vectors in the original constellation, $\mathbf{X}_{all}$, and then to apply the relation $\mathbf{Z}_{all} = \mathbf{T}^{-1}\mathbf{X}_{all}$ for each channel realization. Clearly, this possibility is not feasible since it corresponds to the computational complexity of the ML detector. To avoid this problem, some feasible solutions, more or less efficient, have been proposed in the literature.

6.2 Summary of the lattice reduction-aided neighbourhood study algorithms

While the first idea of combining both the LR and a neighborhood study has been proposed by Zhao *et al.* W. Zhao, and G.B. Giannakis (2006), Qi *et al.* X.-F. Qi, and K. Holt (2007) introduced in detail a novel scheme-Namely LRA-SD algorithm-where a particular attention to neighborhood exploration has been paid. This algorithm has been enhanced by Roger *et al.* S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009) by, among others, associating LR and K-Best. This offers the advantages of the K-Best concerning its complexity and parallel nature, and consequently its implementation. The hot topic of the neighborhood study size

reduction is being widely studied M. Shabany, and P.G. Gulak (2008); S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009). In a first time, let us introduce the basic idea that makes the LR theory appropriate for application in complexity - and latency - limited communication systems. Note that the normalize-shift-scale steps that have been previously introduced, will not be addressed again.

6.3 The problem of the reduced neighborhood study

Starting from Equation (32), both the sides of the lattice-reduced channel and reduced constellation can be left-multiplied by $\tilde{\mathbf{Q}}^H$, where $[\tilde{\mathbf{Q}}, \tilde{\mathbf{R}}] = QRD\{\tilde{\mathbf{H}}\}$. Therefore, a new relation is obtained:

$$\tilde{\mathbf{Q}}^H \mathbf{r}_{\mathbb{Z}} = \tilde{\mathbf{R}} \mathbf{z}_{\mathbb{Z}} + \tilde{\mathbf{n}}, \tag{34}$$

this makes any SD scheme to be introduced, and eventually a K-Best. At this moment, the critical point of neighbours generation in the reduced constellation has to be introduced. As previously presented, the set of possible values in the original constellation is affected by the matrix $\mathbf{T}^{-1}$. In particular, due to $\mathbf{T}$ properties introduced in the LR step, the scaling, rotating, and reflection operations may induce some missing (non-adjacent) or unbounded points in the reduced lattice, despite the regularity and bounds of the original constellation. In presence of noise, some candidates may not map to any legitimate constellation point in the original constellation. Therefore, it is necessary to take into account this effect by discarding vectors with one (or more) entries exceeding constellation boundaries. However, the vicinity of a lattice point in the reduced constellation would be mapped onto the same signal point. Consequently, a large number of solutions might be discarded, leading to inefficiency of any additional neighborhood study. Also note that it is not possible to prevent this aspect without exhaustive search complexity since $\mathbf{T}^{-1}$ applies on the whole $\hat{\mathbf{z}}$ vector while it is treated layer by layer.

Zhao *et al.* W. Zhao, and G.B. Giannakis (2006) propose a radius expression in the reduced lattice from the radius expression in the original constellation through the Cauchy-Schwarz inequality. This idea leads to an upper bound of the explored neighborhood and accordingly a reduction in the number of tested candidates. However, this proposition is not enough to correctly generate a neighborhood because of the classical - and previously introduced - problematic of any fixed radius.

A zig-zag strategy inside of the radius constraint works better S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009); W. Zhao, and G.B. Giannakis (2006). Qi *et al.* X.-F. Qi, and K. Holt (2007) propose a predetermined set of displacement $[\delta_1, \cdots, \delta_N]$ $(N > K)$ generating a neighborhood around the constrained DFD solution $[\mathcal{Q}_{\mathbb{Z}_\mathbb{C}}\{\tilde{\mathbf{z}}_{n_T}\} + \delta_1, \cdots, \mathcal{Q}_{\mathbb{Z}_\mathbb{C}}\{\tilde{\mathbf{z}}_{n_T}\} + \delta_N]$. The N neighbors are ordered according to their norms, by considering the current layer similarly to the SE technique, and the K candidates with the least metrics are stored. The problem of this technique lies in the number of candidates that has to be unbounded, and consequently set to a very large number of candidates N for the sake of feasibility. Roger *et al.* S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009) proposed to replace the neighborhood generation by a zig-zag strategy around the constrained DFD solution with boundaries control constraints. By denoting boundaries in the original DSS constellation $\mathbf{x}_{\mathbb{Z},\,\min}$ and $\mathbf{x}_{\mathbb{Z},\,\max}$, the reduced constellation boundaries can be obtained through the relation $\mathbf{z}_{\mathbb{Z}} = \mathbf{T}^{-1}\mathbf{x}_{\mathbb{Z}}$ that implies $\mathbf{z}_{\max,\,l} = \max\{\mathbf{T}^{-1}_{l,:}\mathbf{x}_{\mathbb{Z}}\}$ for a given layer l. The exact solution is given in S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009) for the real case and can be

extended to the complex case:

$$\mathbf{z}_{\max,\,l} = \mathbf{x}_{\mathbb{Z},\,\max} \sum_{j \in P^l} \mathbf{T}^{-1}_{l,\,j} + \mathbf{x}_{\mathbb{Z},\,\min} \sum_{j \in N^l} \mathbf{T}^{-1}_{l,\,j},\ \mathbf{z}_{\min,\,l} = \mathbf{x}_{\mathbb{Z},\,\min} \sum_{j \in P^l} \mathbf{T}^{-1}_{l,\,j} + \mathbf{x}_{\mathbb{Z},\,\max} \sum_{j \in N^l} \mathbf{T}^{-1}_{l,\,j}, \tag{35}$$

where P^l and N^l stands for the set of indices j corresponding to positive and negative entries $(l,\ j)$ of $\mathbf{T}^{-1}$, respectively. By denoting the latter algorithm as the LRA-KBest-Candidate Limitation (LRA-KBest-CL), note that this solution is exact and does not induce any performance degradation.

The main advantages in the LRA-KBest are highlighted. While it has been shown that the LRA-KBest achieves the ML performance for a reasonable K, even for 16QAM and 64QAM constellations, as depicted in Figure 17, the main favorable aspect lies in the neighborhood study size that is independent of the constellation order. So the SD complexity has been reduced though the LR-Aid and would be feasible, in particular for 16QAM and 64QAM constellations that are required in the 3GPP LTE-A norm 3GPP (2009). Also, such a detector is less sensitive to ill-conditioned channel matrices due to the LR step. However, the detector offers limited benefits with the widely used QPSK modulations, due to nearby lattice points elimination during the quantization step, and the infinite lattice problematic in the reduced domain constellation search has not been solved convincingly and is up to now an active field of search.

Let us introduce the particular case of Zhang *et al.* W. Zhang, and X. Ma (2007a;b) that proposes to combine both LR and a neighborhood study in the original constellation.

6.4 A particular case

In order to reduce the SNR offset by avoiding the problematic neighborhood study in the reduced constellation, a by-solution has been provided W. Zhang, and X. Ma (2007a) based on the unconstrained LRA-ZF result. The idea here was to provide a soft-decision LRA-ZF detector by generating a list of solutions. This way, Log-Likelihood Ratios (LLR) can be obtained through the classical max-log approximation, if both hypothesis and counter-hypothesis have been caught, or through-among others-a LLR clipping B.M. Hochwald, and S. ten Brink (2003); D.L. Milliner, E. Zimmermann, J.R. Barry and G. Fettweis (2008).

The idea introduced by Zhang *et al.* corresponds in reality to a SD-like technique, allowing to provide a neighborhood study around the unconstrained LRA-ZF solution: $\mathbf{r}_{LRA-ZF} = \tilde{\mathbf{H}}^{\dagger}\mathbf{r}$. The list of candidates, that corresponds to the neighborhood in the reduced constellation, can be defined using the following relation:

$$\mathcal{L}_z = \{\tilde{\mathbf{z}} :\ \|\tilde{\mathbf{z}} - \mathbf{r}_{LRA-ZF}\|^2 < d_z\}, \tag{36}$$

where $\tilde{\mathbf{z}}$ is a hypothetical value for $\mathbf{z}$ and $\sqrt{d_z}$ is the sphere constraint. However, a direct estimation of $\hat{\mathbf{x}}$ may be obtained by left-multiplying by correct lines of $\mathbf{T}^{-1}$ at each detected symbol:

$$\mathcal{L}_x = \{\tilde{\mathbf{x}} :\ \|\mathbf{T}^{-1}\tilde{\mathbf{x}} - \mathbf{r}_{LRA-ZF}\|^2 < d_z\}, \tag{37}$$

where $\tilde{\mathbf{x}}$ is a hypothetical value for $\mathbf{x}$ and by noting that the sphere constraint remains unchanged.

The problem introduced by such a technique is how to obtain $\tilde{\mathbf{x}}$ layer by layer, since it would lead to non-existing symbols. A possible solution is the introduction of the QRD of $\mathbf{T}^{-1}$ in

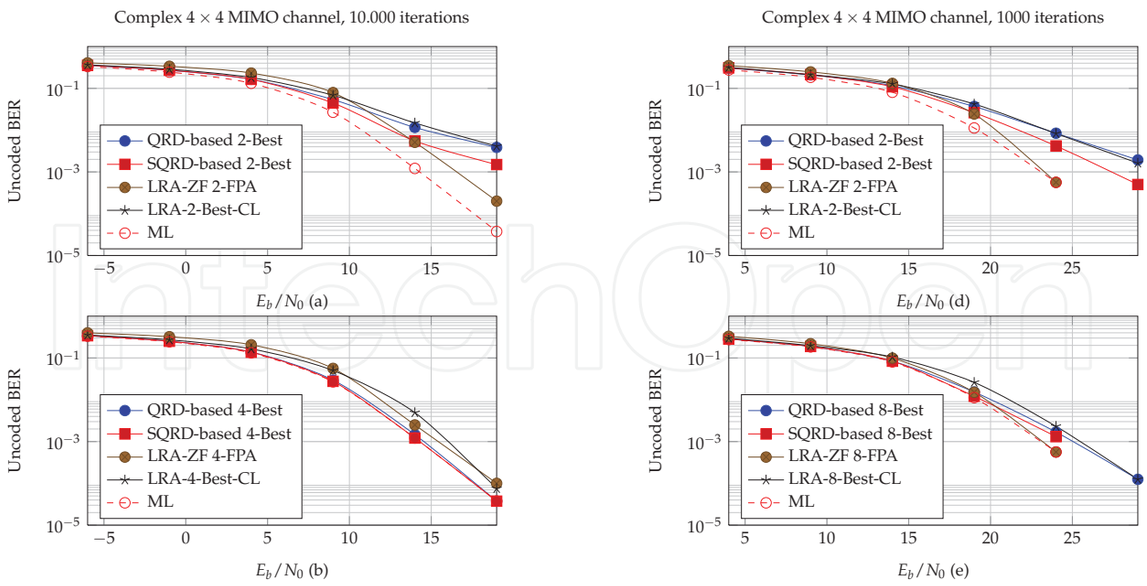


Fig. 17. Uncoded BER as a function of E_b/N_0, Complex Rayleigh 4×4 MIMO channel, QRD-based 2/4(8)-Best, SQRD-based 2/4(8)-Best, LRA-ZF 2/4(8)-FPA, LRA-2/4(8)-Best-CL and ML detectors, QPSK modulations at each layer (a-c) and 16QAM modulations at each layer (d-f).

order to make the current detected symbol within the symbols vector independent of the remaining to-detect symbols. This idea leads to the following expression:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\Omega_{\mathbb{C}}^{n_T}}{\operatorname{argmin}} \left\| \mathbf{Q}_{\mathbf{T}^{-1}}^{H} \mathbf{r}_{LRA-ZF} - \mathbf{R}_{\mathbf{T}^{-1}} \mathbf{x} \right\|^2 < d_z, \tag{38}$$

where $[\mathbf{Q}_{\mathbf{T}^{-1}}, \mathbf{R}_{\mathbf{T}^{-1}}] = QRD\{\mathbf{T}^{-1}\}$. Due to the upper triangular form of $\mathbf{R}_{\mathbf{T}^{-1}}$, $\hat{\mathbf{x}}$ can be detected layer by layer through the K-Best scheme such as the radius constraint can be eluded. Consequently, the problematic aspects of the reduced domain constellation study are avoided, and the neighborhood study is provided at the *cheap* price of an additional QRD. By denoting the latter algorithm as the LRA-ZF Fixed Point Algorithm (LRA-ZF-FPA), note that the problem of this technique lies in the Euclidean Distance expression which is not equivalent to the ML equation. The technique only aims at generating a neighborhood study for the Soft-Decision extension. There will be no significant additional performance improvement for larger K, as depicted in Figure 17.

6.5 Simulation results

In Figure 17 and in the case of a neighborhood study in the reduced domain, near-ML performance is reached for small K values, in both QPSK and 16QAM cases.

It is obvious to the reader that K is independent of the constellation order, which can be demonstrated. This aspect is essential for the OFDM extension since any SD-like detector has to be fully processed for each to-be-estimated symbols vector. Also, the solution offered by LRA-ZF-FPA is interesting in the sense that it allows to make profit of the LRA benefit with an additional neighborhood study in the original constellation. However, it does not reach the ML performance because of the non-equivalence of the metrics computation even in the case of a large K.

7. Conclusion

In this chapter, we have presented an up-to-date review, as well as several prominent contributions, of the detection problematic in MIMO-SM systems. It has been shown that, theoretically, such schemes linearly increase the channel capacity. However, in practice, achieving such increase in the system capacity depends, among other factors, on the employed receiver design and particularly on the de-multiplexing algorithms, a.k.a. detection techniques. In the literature, several detection techniques that differ in their employed strategies have been proposed. This chapter has been devoted to analyze the structures of those algorithms. In addition to the achieved performance, we pay a great attention in our analysis to the computational complexities since these algorithms are candidates for implementation in both latency and power-limited communication systems. The linear detectors have been introduced and their low performances have been outlined despite of their attracting low computational complexities. DFD techniques improve the performance compared to the linear detectors. However, they might require remarkably higher computations, while still being far from achieving the optimal performance, even with ordering. Tree-search detection techniques, including SD, QRD-M, and FSD, achieve the optimum performance. However, FSD and QRD-M are more favorable due to their fixed and realizable computational complexities. An attractive pre-detection process, referred to as lattice-basis reduction, can be considered in order to apply any detector through a close-to-orthogonal channel matrix. As a result, a low complexity detection technique, such as linear detectors, can achieve the optimum diversity order. In this chapter, we followed the lattice reduction technique with the K-best algorithm with low K values, where the optimum performance is achieved. In conclusion, in this chapter, we surveyed the up-to-date advancements in the signal detection field, and we set the criteria over which detection algorithms can be evaluated. Moreover, we set a clear path for future research via introducing several recently proposed detection methodologies that require further studies to be ready for real-time applications.

8. References

3GPP (2009). Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception v8.8.0, *TS 36.101*, 3rd Generation Partnership Project (3GPP).

A. Paulraj, R. Nabar, and D. Gore (2003). Introduction to Space-Time Wireless Communications, *UK, Cambridge: Cambridge University Press* .

A. Telatar (1999). Capacity of multi-antenna Gaussian channels, *Telecommunications, European Transactions on* 10(6): 585–595.

A.K. Lenstra, H.W. Lenstra, and L. Lovász (1982). Factoring polynomials with rational coefficients, *Mathematische Annalen* 261(4): 515–534.

B. Gestner, W. Zhang, X. Ma, and D.V. Anderson (2008). VLSI Implementation of a Lattice Reduction Algorithm for Low-Complexity Equalization, *Circuits and Systems for Communications, IEEE International Conference on* pp. 643–647.

B. Hassibi (2000). An efficient square-root algorithm for BLAST, *Acoustics, Speech, and Signal Processing, IEEE International Conference on* pp. 737–740.

B. Hassibi, and H. Vikalo (2001). On the expected complexity of sphere decoding, *Signals, Systems and Computers, Asilomar Conference on* pp. 1051–1055.

B. Schubert (2006). Analysis of sphere decoding in linear cooperative wireless relay networks., *Master Thesis, Berlin University of Technology* .

B.A. Lamacchia (1991). Basis Reduction Algorithms and Subset Sum Problems, *Technical report*, MSc Thesis, Massachusetts Institute of Technology.

B.M. Hochwald, and S. ten Brink (2003). Achieving Near-Capacity on a Multiple-Antenna Channel, *Communications, IEEE Transactions on* 51(3): 389–399.

C. Ling (2006). Approximate lattice decoding: Primal versus dual lattice reduction, *Information Theory, International Symposium on* .

C. Schnorr, and M. Euchner (1994). Lattice basis reduction: improved practical algorithms and solving subset sum problems, *Mathematical Programming* 66: 181–199.

C. Windpassenger (2004). Detection and precoding for multiple input multiple output channels., *PhD Dissertation, Erlangen-Nurnberg University* .

C. Windpassinger, and R.F.H. Fischer (2003). Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction, *Information Theory Workshop, IEEE* pp. 345–348.

D. Seethaler, G. Matz, and F. Hlawatsch (2007). Low-complexity MIMO data detection using Seysenï£¡s lattice reduction algorithm, *Acoustics, Speech, and Signal Processing, IEEE International Conference on* 3: 53–56.

D. Shiu, and J. Kahn (1999). Layered space-time codes for wireless communications using multiple transmit antennas, *Communications, International Conference on* pp. 436–440.

D. Wübben, and D. Seethaler (2007). On the Performance of Lattice Reduction Schemes for MIMO Data Detection, *Signals, Systems and Computers, Asilomar Conference on* pp. 1534–1538.

D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K.-D. Kammeyer (2001). Efficient algorithm for decoding layered space-time codes, *Electronics Letters, IEEE* 37(22): 1348–1350.

D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2003). MMSE extension of V-BLAST based on sorted QR decomposition, *Vehicular Technology Conference, IEEE* pp. 508–512.

D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer (2004). MMSE-based lattice-reduction for near-ML detection of MIMO systems, *Smart Antennas, ITG Workshop on* pp. 106–113.

D.L. Milliner, E. Zimmermann, J.R. Barry and G. Fettweis (2008). Channel State Information Based LLR Clipping in List MIMO Detection, *Personal, Indoor and Mobile Radio Communications International, IEEE International Symposium on* pp. 1–5.

E. Agrell , T. Eriksson, A. Vardy, and K. Zeger (2002). Closest point search in lattices, *Information Theory, IEEE Transactions on* 48(8): 2201–2214.

H. Kawai, K. Higuchi, N. Maeda, and M. Sawahashi (2006). Adaptive control of surviving symbol replica candidates in QRM-MLD for OFDM-MIMO multiplexing, *IEEE Journal of Selected Areas in Communications* 24(6): 1130–1140.

H. Yao, and G.W. Wornell (2002). Lattice-reduction-aided detectors for MIMO communication systems, *Global Telecommunications Conference, IEEE* 1: 424–428.

H. Zhu, Z. Lei, and F. Chin (2004). An improved square-root algorithm for BLAST, *Signal Processing Letters, IEEE* 11(9): 772–775.

J. Anderson, and S. Mohan (1984). Sequential coding algorithms: a survey and cost analysis, *Communications, IEEE Transactions on* 32(2): 169–176.

J. Benesty, Y. Huang, and J. Chen (2003). A fast recursive algorithm for optimum sequential signal detection in a BLAST system, *Signal Processing, IEEE Transactions on* 51(7): 1722–1730.

J. Boutros, N. Gresset, L. Brunel, and M. Fossorier (2003). Soft-input soft-output lattice sphere decoder for linear channels, *Global Telecommunications Conference, IEEE* pp. 1583–1587.

J. Jaldén, and B. Ottersten (2005). On the complexity of sphere decoding in digital communications, *Signal Processing, IEEE Transactions on* 53(4): 1474–1484.

J. Jaldén, D. Seethaler, and G. Matz (2008). Worst-and average-case complexity of LLL lattice reduction in MIMO wireless systems, *Acoustics, Speech and Signal Processing, IEEE International Conference on* pp. 2685–2688.

J. Wang, and B. Daneshrad (2005). A comparative study of MIMO detection algorithms for wideband spatial multiplexing systems, *Wireless Communications and Networking, IEEE International Conference on* 1: 408–413.

K. Jeon, H. Kim, and H. Park (2006). An efficient QRD-M algorithm using partial decision feedback detection, *Signal, Systems, and Computers, Asilomar Conference on* pp. 1658–1661.

K.-W. Wong, C.-Y. Tsui, S.-K. Cheng, and W.-H. Mow (2002). A VLSI Architecture of a K-Best Lattice Decoding Algorithm For MIMO Channels, 3: 272–276.

K.J. Kim, J. Yue, R.A. Iltis, and J.D. Gibson (2005). A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems, *Wireless Communications, IEEE Transactions on* 4(2): 710–721.

L. Barbero, and J. Thompson (2008a). Extending a fixed-complexity sphere decoder to obtain likelihood information for turbo-MIMO systems, *Vehicular Technology, IEEE Transactions on* 57(5): 2804–2814.

L. Barbero, and J. Thompson (2008b). Fixing the complexity of the sphere decoder for MIMO detection, *Wireless Communications, IEEE Transactions on* 7(6): 2131–2142.

L. Lovász (1986). An Algorithmic Theory of Numbers, Graphs and Convexity, *PA: Society for Industrial and Applied Mathematics* .

L. Wang, L. Xu, S. Chen, and L. Hanzo (2008). MMSE Soft-Interference-Cancellation Aided Iterative Center-Shifting K-Best Sphere Detection for MIMO Channels, *Communications, IEEE International Conference on* pp. 3819–3823.

L.G. Barbero, T. Ratnarajah, and C. Cowan (2008). A comparison of complex lattice reduction algorithms for MIMO detection, *Acoustics, Speech and Signal Processing, IEEE International Conference on* pp. 2705–2708.

M. Mohaisen, and K.H. Chang (2009a). On improving the efficiency of the fixed-complexity sphere decoder, *Vehicular Technology Conference, IEEE* pp. 1–5.

M. Mohaisen, and K.H. Chang (2009b). On the achievable improvement by the linear minimum mean square error detector, *Communication and Information Technology, International Symposium on* pp. 770–774.

M. Mohaisen, and K.H. Chang (2010). Upper-lower bounded-complexity QRD-M for spatial multiplexing MIMO-OFDM systems, *Wireless Personal Communications* .

M. Mohaisen, H.S. An, and K.H.Chang (2009). Detection techniques for MIMO multiplexing: a comparative review, *Internet and Information Systems, KSII Transactions on* .

M. Mohaisen, K.H. Chang, and B.T. Koo (2009). Adaptive parallel and iterative QRDM algorithms for spatial multiplexing MIMO systems , *Vehicular Technology Conference* pp. 1–5.

M. Seysen (1993). Simultaneous reduction of a lattice basis and its reciprocal basis, *Combinatorica* 13(3): 363–376.

M. Shabany, and P.G. Gulak (2008). The Application of Lattice-Reduction to the K-Best Algorithm for Near-Optimal MIMO Detection, *Circuits and Systems, IEEE International Symposium on* pp. 316–319.

M. Taherzadeh, A. Mobasher, and A.K. Khandani (2005). LLL lattice-basis reduction achieves the maximum diversity in MIMO systems, *Information Theory, International Symposium on* .

P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela (1998). V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel, *Signals, Systems, and Electronics, URSI International Symposium on* pp. 295–300.

S. Aubert, M. Mohaisen, F. Nouvel, and K.H. Chang (2010). Parallel QR decomposition in LTE-A systems, *IEEE International Workshop on Signal Processing Advances in Wireless Communications* .

S. Haykin, and M. Moher (2005). Modern Wireless Communications, *USA, NJ, Pearson Prentice Hall* .

S. Roger, A. Gonzalez, V. Almenar, and A.M. Vidal (2009). On Decreasing the Complexity of Lattice-Reduction-Aided K-Best MIMO Detectors, *European Signal Processing Conference* pp. 2411–2415.

T. Cui, and C. Tellambura (2005). An efficient generalized sphere decoder for rank-deficient MIMO systems, *Communications Letters, IEEE* 9(5): 423–425.

U. Fincke, and M. Pohst (1985). Improved methods for calculating vectors of short length in a lattice, including complexity analysis, *Mathematics of Computation* 44(170): 463–471.

W. Van Etten (1976). Maximum likelihood receiver for multiple channel transmission systems, *Communications, IEEE Transactions on* pp. 276–283.

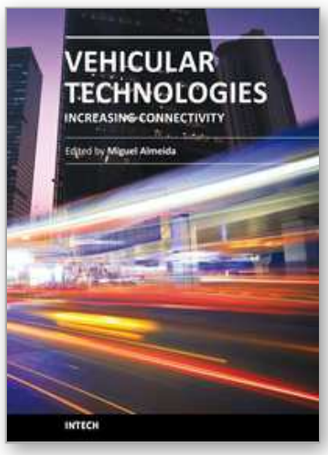
W. Zhang, and X. Ma (2007a). Approaching Optimal Performance By Lattice-Reduction Aided Soft Detectors, *Information Sciences and Systems, Conference on* pp. 818–822.

W. Zhang, and X. Ma (2007b). Designing Soft Detectors Based On Seysen's Algorithm, *Military Communications Conference, IEEE* 14(5): 1–7.

W. Zhao, and G.B. Giannakis (2006). Reduced Complexity Closest Point Decoding Algorithms for Random Lattices, *Wireless Communications, IEEE Transactions on* 5(1): 101–111.

X.-F. Qi, and K. Holt (2007). A Lattice-Reduction-Aided Soft Demapper for High-Rate Coded MIMO-OFDM Systems, *Signal Processing Letters, IEEE* 14(5): 305 –308.

Y.H. Gan, C. Ling, and W.H. Mow (2009). Complex Lattice Reduction Algorithm for Low-Complexity Full-Diversity MIMO Detection, *Signal Processing, IEEE Transactions on* 57(7): 2701–2710.



Vehicular Technologies: Increasing Connectivity
Edited by Dr Miguel Almeida

ISBN 978-953-307-223-4
Hard cover, 448 pages
Publisher InTech
Published online 11, April, 2011
Published in print edition April, 2011

This book provides an insight on both the challenges and the technological solutions of several approaches, which allow connecting vehicles between each other and with the network. It underlines the trends on networking capabilities and their issues, further focusing on the MAC and Physical layer challenges. Ranging from the advances on radio access technologies to intelligent mechanisms deployed to enhance cooperative communications, cognitive radio and multiple antenna systems have been given particular highlight.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sébastien Aubert and Manar Mohaisen (2011). From Linear Equalization to Lattice-Reduction-Aided Sphere-Detector as an Answer to the MIMO Detection Problematic in Spatial Multiplexing Systems, Vehicular Technologies: Increasing Connectivity, Dr Miguel Almeida (Ed.), ISBN: 978-953-307-223-4, InTech, Available from: http://www.intechopen.com/books/vehicular-technologies-increasing-connectivity/from-linear-equalization-to-lattice-reduction-aided-sphere-detector-as-an-answer-to-the-mimo-detecti

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen