RESEARCH　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

A simplified hard output sphere decoder for large MIMO systems with the use of efficient search center and reduced domain neighborhood study

Youssef Nasser[1*], Sebastien Aubert[2,3], Fabienne Nouvel[3], Karim Y. Kabalan[1] and Hassan A. Artail[1]

Abstract

Multiple-input multiple-output (MIMO) with a spatial-multiplexing (SM) scheme is a topic of high interest for the next generation of wireless communications systems. At the receiver, neighborhood studies (NS) and lattice reduction (LR)-aided techniques are common solutions in the literature to approach the optimal and computationally complex maximum likelihood (ML) detection. However, the NS and LR solutions might not offer optimal performance for large dimensional systems, such as large number of antennas, and high-order constellations when they are considered separately. In this paper, we propose a novel equivalent metric dealing with the association of these solutions by introducing a reduced domain neighborhood study. We show that the proposed metric presents a relevant complexity reduction while maintaining near-ML performance. Moreover, the corresponding computational complexity is shown to be independent of the constellation size, but it is quadratic in the number of transmit antennas. For instance, for a 4×4 MIMO system with 16-QAM modulation on each layer, the proposed solution is simultaneously near-ML with perfect and real channel estimation and ten times less complex than the classical neighborhood-based K-best solution.

1 Introduction

Multiple-input multiple-output (MIMO) technology has taken a lot of attention in the last decade since it can improve link reliability without sacrificing bandwidth efficiency or, contrariwise, it can improve the bandwidth efficiency without losing link reliability [1]. Recently, the concept of large MIMO systems, i.e., high number of antennas, has also gained research interests, and it is well seen as a part of next-generation wireless communication systems [2, 3].

However, the main drawback of MIMO technology is the increased complexity at the receiver side when a non-orthogonal (NO) MIMO scheme with a large number of antennas and/or large constellation size is implemented [4, 5]. For the detection process, although the performance of the maximum likelihood (ML) detector is optimal, its computational complexity increases exponentially with the number of transmit antennas and with the constellation size. In literature, different MIMO detection techniques have been proposed. The linear-like detection

(LD) [6] and decision-feedback detection (DFD) [7] are the baseline detection algorithms. Here, we distinguish the conventional linear MIMO detection techniques zero forcing (ZF) [8] and minimum mean square error (MMSE) [8]. Although linear detection approaches are attractive in terms of their computational complexity, they might lead to a non-negligible degradation in terms of performance [9].

Some non-linear detectors have been also introduced. The sphere decoder (SD), one of the most famous MIMO detectors, is based on a tree search and is very popular due to its quasi-optimal performance [10]. However, this performance is reached at the detriment of an additional implementation complexity. Indeed, the SD achieves quasi-ML performance while its average complexity is shown to be polynomial (roughly cubic) in constellation size and in the number of transmit antennas over a certain range of signal-to-noise ratio (SNR) while the worst case complexity is still exponential [11]. From a hardware implementation point of view, the SD algorithm presents two main drawbacks. Firstly, its complexity coefficients can become large when the problem dimension is high, i.e., at the high spectral efficiency, high number of antennas, and high number of

* Correspondence: youssef.nasser@aub.edu.lb
[1]ECE Department, American University of Beirut, Bliss Street, Beirut, Lebanon
Full list of author information is available at the end of the article

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 2 of 17

users in multi-user MIMO (MU-MIMO) context. Secondly, the variance of its computation time can be also large leading to undesirable highly variable decoding delays. Despite classical optimizations such as the Schnorr-Euchner (SE) enumeration [12], the SD originally presented in [11] offers by nature a sequential tree search phase, which is an additional drawback for implementation. In order to deal with these two aspects, the authors in [13] have proposed a sub-optimal solution denoted as the K-best [13, 14], where K is the number of stored neighbors given a layer. However, even with a fixed computational complexity and a parallel nature of implementation, some optimizations are required especially for high-order constellation and large number of antennas (due to the large K required in this case) [15–18]. Aiming at reducing the neighborhood size (namely K, over all layers), different solutions are proposed. For instance, the sorted QR decomposition (SQRD)-based dynamic K-best which leads to the famous SQRD-based fixed throughput SD (FTSD) is proposed in [16]. Even with these efforts, the neighborhood size still induces a computationally expensive solution for achieving quasi-ML performance. An alternative trend has been firstly presented in the literature by Wuebben et al. in [19]. It consists in adding a pre-processing step, namely the lattice reduction (LR), aiming at applying a classical detection through a better-conditioned channel matrix [19–21]. This solution has been shown to offer the full reception diversity at the expense of a SNR offset in the system performance. This offset increases with large dimensional transmit antenna systems and high-order modulations.

Recently, a promising—although complex—association of the K-best and LR solutions has been considered. It provides a convenient performance-complexity tradeoff. The general idea consists in reducing the SNR offset through a neighborhood study which yields a near-ML performance for a reasonable K. The concept has been introduced first by the authors of [22]. Later on, their basic solution has been improved by proposing to model the sphere constraint in a reduced domain or by introducing efficient symbols' enumeration algorithm [23]. However, a major aspect of this combination has not been considered yet. In particular, any SD, including the K-best, may be advantageously applied by considering a better-conditioned channel matrix through the introduction of a Reduced Domain Neighborhood (RDN) study and a judicious search center. In [5], an improved LR technique dealing with the RDN has been proposed in the context of large MIMO systems. It is based on the decomposition of the spanned space of the channel matrix into small subspaces in order to improve orthogonality of the quantization. In [24], the search center is found through an ant colony optimization and initial search through the output of the MMSE detector.

In this paper, we adopt the K-best solution with fixed complexity as the basic solution of the SD. We propose to reduce the neighborhood size through an efficient pre-processing step which allows the SD process to apply a neighborhood study in a modified constellation domain. Then, using the modified domain, we propose a modified novel ML equation with an efficient search center. We show that the proposed metric presents a large complexity reduction while maintaining near-ML performance. Moreover, the corresponding complexity is shown to be independent of the constellation size and polynomial in the number of transmit antennas. In particular, for a 4×4 MIMO system with 16-QAM modulation on each layer, the proposed association presents near-ML performance while it is ten times less complex than the classical K-best solution. We note that because the complexity is fixed with such a detector, the exposed optimizations guarantee a performance gain for a given neighborhood size or a reduction of the neighborhood size for a given bit error rate (BER) target. The contributions of this paper are summarized as follows:[1]

- A promising association of the K-best and LR solutions is proposed.
- Modification of the SD neighborhood study by applying a pre-processing step. This is accompanied with a new and efficient search center and MMSE detector. The equivalent expression of the lattice reduction-aided (LRA)-MMSE-centered SD, which corresponds to an efficient LRA-MMSE-successive interference canceller (SIC) Babai point, is proposed to improve the performance or reduce the complexity of the detector.
- The (S)QRD is introduced in formulas. It provides—to the best of the authors' knowledge—the best known pseudo-linear hard detector as a Babai point, for large number of antennas as well as for high-order modulations.
- The proposed expression is robust by nature to any search center and constellation order and offers close-to-optimal performance with medium K values. This applies for both perfect and real channel estimation.
- The proposed solution offers a computational complexity that is independent of the constellation order. Therefore, it outperforms classical SD techniques for a reasonable complexity in the case of high-order constellations. We show for example that a number of neighbors $K = 2$ is sufficient for a 4×4 MIMO system with 16-QAM modulation on each layer, and it is less than 0.5 dB for a 64×64 and 128×128 MIMO system from the ML solution.
- The proposed solution offers a computational complexity that is quasi-constant for large number of antennas, showing the evidence of its importance.

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 3 of 17

This paper is organized as follows. Section 2 presents the problem statement of the SD. In Section 3, the different existing solutions are described and analyzed. In Section 4, we propose our generalized solution based on LR with the use of an efficient search center and reduced domain neighborhood. In Section 5, the performance of the presented detectors are provided, compared, and discussed. In Section 6, we consider the computational complexity of the proposed solution in comparison with some reference detection techniques. Conclusions are drawn in Section 7.

2 Problem statement
2.1 Sphere decoder detector
Let us introduce a $n_R \times n_T$ MIMO system model with n_T transmit and n_R receive antennas. Then, the received symbols vector could be written as

$$y = Hx + n, \tag{1}$$

where H represents the (n_R, n_T) complex channel matrix assumed to be perfectly known at the receiver, x is the transmit symbol vector of dimensions $(n_T, 1)$ where each entry is independently withdrawn from a constellation set ξ, and n is the additive white Gaussian noise of dimensions $(n_R, 1)$ and of variance $\sigma^2/2$ per dimension. The basic idea of the SD, to reach the optimal estimate $\hat{x}_{ML}$ (given by the ML detector) while avoiding an exhaustive search, is to observe the lattice points that lie inside a sphere of radius d.

The SD solution starts from the ML equation $\hat{x}_{ML} = \text{argmin}_{x \in \xi^{n_T}} \|y - Hx\|^2$ and reads

$$\hat{x}_{SD} = \underset{x \in \xi^{n_T}}{\text{argmin}} \|Q^H y - Rx\|^2 \leq d^2 \tag{2}$$

where $H = QR$, with the classical QRD definitions.

The classical SD formula in (2) is centered on the received signal y. From now on, this detection will be denoted as the naïve SD. In the case of a depth-first search algorithm [13], the first solution given by this algorithm is defined as the Babai point [25, 26]. In order to write it, the classical SD expression may be re-arranged, leading then to an exact formulation through an efficient partial Euclidean distance (PED) expression and early pruned nodes [27].

In the literature, the SD principle leads to numerous implementation problems. In particular, it is a non-deterministic polynomial-time (NP-hard) problem [28]. This aspect has been partially solved through the introduction of an efficient solution that lies in a fixed neighborhood size algorithm (FNSA), commonly known as the *K*-best solution. However, this solution makes the detector sub-optimal since it leads to a performance loss compared to the ML detector. It is particularly true in the case of an inappropriate choice of K according to the MIMO channel condition number and in the case of an inappropriate choice of d in (2). Indeed, an inappropriate choice of d could lead to a ML solution excluded from the search tree. On the other hand, although a neighborhood study remains the one and only one solution that achieves near-ML performance, it may lead to the use of a large-size neighborhood scan which would correspond to a dramatic increase of the computational complexity. This complexity's increase will be prohibitive for high-order modulations.

2.2 Lattice reduction
Through the aforementioned considerations and by using the lattice definition in [26], the system model given in (1) rewrites

$$y = \tilde{H}z + n, \tag{3}$$

where $\tilde{H} = HT$ and $z = T^{-1}x$. The $n_T \times n_T$ complex matrix T (with $|\det\{T\}| = 1$) is unimodular, i.e., its entries belong to the set of complex integers which reads $\mathbb{Z}_\mathbb{C} = \mathbb{Z} + j\mathbb{Z}$, with $j^2 = 1$. The key idea of any LR-aided (LRA) detection scheme is to understand that the finite set of transmitted symbols ξ^{n_T} can be interpreted as a de-normalized, shifted then scaled version of the infinite set of complex integers subset $\subset \mathbb{Z}_C^{n_T}$ according to the relations offered in [29].

To this end, various reduction algorithms have been proposed [19, 30–32]. In the following, we focus on the well-known Lenstra-Lenstra-Lovász (LLL) algorithm due to considerations presented in [30, 33]. The lattice aided (LA) is a local approach [34] that transforms the channel matrix into an LLL-reduced basis that satisfies both of the orthogonality and norm reduction conditions [31]. While it has been shown in [33] that the QRD outputs of the channel matrix is a possible starting point for the LLL, it has been subsequently introduced that the SQRD provides a better starting point [34]. In particular, it leads to a significant reduction of its computational complexity [35]. That is, the detection process in (3) is performed on z instead of x through the better-conditioned matrix $\tilde{H}$. Wuebben et al. [19] proposed a full description of some reference solutions, namely the LRA-ZF and LRA-ZF-SIC without noise power consideration and the LRA-MMSE, LRA-MMSE extended, and LRA-MMSE-SIC. LRA detectors constitute efficient detectors in the sense of the high quality of their hard outputs. Indeed, they offer a low overall computational complexity while the ML diversity is reached within a constant offset. However, some important drawbacks exist. In particular, the aforementioned SNR offset is important in the case of high-order modulations and of large number of antennas.

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 4 of 17

This issue is expected to be bypassed through an additional neighborhood study.

2.3 Lattice reduction-aided sphere decoder

Contrary to the LRA-(O)SIC receivers, the application of the LR preprocessing step followed by any SD detector is not straightforward. The main problem lies in the consideration of the possibly transmit symbol vector in the reduced constellation since, unfortunately, the set of all possible transmit symbols vectors cannot be predetermined. The reason for that is because the solution does not only depend on the employed constellation but also on the T^{-1} matrix of (3). Hence, the number of children in the search tree and their values are not known in advance. A brute-force solution is then to determine the set of all possible transmit vectors in the reduced constellation, starting from the set of all possible transmit vectors in the original constellation and by switching to the reduced domain, thanks to the T^{-1} matrix.

3 Detection process in the original domain neighborhood

3.1 Zero forcing-centered sphere decoder with original domain neighborhood study detection process

In order to deal with the detection process, we firstly introduce the sphere center x_C search algorithm. It concerns any signal of the form $\|x_C - x\|^2 \leq d^2$ where x is a possible signal.

Based on this search algorithm, different possible sphere centers could be introduced. Using a ZF detector, the received symbols given in (2) are then estimated through

$$\hat{x}_{\text{ZF-SIC}} = \operatorname*{argmin}_{x \in \xi^{n_T}} \|Re_{\text{ZF}}\|^2 \tag{4}$$

where $e_{\text{ZF}} = x_{\text{ZF}} - x$ and $x_{\text{ZF}} = (H^H H)^{-1} H^H y$.

Equation (4) clearly shows that the naïve SD is unconstrained ZF-centered. It implicitly corresponds to a ZF-SIC solution with an Original Domain Neighborhood (ODN) study at each layer where each layer is defined as the number of spatial multiplexed data streams. It can be noticed that, in the case of a large ODN study, the ML performance is achieved since the computed metrics are exactly the ML metrics. However, this occurs at the detriment of a large neighborhood study and subsequently a large computational complexity.

3.2 Minimum mean square error-centered sphere decoder with original domain neighborhood study detection process: equivalent formula

In this section, we introduce the minimum mean square error successive interference cancellation (MMSE-SIC), a closer-to-ML Babai point than the ZF-SIC. For the sake of

clearness with definitions, we firstly give a general definition of the equivalence between two ML metrics.

Definition Two ML equations are equivalent if the lattice point argument outputs of the minimum distance are the same, even in the case of different metrics. Two ML equations are equivalent iff:

$$\operatorname*{argmin}_{x \in \xi^{n_T}} \left\{ \|y - Hx\|^2 \right\} = \operatorname*{argmin}_{x \in \xi^{n_T}} \left\{ \|y - Hx\|^2 + c \right\} \tag{5}$$

where c is a constant.

Using (5), Cui et al. [36] proposed a general equivalent minimization problem given by

$$\hat{x}_{\text{ML}} = \operatorname*{argmin}_{x \in \xi^{n_T}} \left\{ \|y - Hx\|^2 + \alpha x^H x \right\} \tag{6}$$

where the signals x have to be of constant modulus, i.e., $x^H x$ is a constant.

This assumption is respected in the case of quadrature phase-shift keying (QPSK) modulations, but it is not directly applicable to 16-QAM and 64-QAM modulations. However, this assumption is not limiting in practice since a QAM constellation can be considered as a linear sum of QPSK points [36]. In Appendix 1, we discuss the constant modulus constraint on the signal x.

The authors of [37] proposed to apply this solution to the FNSA detection technique of the unconstrained MMSE center, leading to a MMSE-SIC procedure with an ODN study at each layer [37]. In this case, the equivalent ML equation reads

$$
\begin{aligned}
\hat{x}_{\text{MMSE-SIC}} = \operatorname*{argmin}_{x \in \xi^{n_T}} & (x_{\text{MMSE}} - x)^H (H^H H + \sigma^2 I) \\
& \times (x_{\text{MMSE}} - x)
\end{aligned} \tag{7}
$$

Through the use of the Cholesky factorization (CF) of $H^H H + \sigma^2 I = U^H U$ in the MMSE case ($H^H H = U^H U$ in the ZF case), the ML expression equivalently rewrites, using the proof in Appendix 2, as

$$\hat{x}_{\text{SIC}} = \operatorname*{argmin}_{x \in \xi^{n_T}} \left\{ \|U(\tilde{x} - x)\|^2 \right\} \tag{8}$$

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

4 Proposed detection process in the reduced domain neighborhood

Due to the implementation drawbacks, the optimal SD has been proposed to be replaced by a sub-optimal FNSA. Hassibi et al. have discussed and shown in [11] that the detector performance is impacted by the noise power and the channel condition number. Hence, the presence of a well-conditioned channel could highly reduce the neighborhood. This means that realizing a
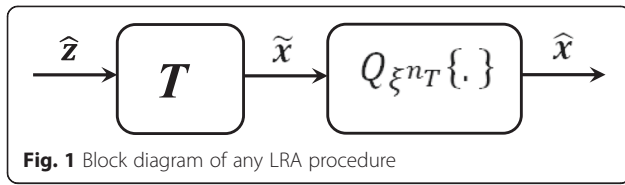
Fig. 1 Block diagram of any LRA procedure

LR step followed by a neighborhood study is a very interesting solution in a good-conditioned channel matrix. Accordingly, our proposed combined solution will be detailed in the next subsections.

4.1 Preprocessing

All existing solutions rely on the utilization of the efficient CF pre-processing step. However, these solutions are only functional in the case of a factorized formulation form. Although it is the case in our context, most of the advanced studies have been provided with the applicable QRD. In particular, the advantageous SIC performance optimizations such as ordering according to the corresponding decreasing SNR (from n_T to 1) in the ZF-SQRD case and SINR in the MMSE-SQRD case have been proposed in [33]. Moreover, a complexity reduction of the LLL-based LR algorithm has been proposed by the same authors in [33]. In our work, we propose to modify the classical detectors by introducing the QRD instead of the CF, and subsequently of the SQRD, in the (LRA-)MMSE-(O)SIC cases.

The MMSE criterion is introduced through the consideration of an extended system model [27], by introducing the $(n_R + n_T)$ - by - n_T matrix $\boldsymbol{H}_{\text{ext}}$ and the $(n_R + n_T)$ vector $\boldsymbol{y}_{\text{ext}}$ such as

$$\boldsymbol{H}_{\text{ext}} = \begin{bmatrix} \boldsymbol{H} \\ \sigma \boldsymbol{I} \end{bmatrix} \text{ and } \boldsymbol{y}_{\text{ext}} = \begin{bmatrix} \boldsymbol{y} \\ 0 \end{bmatrix}. \tag{9}$$

In this way, the pre-processing step is similar to the ZF-SQRD and the detection procedure equals that of LRA-ZF-SIC. The SQRD interest lies in the ordering of the detection symbols as a function of their S(I)NR, and consequently, it limits the error propagation in SIC procedures. Indeed, it has been shown by Wübben et al.

[19] that the optimum order offers a performance improvement even if the ML diversity is not reached. On the other hand, it was shown that once the ML diversity is achieved through a LRA technique, the performance may be significantly improved with this solution [19]. Thus, The LRA-MMSE-OSIC corresponds, to the best of the authors' knowledge, to the best pseudo-linear detector proposed in the literature, in particular in the case of 4×4 MIMO systems with QPSK modulations on each layer [19]. For higher order constellations or larger number of antennas, it may be shown that our proposed solution offers convenient hard-decision performance with a highly reduced complexity. In order to deal with these statements, we introduce the reduced domain neighborhood by using the following notations:

- $Q_{\xi^{n_T}}\{.\}$ is the quantification operator in the original domain constellation,
- $Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{.\}$ is the quantification operator in the reduced domain constellation,
- a is the power normalization and scaling coefficient (i.e., $2/\sqrt{2}$, $2/\sqrt{10}$, and $2/\sqrt{42}$ for QPSK, 16-QAM, and 64-QAM constellations, respectively)
- $\boldsymbol{d} = \frac{1}{2}\boldsymbol{T}^{-1}[1+j \quad \dots \quad 1+j]^T$ is a complex displacement vector.

The classical LRA-FNSA is implicitly unconstrained LRA-ZF-centered, which leads to a LRA-ZF-SIC procedure with a RDN study at each layer. The exact formula has not been clearly provided but is implicitly used by any LRA-FNSA [21] and may even be considered as an incremental extension of (4):

$$\hat{\boldsymbol{z}}_{\text{LRA-ZF-SIC}} = \underset{\boldsymbol{z}\in\mathbb{Z}_{\mathbb{C}}^{n_T}}{\arg\min} \|\tilde{\boldsymbol{R}}\boldsymbol{e}_{\text{LRA-ZF}}\|^2 \tag{10}$$

where $\tilde{\boldsymbol{R}}$ is the LLL-based LR algorithm output, $\boldsymbol{e}_{\text{LRA-ZF}} = \boldsymbol{z}_{\text{LRA-ZF}} - \boldsymbol{z}$, and $\mathbb{Z}_{\mathbb{C}}^{n_T}$ is the n_T-dimensional infinite set of complex integers.

Fig. 2 Block diagram of any RDN LRA-SIC FNSA procedure

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227
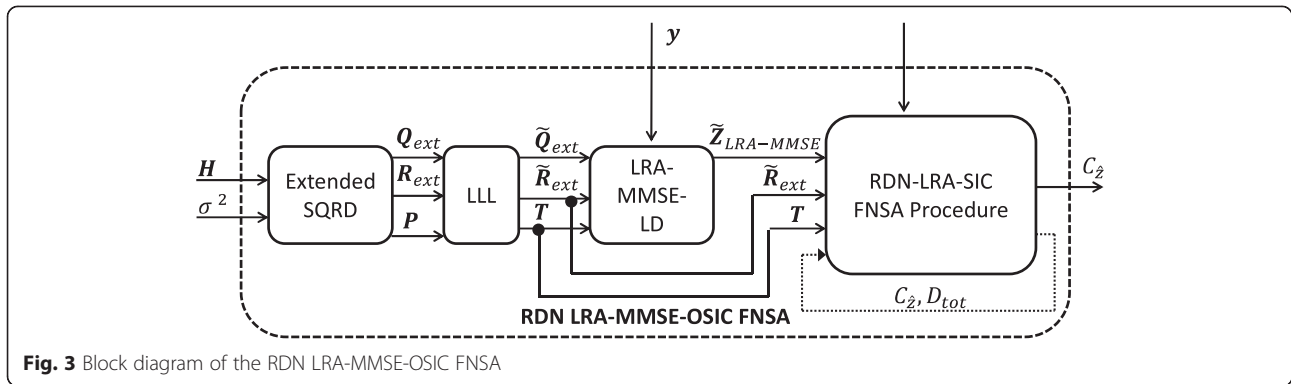
Page 6 of 17

Fig. 3 Block diagram of the RDN LRA-MMSE-OSIC FNSA

4.2 Lattice reduction-aided minimum mean square error-centered sphere decoder with reduced domain neighborhood study detection process

To the best of the author's knowledge, no convincing formula has been proposed until now. Even if Jalden et al. [38] proposed a LRA-MMSE-centered solution, the introduced metrics are not equivalent to the ML expression. The solution of [38] is given by

$$\hat{z}_{\alpha,\ \mathrm{ML}} = \underset{z\in\mathbb{Z}_{\mathbb{C}}^{n_{\mathrm{T}}}}{\mathrm{argmin}} \|\tilde{R}^{-1}\,R^{-1\dagger}\,H^{\dagger}y - z\|^2$$
$$= \underset{z\in\mathbb{Z}_{\mathbb{C}}^{n_{\mathrm{T}}}}{\mathrm{argmin}} \|z_{\text{LRA-MMSE}} - z\|^2 \qquad (11)$$

The corresponding detector is a sub-optimal solution that consists in a RDN study around the unconstrained LRA-MMSE solution, obtained through QR decomposition. This solution's output is the constrained LRA-MMSE detection plus a list of solutions in the neighborhood. The latter is generated according to a non-equivalent metric, which would be subsequently reordered according to the exact ML metric. However, the list is not generated according to the correct distance minimization criterion and would not lead to a near-ML solution. Consequently, the proposed detector does not offer an acceptable uncoded BER performance in the sense that it would not lead to a near-ML solution. In particular, the ML performance is not reached in the case of a large neighborhood study.

An efficient solution is derived from (11) and consists in an unconstrained LRA-MMSE center which leads to a LRA-MMSE-SIC procedure with a RDN study at each layer. The equivalent ML equation reads

$$\hat{z}_{\text{LRA-SIC}} = \underset{z\in\mathbb{Z}_{\mathbb{C}}^{n_{\mathrm{T}}}}{\mathrm{argmin}} \|\tilde{U}(\tilde{z} - z)\|^2, \qquad (12)$$

where $\tilde{H}^H\tilde{H} + \sigma^2 T^H T = \tilde{U}^H\tilde{U}$ in the MMSE case ($\tilde{H}^H\tilde{H} = \tilde{U}^H\tilde{U}$ in the ZF case) and by noting that $\tilde{U}$ is upper triangular with real diagonal elements and $\tilde{z}$ is any LRA (ZF or MMSE) unconstrained linear estimate. The proof of this detector formula is given in Appendix 3.

The formula introduced in (12) offers an equivalent metric to the MMSE one introduced in (11), which has been shown to be near-ML performance. The difference, and in particular the interest in the LRA case in (12), relies on the neighborhood study nature. In the case of a RDN study, the equivalent channel matrix $\tilde{H}$ is considered and is remembered to be only roughly, and not exactly, orthogonal. Consequently, the detection, layer by layer, of the symbol vector x does not exactly correspond to its joint detection since the mutual influence of the transformed z signal is still present. This discussion not only exhibits the interest of SD-like techniques to still improve such a detector performance but also puts a big challenge to achieve the ML performance.

The general principle of RDN LRA-MMSE-OSIC-centered solution key points is depicted as a block diagram in Fig. 1. The detailed block diagram description of the proposed solution is addressed in Fig. 2.

Table 1 ODN naïve (O)SIC FNSA, ODN ZF-(O)SIC FNSA, ODN MMSE-(O)SIC FNSA, RDN LRA-ZF-(O)SIC FNSA, RDN LRA-MMSE-(O)SIC FNSA, and ML formulas

Technique designation	Corresponding metric
ODN naïve (O)SIC FNSA	$\|Q^H y - Rx\|^2$, $x\in\xi^{n_{\mathrm{T}}}$
ODN exact ZF-(O)SIC FNSA	$\|R(y_{\mathrm{ZF}} - x)\|^2$, $x\in\xi^{n_{\mathrm{T}}}$ [14]
ODN equivalent MMSE-(O)SIC FNSA	$(y_{\mathrm{MMSE}} - x)^H(H^H H + \sigma^2 I)(y_{\mathrm{MMSE}} - x)$, $x\in\xi^{n_{\mathrm{T}}}$ [38]
RDN exact LRA-ZF-(O)SIC FNSA	$\|\tilde{R}(z_{\text{LRA-ZF}} - z)\|^2$, $z\in\mathbb{Z}_{\mathbb{C}}^{n_{\mathrm{T}}}$
RDN equivalent LRA-MMSE-(O)SIC FNSA (proposed)	$(z_{\text{LRA-MMSE}} - z)^H(\tilde{H}^H\tilde{H} + \sigma^2 T^H T)(z_{\text{LRA-MMSE}} - z)$, $z\in\mathbb{Z}_{\mathbb{C}}^{n_{\mathrm{T}}}$
ML	$\|y - Hx\|^2$, $x\in\xi^{n_{\mathrm{T}}}$

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 7 of 17

In Fig. 2, the mapping of any estimate (or list of estimates) from the reduced domain $\hat{z}$ to the original domain $\tilde{\tilde{x}}$ is processed through the $\boldsymbol{T}$ matrix multiplication (see Equation (3)). The additional quantification step aims at removing duplicate symbol vector outputs in the case of a list of solutions.

For the sake of simplicity, let us consider any LRA-SIC procedure with no neighborhood study. The search center is updated at each layer as follows. By considering the k-th layer and with the knowledge of the $\hat{z}_{k+1:n_T}$ estimates at previous layers, the $\hat{z}_k$ unconstrained Babai point can be provided. Then, it has to be de-normalized and shifted to make it belong to $\mathbb{Z}_{\mathbb{C}}^{n_T}$. After quantization, and de-shifting and normalization, the $\hat{z}_k$ estimate at the k-th layer is obtained such as the next $(k-1)$-th layer can be considered, until the

whole symbol vector is detected. As previously introduced, the neighborhood generation is a problematic step due to the infiniteness and non-regular natures of the constellations in the reduced domain. This point is transparent with classical detectors such as LD and DFD, thanks to the straight-forward quantification step in the reduced domain [39].

However, the issue of infinite lattices, addressed through a sphere constraint, appears when working with the classical considerations. It presents a performance loss or a NP-hard complexity solution. Hence, our proposed solution relies on a SE enumeration. Starting from the LRA-SIC principle, a neighborhood is considered at each layer and leads to the RDN LRA-SIC FNSA principle. In particular and due to the implementation

Fig. 4 Uncoded BER of the ODN ZF-SIC-centered FNSA (*curve 1*), of the ODN MMSE-SIC-centered FNSA (*curve 2*), of the RDN LRA-ZF-SIC-centered FNSA (*curve 3*), of the RDN LRA-MMSE-SIC-centered FNSA (*curve 4*), of the RDN LRA-MMSE-OSIC-centered FNSA (*curve 5*), and of the ML (*curve 6*), for K = {1, 2, 4, 16} (*top left*, *top right*, *bottom left*, and *bottom right*, respectively), 4 × 4 complex Rayleigh channel, QPSK modulation on each layer

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 8 of 17

constraints, the RDN generation is processed for bounded number of N possibilities and in a SE fashion, namely with ordered PEDs according to an increasing distance from $\tilde{z}_k$ at each layer as follows:

$$
\begin{aligned}
\mathbf{z}_k = \ & Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\},\ Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\} + 1,\ Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\} \\
& + j,\ Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\},\ Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\} - 1,\ Q_{\mathbb{Z}_{\mathbb{C}}^{n_T}}\{\tilde{z}_k\} - j,\ \dots
\end{aligned}
$$

$$(13)$$

The SE strategy aims at finding the correct decision early, leading to a safe early termination criterion, which is not considered here for the sake of readability in performance comparison. Also, all the corresponding PEDs are computed and then ordered. The K-best solutions, namely with the lowest PED, in the reduced domain are stored ($C_{\hat{z}}$) similarly to their corresponding cumulative Euclidean distances (CED) (D_{tot}). The whole procedure is depicted in Fig. 2.

By adding the pre-processing steps, i.e., the SQRD-based then LLL-based LR blocks, and the computation of a close-to-ML unconstrained estimate (although linear) such as LRA-MMSE extended, a complete description of the detection may be obtained. Figure 3 shows the detailed block diagram of the complete proposed solution. The SQRD block offers an efficient layer
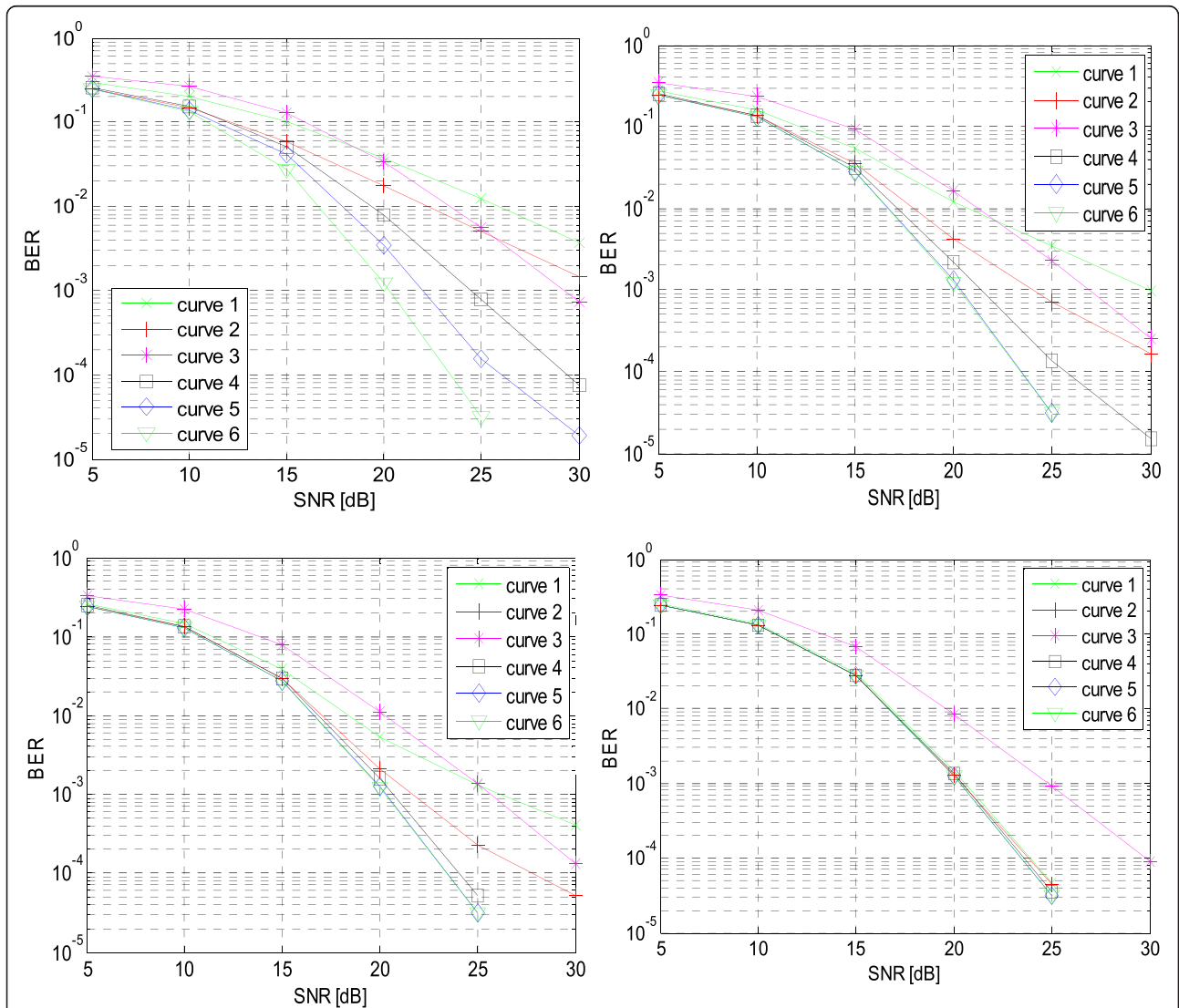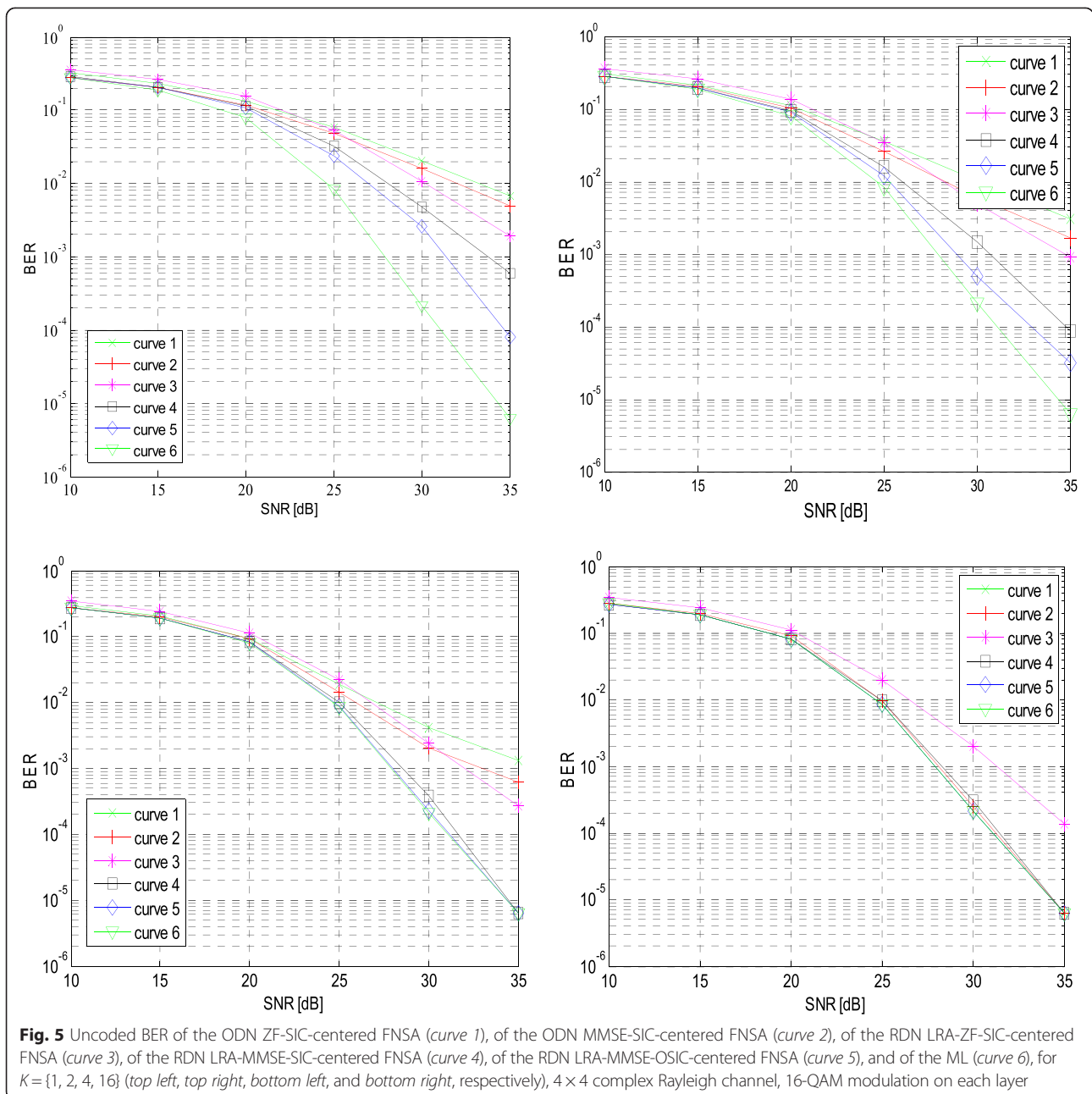

Fig. 5 Uncoded BER of the ODN ZF-SIC-centered FNSA (*curve 1*), of the ODN MMSE-SIC-centered FNSA (*curve 2*), of the RDN LRA-ZF-SIC-centered FNSA (*curve 3*), of the RDN LRA-MMSE-SIC-centered FNSA (*curve 4*), of the RDN LRA-MMSE-OSIC-centered FNSA (*curve 5*), and of the ML (*curve 6*), for $K = \{1, 2, 4, 16\}$ (*top left, top right, bottom left,* and *bottom right,* respectively), 4×4 complex Rayleigh channel, 16-QAM modulation on each layer
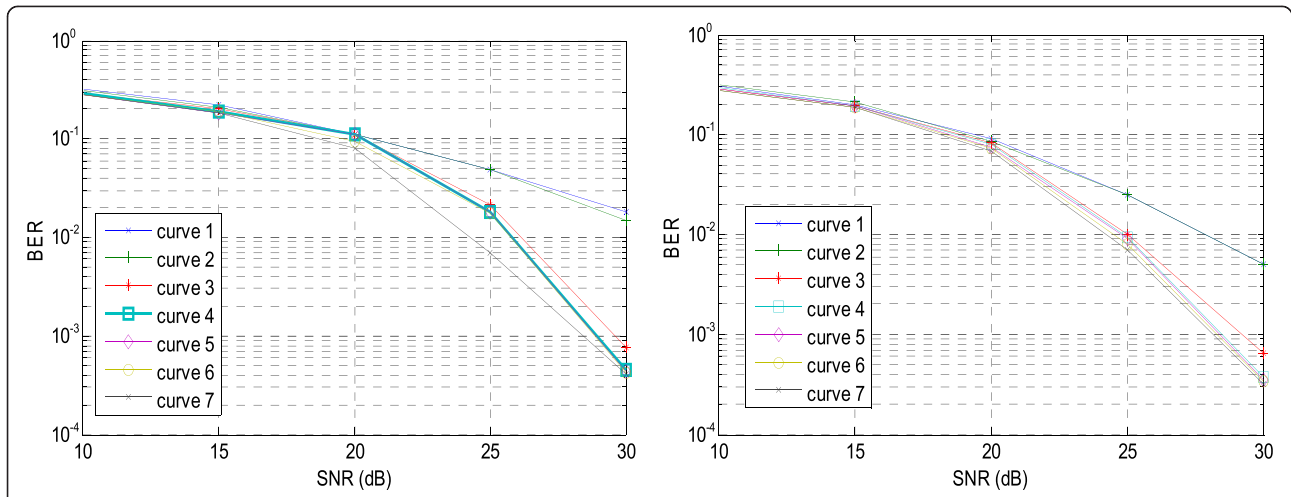
Fig. 6 Uncoded BER of the strictly equivalent ODN MMSE-SIC-centered FNSA, of the strictly equivalent RDN LRA-MMSE-SIC-centered FNSA, of the strictly equivalent RDN LRA-MMSE-OSIC-centered FNSA, compared to the assumption respect in mean, and of the ML, for $K = \{2, 4\}$, 4×4 complex Rayleigh channel, 16-QAM modulation on each layer. *Some curves are coincident*

re-ordering [19] that lies on the noise power. The latter is taken into account in the rest of the detector through the T matrix.

As a final step of the detector and in the case of a RDN-based SD, the list of possible symbols output has to be re-ordered according to the ML metrics in the original domain and duplicate solutions are removed. It is due to the presence of noise that makes some candidates to be mapped on non-legitimate constellation points in the reduced constellation, leading to non-acceptable points in the original constellation. The symbol vector associated to the minimal metric becomes the hard decision output of the detector and offers a near-ML solution. The proposed algorithm is described in detail in Appendix 4. The reader may refer to this appendix for more details.

5 System performance

In this section, we present and compare the system performance of the different techniques previously presented, and we compare them with our proposed solution. For clearness target, we summarize the detection metrics for each solution in Table 1.

We should note that the RDN LRA-MMSE-OSIC FNSA, to which this paper relates, is particularly efficient in the case of rank-deficient MIMO systems, i.e., spatially correlated antenna systems, for high-order modulation which are considered points of the LTE-A norm and for
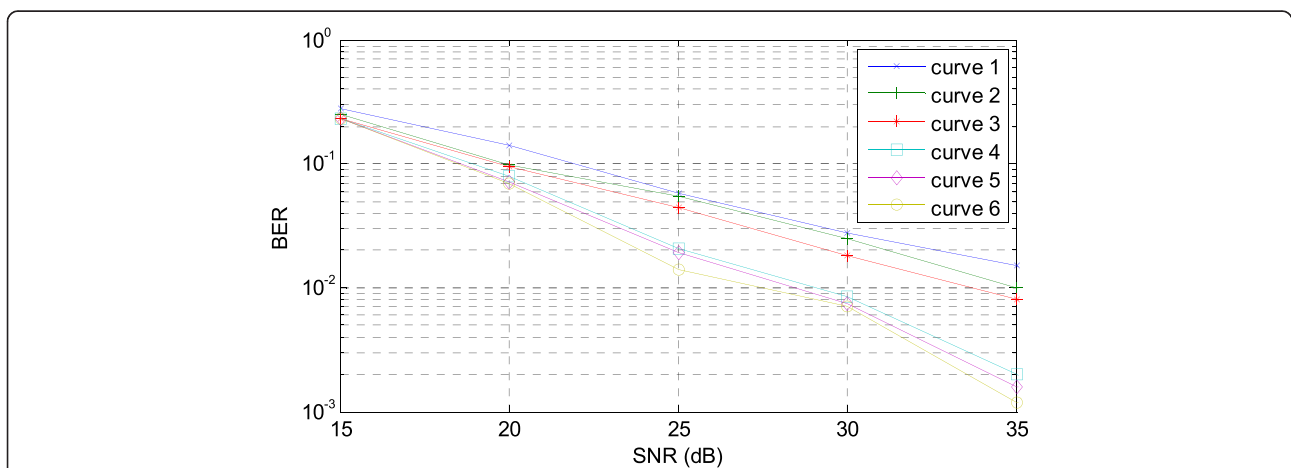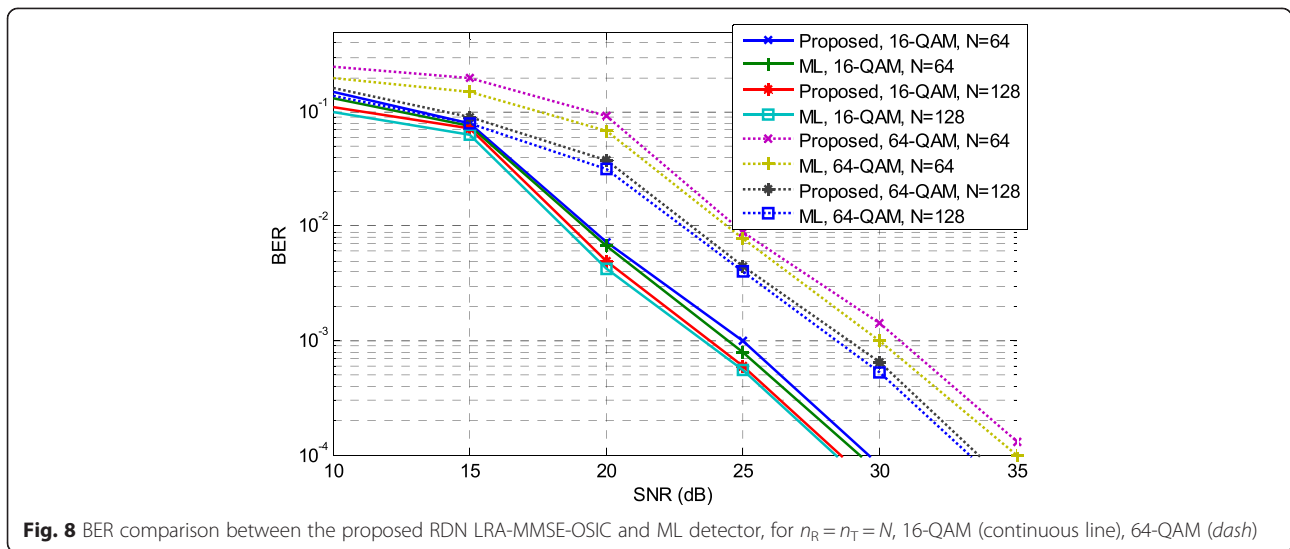

Fig. 7 Uncoded BER of the ODN ZF-SIC-centered FNSA (*curve 1*), of the ODN MMSE-SIC-centered FNSA (*curve 2*), of the RDN LRA-ZF-SIC-centered FNSA (*curve 3*), of the RDN LRA-MMSE-SIC-centered FNSA (*curve 4*), of the RDN LRA-MMSE-OSIC-centered FNSA (*curve 5*), and of the ML (*curve 6*), for $K = 4$, 4×4 complex Rayleigh channel, 64-QAM modulation on each layer

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 10 of 17

Fig. 8 BER comparison between the proposed RDN LRA-MMSE-OSIC and ML detector, for $n_R = n_T = N$, 16-QAM (continuous line), 64-QAM (*dash*)

large number of antennas as in the future generation of cellular systems (beyond 4G networks). Moreover, since the equivalent channel matrix in the LRA case is only roughly orthogonal, the mutual influence of the transformed z is small but still present. Hence, a neighborhood study in the original constellation domain improves the performance compared to a SIC. However, contrarily to classical solutions that are not LRA, the necessary size for achieving the optimal performance is smaller.

Figure 4 depicts the BER for the aforementioned techniques. Some notable points have to be highlighted from this figure. Contrary to the RDN LRA-ZF/MMSE-(O)SIC FNSA, the ODN ZF/MMSE-SIC FNSA does not reach the ML diversity for a reasonable neighborhood size, even if there is a decrease of the SNR offset in the MMSE-SIC

case. However, a BER offset can be observed in the low SNR range, due to error propagation. Consequently, there exists a switching point from low to high SNR between LRA detectors and others. This aspect is removed through the use of better techniques. In particular, the SQRD in the RDN LRA-MMSE-OSIC FNSA presented in this work offers ML diversity, and the BER offset in low SNR has been highly reduced compared to the RDN LRA-MMSE-SIC FNSA and is now close-to-ML.

It may also be noticed in Fig. 4 that the RDN LRA-ZF-SIC-centered FNSA does not reach the ML performance, contrarily to other techniques. It is due to the chosen neighborhood size in the reduced constellation value ($N = 5$) that is not sufficient for this detector but that is sufficient for the proposed LRA-MMSE-(O)SIC

Fig. 9 Uncoded BER with imperfect channel estimation, of the ODN MMSE-SIC-centered FNSA (FNSA curve), of the proposed RDN LRA-MMSE-OSIC-centered FNSA (proposed), and of the ML, with perfect channel estimation $\Delta = 0$ and real channel estimation, $\Delta = 0.001$ (*left*) and $\Delta = 0.005$ (*right*), for $K = 4$, (4×4 complex Rayleigh channel, QPSK modulation on each layer)

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 11 of 17

Table 2 Computational complexity equivalences

Complex operations	Real operations	MUL
ADD_{CC}	2ADD	0
ADD_{RC}	ADD	0
ADD_{RR}	ADD	0
MUL_{CC}	4MUL + 4ADD	4
MUL_{RC}	2MUL	2
MUL_{RR}	1MUL	1
DIV_{CC}	6DIV + 6ADD	96
DIV_{RC}	2DIV	32
DIV_{RR}	1DIV	16
$SQRT_{RR}$	1SQRT	32

Babai points. With a larger N value, the RDN LRA-ZF-SIC-centered FNSA achieves the ML performance, similarly to other presented detectors.

Similarly to Fig. 4, some notable points have to be highlighted from Fig. 5. There still exists a switching point from low to high SNR regime between LRA detectors and others. This aspect is removed through the use of better techniques. In particular, the SQRD in the RDN LRA-MMSE-OSIC FNSA offers ML diversity and the BER offset in low SNR has been importantly reduced compared to the RDN LRA-MMSE-SIC FNSA, leading now to a close-to-ML solution. We can observe from both Figs. 4 and 5 that even though when ZF-SIC and equivalent MMSE-SIC are not LRA, they achieve the ML performance at the detriment of a very large neighborhood study size; it is of the order of the number of symbols contained in the employed constellation. By comparing the impact on LRA detector performance of QPSK and 16-QAM modulations, two fundamental points must be discussed. Firstly, there implicitly exists a constraint from the QPSK constellation construction that eliminates nearby lattice points that do not belong to ξ^{n_T}, due to the quantization operation $Q_{\xi^{n_T}}\{.\}$. This aspect annihilates a large part of the LR-aid benefit and cannot be corrected despite the increase of the neighborhood study size since many lattice points considered in the RDN would be associated with the same constellation point after quantization in the original constellation. In the case of larger constellation orders, the LRA solution is more effective, as depicted in Fig. 5.

Secondly, we recall that the constant modulus constellation assumption has, in theory, to be fulfilled. It was not the case in Fig. 5 with 16-QAM modulation on each layer. However, it could be assumed that this constraint would be almost respected in mean value as shown in Appendix 1 (Fig. 12). In Fig. 6, the performance of R(O)DN (LRA)-MMSE-(O)SIC FNSA detectors with or without respect of this assumption are depicted, but only for a neighborhood scan of 1 and 2 neighbors for the sake of consistency between QPSK and 16-QAM performance.

As depicted in Fig. 6 and with 16-QAM modulation, the performance is impacted by the fact that the strict equivalence assumption is not true, i.e., the term $x^H x$ (or $z^H z$) is not exactly constant but only constant in average. As shown in this figure, this assumption is not constraining in terms of performance loss. Moreover, it is insignificant compared to the advantage of the LRA in high-order constellation, which would be annihilated by the use of QPSK constellation.

The proposed solution is particularly efficient for a large number of antennas and for high-order constellations. It was not the case of the LRA-MMSE-OSIC that has been shown worse BER performance in 4×4 MIMO systems with a 16-QAM modulation on each layer, compared to the ML detection [40], while it was the case for 4×4 MIMO systems with QPSK modulation on each layer [41]. For the sake of completeness of this work, Fig. 7 shows the same results with 64-QAM modulation as those given in Fig. 5. Again this figure shows the outperformance of the proposed detection algorithm with high-order constellation.

Figure 8 shows the comparison between the proposed RDN LRA-MMSE-OSIC-centered FNSA and the ML detection for high number of antennas, such that $n_R = n_T = N = 64$ and $N = 128$ and, $K = 2$. First, there is no doubt that increasing the number of antennas increases the performance gain. Secondly, the proposed solution shows a comparable performance with respect to the ML decoder. At a BER = 10^{-4}, the SNR loss is less than 0.4 dB

Table 3 ODN ZF-(O)SIC FNSA, ODN MMSE-(O)SIC FNSA, RDN LRA-ZF-(O)SIC FNSA, RDN LRA-MMSE-(O)SIC FNSA, and ML formulas

Technique designation	Corresponding computational complexity in MUL
ODN exact ZF-(O)SIC	$2MKn_T^2 + 2MKn_T - 4MK + 3M$
ODN equivalent MMSE-(O)SIC	$2MKn_T^2 + 2MKn_T - 4MK + 3M$
RDN exact LRA-ZF-(O)SIC	$2N \min\{K,N\}n_T^2 + 30\min\{K,N\}n_T + 2N \min\{K,N\}n_T - 4N\min\{K,N\}$
	$+ 6\min\{K,N\}n_T^2 + 4\min\{K,N\}n_R n_T + 2\min\{K,N\}n_R + 4n_T^2 - 32\min\{K,N\} + 2N$
RDN equivalent LRA-MMSE-(O)SIC	$2N \min\{K,N\}n_T^2 + 30\min\{K,N\}n_T + 2N \min\{K,N\}n_T - 4N\min\{K,N\}$
	$+ 6\min\{K,N\}n_T^2 + 4\min\{K,N\}n_R n_T + 2\min\{K,N\}n_R + 4n_T^2 - 32\min\{K,N\} + 2N$
ML	$4n_R n_T M^{n_T}$

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 12 of 17

Table 4 SNR loss at BER = 10^{-4}, ODN ZF-SIC FNSA, ODN MMSE-SIC FNSA, RDN LRA-ZF-SIC FNSA, RDN LRA-MMSE-SIC FNSA, and RDN LRA-MMSE-OSIC FNSA compared to ML

Technique	SNR loss (QPSK)				SNR loss (16-QAM)			
	K = 1	K = 2	K = 3	K = 4	K = 1	K = 2	K = 4	K = 16
ODN exact ZF-SIC FNSA	>7.6	>7.6	>7.6	0.36	>5.0	>5.0	>5.0	0
ODN equivalent MMSE-SIC FNSA	>7.6	>7.6	6.21	0.30	>5.0	>5.0	>5.0	0.09
RDN exact LRA-ZF-SIC FNSA	4.43	2.90	1.92	1.71	3.21	2.04	1.27	0.62
RDN equivalent LRA-MMSE-SIC FNSA	2.90	0.73	0.52	0.27	2.12	0.76	0.53	0.40
RDN equivalent LRA-MMSE-OSIC FNSA	0.80	0.01	0	0	1.62	0.02	0	0

for 16-QAM and less than 0.5 dB for 64-QAM while the complexity of the proposed RDN LRA-MMSE-OSIC-centered FNSA solution is by far much lower than the ML decoder. This will be discussed in the next section.

Finally, even though it is not the target of the paper, we have drawn the simulation results of the proposed solution with real channel estimation. Figure 9 shows the simulation results when the channel estimation error variance Δ is equal to 0.001 and 0.005, assuming that the channel coefficients power is normalized by the number of antennas. This figure shows that the proposed LRA-MMSE solution still presents quasi-ML detection even with real channel estimation.

6 Complexity evaluation

Based on the assumptions presented in Table 1, the computational complexities introduced in Table 2 can be demonstrated. The RDN study is processed in an infinite lattice which would not lead to boundary control; however, a finite set of displacements has been generated in a SE fashion in simulations. Its size has been fixed to an arbitrary value ($N = 5$)—decided through simulations. Although an SE technique is used, the proposed solution does not consider any complexity reduction like early termination.

As shown in Table 3, the computational complexities of RDN LRA-ZF/MMSE-(O)SIC FNSA detectors do not depend on the constellation order $\log_2\{M\}$. It may be checked in the numerical applications in Table 4, and it is the key point of the paper advantage over classical techniques for high-order modulations such as 16(64)-

QAM. The SNR loss compared to ML are given in Table 4. They have been measured for an uncoded BER of 10^{-4} in the case of the ML decoder. For all the configurations given in Table 4, the numerical application of the corresponding computational complexity is given in Table 5 for a RDN size $N = 5$.

Even if the proposed solution is two times more complex in the QPSK case, it offers near-ML performance and in particular a SNR gain of 0.3 dB at a BER of 10^{-4}. The interesting point concerns higher order modulations: starting from the 16-QAM modulation, the estimated complexity of the proposed solution is ten times less complex than the classical one, for the same performance result. Identically, same conclusions are obtained for a 64-QAM modulation. In such case, the complexity gain will increase importantly to reach a hundred times. Similarly, the numerical application of the 16-QAM extension complexity is given in Table 6. As an example, in the case of 16-QAM modulations, the computational complexities read $8MKn_{\mathrm{T}}^2 + 4MKn_{\mathrm{T}} - 4MK + 3M$ for the ODN equivalent MMSE-(O)SIC and $8N\min\{K,N\}n_{\mathrm{T}}^2 + 60\min\{K,N\}n_{\mathrm{T}} + 4N\min\{K,N\}n_{\mathrm{T}} - 4N\min\{K,N\} + 24\min\{K,N\}n_{\mathrm{T}}^2 + 8\min\{K,N\}n_{\mathrm{R}}n_{\mathrm{T}} + 2\min\{K,N\}n_{\mathrm{R}} + 16n_{\mathrm{T}}^2 - 32\min\{K,N\} + 2N$ for the RDN equivalent LRA-MMSE-(O)SIC, and with $M = 4$ since a QPSK modulation is considered in this case. As depicted in Table 6, the computational complexity of the 16-QAM extension with respect to the constant modulus criterion is more important compared to the straightforward but not strictly correct solution. Since no significant gain is provided, we consequently claim it does not offer high advantages.

Table 5 ODN ZF-SIC, ODN MMSE-SIC, RDN LRA-ZF-SIC, RDN LRA-MMSE-SIC, RDN LRA-MMSE-OSIC, and ML computational complexities in MUL

Technique	MUL (QPSK)				MUL (16-QAM)			
	K = 1	K = 2	K = 3	K = 4	K = 1	K = 2	K = 4	K = 16
ODN ZF-(O)SIC FNSA	156	300	444	588	624	1200	2352	9264
ODN MMSE-(O)SIC FNSA	156	300	444	588	624	1200	2352	9264
RDN LRA-ZF-(O)SIC FNSA	510	946	1382	1818	510	946	1818	2254
RDN LRA-MMSE-(O)SIC FNSA	510	946	1382	1818	510	946	1818	2254
ML	16,384				4,194,304			

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 13 of 17

Table 6 ODN MMSE-SIC, RDN LRA-MMSE-SIC, and RDN LRA-MMSE-OSIC computational complexities in MUL

Technique	MUL (16-QAM extension)				MUL (16-QAM)			
	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 1$	$K = 2$	$K = 4$	$K = 16$
ODN equivalent MMSE-(O)SIC FNSA	560	1120	1680	2240	624	1200	2352	9264
RDN equivalent LRA-MMSE-(O)SIC FNSA	1694	3122	4550	5978	510	946	1818	2254

Figure 10 shows the "measured" complexity of all solutions explored in this work versus the constellation size, expressed in terms of the exponent (in base 10) of the computational capacity in MUL, for $n_R = n_T = 8$ and $K = 2$. This figure shows, as explained earlier, that the proposed solution is independent of the constellation size. This is very crucial in the future large MIMO systems exploiting large dimensions. Figure 11 is in line with the previous conclusion. It provides the computation complexity of the different MIMO detection solutions, expressed as a function of the number of antennas. This figure shows that the proposed solution is almost ten times less complex than the classical K-best solutions. Moreover, it presents almost equal complexity for $n_T \geq 32$ yielding another important characteristic for large MIMO decoding.

Finally, to give some concrete example, Table 6 compares between ODN and RDN cases. It shows that the proposed solution offers an advantage over existing solutions when applied to any OFDM standard supporting MIMO spatial-multiplexing mode, e.g., IEEE 802.16, IEEE 802.11, 3GPP LTE, and 3GPP LTE-A. It may be advantageously considered in the case of a large number of antennas and consequently in the case of the 3GPP LTE-A standard. The main advantages reside in the following points:

- The equivalent expression of the LRA-MMSE-centered SD, which corresponds to an efficient LRA-MMSE-

OSIC Babai point, improves the performance or reduces the complexity of the detector.
- The proposed (S)QRD formulation with reduced domain neighborhood induces the use of the best known hard detector as a Babai point, for both large number of antennas and high-order modulations.
- The proposed expression is robust by nature to any search center and constellation order and offers close-to-optimal performance for large K. Likewise, the proposed solution offers a computational complexity that is independent of the constellation order which consequently offers a solution that outperforms classical SD techniques for a reasonable computational complexity in the case of high-order constellations. For instance, the neighborhood study size K has been reduced to $K = 2$ for a 16-QAM modulation compared to classical SD techniques.

7 Conclusions

In this paper, the LRA-MMSE-centered SD has been proposed with a K-best neighborhood generation. A detailed and hardware implementation-oriented computational complexity estimation has been provided and combined with performance results. It has been shown that the proposed detection technique outperforms the existing solutions. In particular, the corresponding
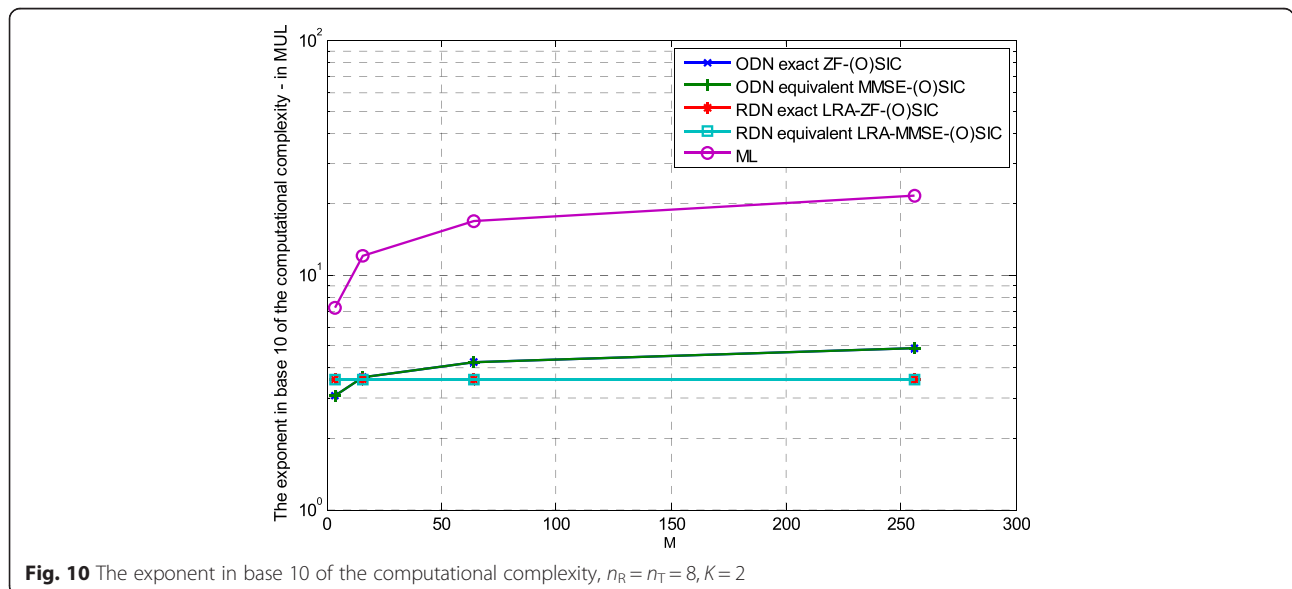

Fig. 10 The exponent in base 10 of the computational complexity, $n_R = n_T = 8$, $K = 2$
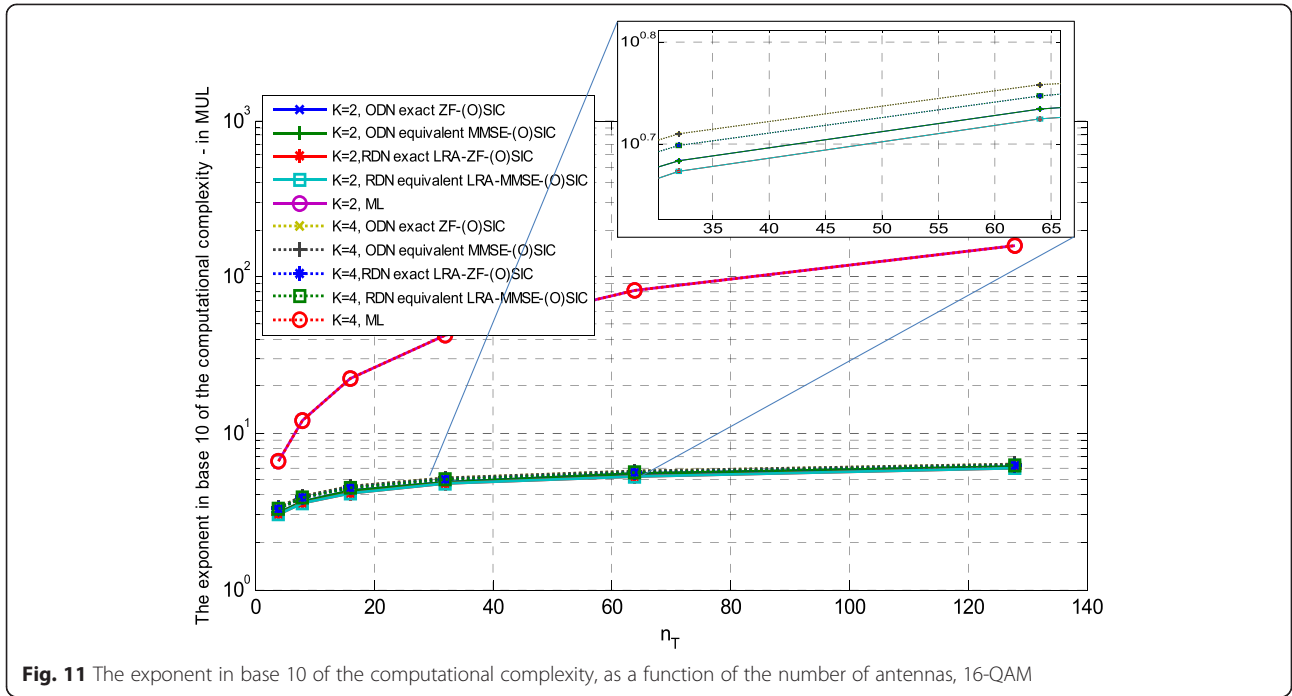
Fig. 11 The exponent in base 10 of the computational complexity, as a function of the number of antennas, 16-QAM

implementation complexity has been shown to be independent of the constellation size and polynomial in the number of antennas while reaching the ML performance with both real and perfect channel estimation. It implies a ten times lower computational complexity compared to the classical K-best, even for a large MIMO system, with 16-QAM modulation on each layer.

Endnotes
[1]It is worth mentioning that, with respect to our previous work in [1], this paper presents a detailed technical description of the proposed methodology, a detailed complexity analysis, and more results. This particularly includes a step by step implementation of the proposed algorithm in Appendix 4.

Appendices
Appendix 1: the constant modulus constraint in (6)
The authors of [42] discussed the constant modulus constraint of $\boldsymbol{x}$ in (6) when n_T is large. It has been shown that the constant modulus signal assumption becomes the time average of the n_T $\boldsymbol{x}_i$ entries. Figure 12 presents the probability density functions (PDF) of $\boldsymbol{x}^H \boldsymbol{x}$ for different number of transmit antennas and different modulations. This figure shows that, due to the weak law of large numbers, the term is Gaussian centered to a mean value that is constant in time. Consequently, the assumption may still be considered as fulfilled as n_T

increases. It is worth mentioning that, in order to make (6) strictly equivalent to the ML metric, any M-QAM constellation may be represented as a weighted sum of QPSK constellations:
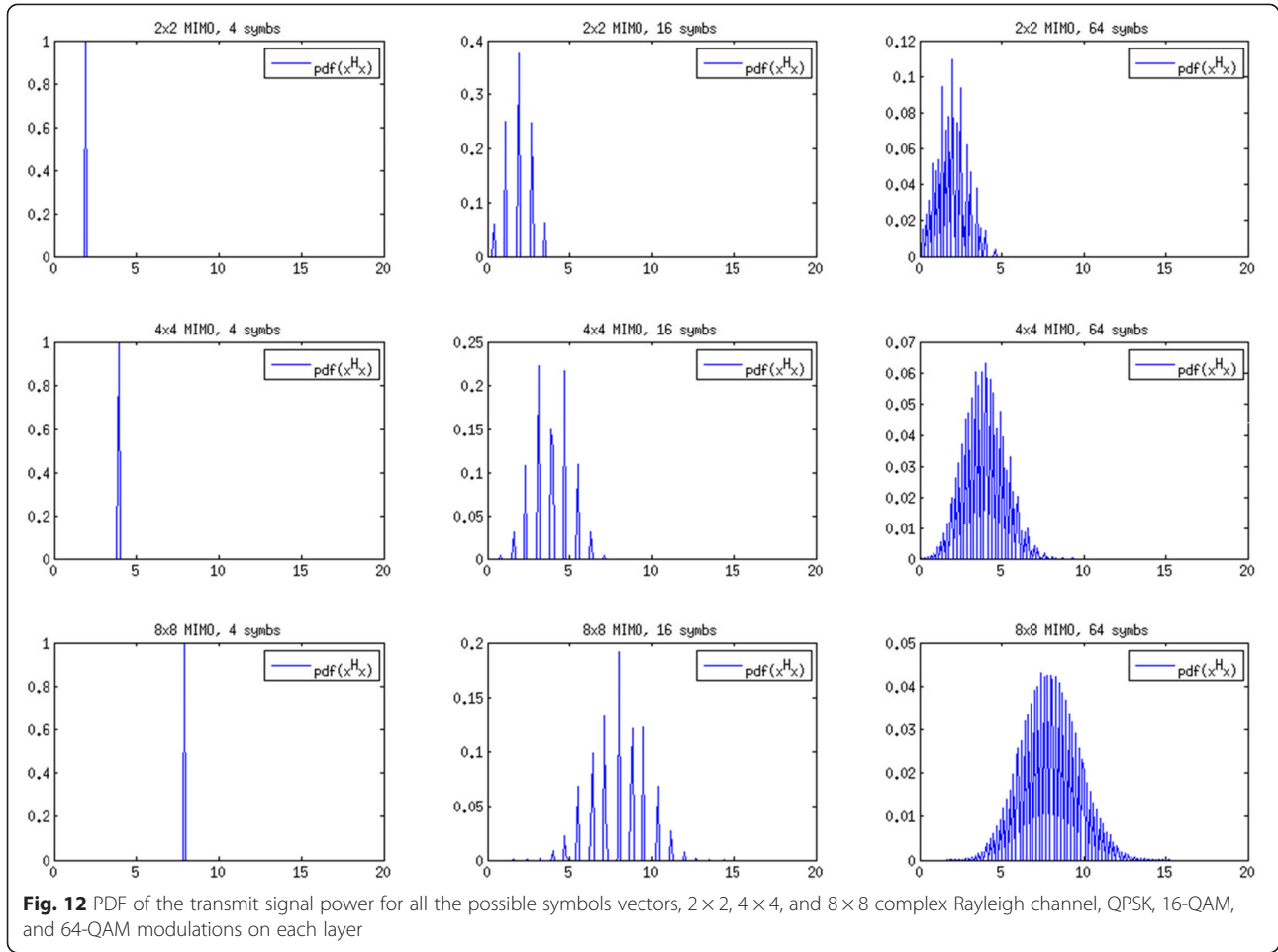
$$\boldsymbol{x}^{(M\text{-QAM})} = \sum_{i=0}^{\log_2\{M\}-1} 2^i \left(\frac{\sqrt{2}}{2}\right) \boldsymbol{x}_i^{(\text{QPSK})} \qquad (14)$$

Where $\boldsymbol{x}^{(M\text{-QAM})}$ is an n_T symbols vector whose entries all belong to a M-QAM constellation and $\boldsymbol{x}_i^{(\text{QPSK})}$ is an n_T symbol vector whose all entries belong to a QPSK constellation.

Appendix 2: proof of Equation (8)
Let us introduce any term c s.t. $\|\boldsymbol{y}-\boldsymbol{Hx}\|^2 + c = \|\boldsymbol{U}(\tilde{\boldsymbol{x}}-\boldsymbol{x})\|^2$, where $\tilde{\boldsymbol{x}}$ is any (ZF or MMSE) unconstrained linear estimate:

$$
\begin{aligned}
c &= \|\boldsymbol{U}(\tilde{\boldsymbol{x}}-\boldsymbol{x})\|^2 - \|\boldsymbol{y}-\boldsymbol{Hx}\|^2 \\
&= (\tilde{\boldsymbol{x}}-\boldsymbol{x})^H \boldsymbol{U}^H \boldsymbol{U}(\tilde{\boldsymbol{x}}-\boldsymbol{x}) - (\boldsymbol{y}-\boldsymbol{Hx})^H(\boldsymbol{y}-\boldsymbol{Hx}) \\
&= \tilde{\boldsymbol{x}}^H \boldsymbol{G}\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^H \boldsymbol{Gx} - \boldsymbol{x}\boldsymbol{G}\tilde{\boldsymbol{x}} + \boldsymbol{x}^H \boldsymbol{Gx} - \boldsymbol{y}^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{Hx} \\
&\quad + \boldsymbol{x}^H \boldsymbol{H}^H\boldsymbol{y} - \boldsymbol{x}^H \boldsymbol{H}^H\boldsymbol{Hx}, \text{ with } \boldsymbol{G} = \boldsymbol{U}^H\boldsymbol{U}, \\
&= \boldsymbol{y}^H \boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{H}^H\boldsymbol{y} - \boldsymbol{y}^H\boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{Gx} - \boldsymbol{x}\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{H}^H\boldsymbol{y} \\
&\quad + \boldsymbol{x}^H\boldsymbol{Gx} - \boldsymbol{y}^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{Hx} + \boldsymbol{x}^H\boldsymbol{H}^H\boldsymbol{y} - \boldsymbol{x}^H\boldsymbol{H}^H\boldsymbol{Hx},
\end{aligned}
$$

Fig. 12 PDF of the transmit signal power for all the possible symbols vectors, 2×2, 4×4, and 8×8 complex Rayleigh channel, QPSK, 16-QAM, and 64-QAM modulations on each layer

by introducing $\tilde{\boldsymbol{x}} = \boldsymbol{G}^{-1}\boldsymbol{H}^H\boldsymbol{y}$ and $\tilde{\boldsymbol{x}}^H = \boldsymbol{y}^H\boldsymbol{H}\boldsymbol{G}^{-1}$ and where $\boldsymbol{G} = \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{H}^H\boldsymbol{H}$ in the ZF case and $\boldsymbol{G} = \boldsymbol{H}^H\boldsymbol{H} + \sigma^2\boldsymbol{I}$ in the MMSE case.

$$c = \boldsymbol{y}^H\boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{H}^H\boldsymbol{y} + \boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}^H\boldsymbol{y} - \boldsymbol{x}^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{x}$$
$$= \boldsymbol{y}^H\boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{H}^H\boldsymbol{y} + \boldsymbol{x}^H\left(\boldsymbol{G} - \boldsymbol{H}^H\boldsymbol{H}\right)\boldsymbol{x} - \boldsymbol{y}^H\boldsymbol{y}$$

In the ZF case, $\boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{H}^H = \boldsymbol{H}\boldsymbol{H}^{-1}(\boldsymbol{H}^H)^{-1}\boldsymbol{H}^H = \boldsymbol{I}$ and $\boldsymbol{G} - \boldsymbol{H}^H\boldsymbol{H} = 0$, consequently $c = 0$ which is a constant term.

In the MMSE case, $c = \boldsymbol{y}^H[\boldsymbol{H}(\boldsymbol{H}^H\boldsymbol{H} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{H}^H - \boldsymbol{I}]\boldsymbol{y} + \sigma^2\boldsymbol{x}^H\boldsymbol{x}$ which is a constant term in $\boldsymbol{x}$ iff the signal $\boldsymbol{x}$ entries are of constant modulus.

Appendix 3: proof of Equation (12)

The proof of Equation (12) is very similar to the proof of (8); however, in this appendix, we work on the LRA-based detector. Let us introduce any term c' s.t.

$\|\boldsymbol{y} - \tilde{\boldsymbol{H}}\boldsymbol{z}\|^2 + c' = \|\tilde{\boldsymbol{U}}\left(\tilde{\tilde{\boldsymbol{z}}} - \boldsymbol{z}\right)\|^2$, where $\tilde{\tilde{\boldsymbol{z}}}$ is any LRA (ZF or MMSE) unconstrained linear estimate:

$$c' = \|\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{z}} - \boldsymbol{z})\|^2 - \|\boldsymbol{y} - \tilde{\boldsymbol{H}}\boldsymbol{z}\|^2$$
$$= (\tilde{\boldsymbol{z}} - \boldsymbol{z})^H\tilde{\boldsymbol{U}}^H\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{z}} - \boldsymbol{z}) - (\boldsymbol{y} - \tilde{\boldsymbol{H}}\boldsymbol{z})^H(\boldsymbol{y} - \tilde{\boldsymbol{H}}\boldsymbol{z})$$
$$= \tilde{\boldsymbol{z}}^H\tilde{\boldsymbol{G}}\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{z}}^H\tilde{\boldsymbol{G}}\boldsymbol{z} - \boldsymbol{z}^H\tilde{\boldsymbol{G}}\tilde{\boldsymbol{z}} + \boldsymbol{z}^H\tilde{\boldsymbol{G}}\boldsymbol{z} - \boldsymbol{y}^H\boldsymbol{y} + \boldsymbol{y}^H\tilde{\boldsymbol{H}}\boldsymbol{z}$$
$$+ \boldsymbol{z}^H\tilde{\boldsymbol{H}}^H\boldsymbol{y} - \boldsymbol{z}^H\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}\boldsymbol{z}, \text{ with } \tilde{\boldsymbol{U}}^H\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{G}},$$
$$= \boldsymbol{y}^H\tilde{\boldsymbol{H}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{H}}^H\boldsymbol{y} - \boldsymbol{y}^H\tilde{\boldsymbol{H}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{G}}\boldsymbol{z} - \boldsymbol{z}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{H}}^H\boldsymbol{y}$$
$$+ \boldsymbol{z}^H\tilde{\boldsymbol{G}}\boldsymbol{z} - \boldsymbol{y}^H\boldsymbol{y} + \boldsymbol{y}^H\tilde{\boldsymbol{H}}\boldsymbol{z} + \boldsymbol{z}^H\tilde{\boldsymbol{H}}^H\boldsymbol{y} - \boldsymbol{z}^H\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}\boldsymbol{z}$$

by introducing $\tilde{\boldsymbol{z}} = \tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{H}}^H\boldsymbol{y}$ and $\tilde{\boldsymbol{z}}^H = \boldsymbol{y}^H\tilde{\boldsymbol{H}}\tilde{\boldsymbol{G}}^{-1}$, where $\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}$ in the LRA-ZF case and $\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}} + \sigma^2\boldsymbol{T}^H\boldsymbol{T}$ in the LRA-MMSE case.

$$c' = \boldsymbol{y}^H\tilde{\boldsymbol{H}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{H}}^H\boldsymbol{y} + \boldsymbol{z}^H\tilde{\boldsymbol{G}}\boldsymbol{z} - \boldsymbol{y}^H\boldsymbol{y} - \boldsymbol{z}^H\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}\boldsymbol{z}$$
$$= \boldsymbol{y}^H\tilde{\boldsymbol{H}}\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{H}}^H\boldsymbol{y} + \boldsymbol{z}^H\left(\tilde{\boldsymbol{G}} - \tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}\right)\boldsymbol{z} - \boldsymbol{y}^H\boldsymbol{y}$$

In the ZF case, $\tilde{H}\tilde{G}^{-1}\tilde{H}^H = \tilde{H}\tilde{H}^{-1}(\tilde{H}^H)^{-1}\tilde{H}^H = I$ and $\tilde{G} - \tilde{H}^H\tilde{H} = 0$, consequently $c' = 0$ is a constant term.

In the MMSE case, $c' = y^H[\tilde{H}(\tilde{H}^H\tilde{H} + \sigma^2 T^H T)^{-1}\tilde{H}^H - I]y + \sigma^2 z^H T^H T z$ which is a constant term in x iff the signal x entries are of constant modulus since $\sigma^2 z^H T^H T z = \sigma^2 x^H x$.

Appendix 4: description of the proposed detection algorithm: RDN LRA-ZF-(O)SIC *K*-best

---

Algorithm 1: RDN LRA-ZF-(O)SIC K-Best

| | Inputs: | H, σ^2, y, K |
| | Outputs: | C_z |

0. $[Q_{ext}, R_{ext}, P] = sqrd\{H, \sigma^2\}$

1. $[\tilde{Q}_{ext}, \tilde{R}_{ext}, T] = lll\{Q_{ext}, R_{ext}, P, \frac{3}{4}\}$

2. $z_{center,1:n_T} = (\tilde{H}^H\tilde{H} + \sigma^2 T^H T)^{-1}\tilde{H}^H y$

3. $d = \frac{1}{2}T^{-1}[1+j, \ldots, 1+j]^T$

4. $a = 2scale$

5. $\tilde{z}_{1:n_T} = z_{center,1:n_T}$

6. $z_: = rdn\_generation\{\tilde{z}_{n_T}, N\}$

7. $for\ l = 1:N, do$

8. $D_{tot,l} = |\tilde{R}_{n_T,n_T}(\tilde{z}_{n_T} - z_l)|^2$

9. end

10.   % Sort the N PED candidates in ascending order, and store the $\min\{K, N\}$ lowest

11. $C_{\hat{z}_{n_T,1:\min\{K,N\}}} = ordered\_candidates_{n_T,1:\min\{K,N\}}$

12. $D_{prev,1:\min\{K,N\}} = ordered\_peds_{n_T,1:\min\{K,N\}}$

13. $for\ n = n_T - 1 : -1 : 1, do$

14. $for\ k = 1 : \min\{K, N\}, do$

15. $sum = \sum_{i=n+1}^{n_T}\{\tilde{R}_{n,i}(\tilde{z}_i - C_{\hat{z}_{i,k}})\}$

16. $z_: = rdn\_generation\{\frac{sum}{\tilde{R}_{n,n}} + \tilde{z}_n, N\}$

17. $for\ l = 1 : N, do$

18. $D_{tot,(k-1)N+l} = D_{prev,k} + |\tilde{R}_{n,n}(z_{center,n} - z_l) + sum|^2$

19. end

20. end

21.   % Sort the $N \min\{K, N\}$ PED candidates in ascending order, and store the $\min\{K, N\}$ lowest

22. $C_{\hat{z}_{n:n_T,1:\min\{K,N\}}} = ordered\_candidates_{n:n_T,1:\min\{K,N\}}$

23. $D_{prev,1:\min\{K,N\}} = ordered\_peds_{n,1:\min\{K,N\}}$

24. end

25. $C_{\tilde{x}_{:,1:\min\{K,N\}}} = TC_{\hat{z}_{:,1:\min\{K,N\}}}$

26. $for\ k = 1 : \min\{K, N\}, do$

27. $C_{\hat{x}_{:,k}} = Q_{\xi^{n_T}}\{C_{\tilde{x}_{:,k}}\}$

28. end

29.   % Sort the $\min\{K, N\}$ candidates in ascending order in the Original Domain

---

Nasser *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:227

Page 17 of 17

Author details
[1]ECE Department, American University of Beirut, Bliss Street, Beirut, Lebanon.
[2]ST-Ericsson, 635 route des Lucioles, Sophia-Antipolis, France. [3]Université
Européenne de Bretagne, INSA, IETR, UMR 6164, 35708 Rennes, France.

References
1. S Aubert, Y Nasser, F Nouvel, Lattice reduction-aided minimum mean square error k-best detection for MIMO systems, in *Proc. of the International Conference Computing, Networking and Communications (ICNC)*, 2012, pp. 1066–1070
2. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Processing Magazine **30**(1), 40–46 (2013)
3. EG Larsson, F Tufvesson, O Edfors, TL Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
4. Y Kong, Q Zhou, X Ma, Lattice reduction aided transceiver design for multiuser MIMO downlink transmissions, in *Proc. of the IEEE Military Communications Conference (MILCOM)*, 2014, pp. 556–562
5. KA Singhal, T Datta, A Chockalingam, Lattice reduction aided detection in large-MIMO systems, in *Proc. of the IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2013, pp. 594–598
6. E Zimmermann, G Fettweis, Linear MIMO receivers vs. tree search detection: a performance comparison overview, in *Proc. of the IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, 2006, pp. 1–7
7. N Prasad, MK Varanasi, Analysis of decision feedback detection for MIMO Rayleigh-fading channels and the optimization of power and rate allocations. IEEE Transactions on Information Theory **50**(6), 1009–1025 (2004)
8. R Xu, FCM Lau, Performance analysis for MIMO systems using zero forcing detector over fading channels. IEE Proceedings on Communications **153**(1), 74–80 (2006)
9. Y Nasser, J-F Hélard, M Crussière. System Level Evaluation of Innovative Coded MIMO-OFDM Systems for Broadcasting Digital TV; in EURASIP International Journal of Digital Multimedia Broadcasting. **2008**(359206), 12 (2008)
10. E Viterbo, J Boutros, A universal lattice code decoder for fading channels. IEEE Trans. on Information Theory **45**, 1639–1642 (1997)
11. B Hassibi, H Vikalo, On the expected complexity of sphere decoding, in *Proc. of the Asimolar Conference on Signal, Systems and Computers*, 2001, pp. 1051–1055
12. C Schnorr, M Euchner, Lattice basis reduction: improved practical algorithms and solving subset sum problems. Mathematical Programming **66**, 181–199 (1994)
13. Z Guo, P Nilsson, Algorithm and implementation of the K-best sphere decoding for MIMO detection. IEEE Journal on Selected Areas in Communications **24**(3), 491–503 (2006)
14. LG Barbero, JS Thompson, A fixed-complexity MIMO detector based on the complex sphere decoder. IEEE 7th Workshop on Signal Processing Advances in Wireless Communications, 2006. SPAWC '06. pp. 1, 5, 2–5 (2006)
15. M Mohaisen, KyungHi Chang, On improving the efficiency of the fixed-complexity sphere decoder. 2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), 20–23 Sept 2009, pp. 1, 5
16. Y Ding, Y Wang, JF Diouris, Z Yao, Robust fixed-complexity sphere decoders for rank-deficient MIMO systems. IEEE Trans. Wireless Commun **12**(9), 4297–4305 (2013)
17. J Fink, S Roger, A Gonzalez, V Almenar, VM Garciay, Complexity assessment of sphere decoding methods for MIMO detection. 2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 14–17 Dec 2009, pp. 9, 14
18. C Hess, M Wenk, A Burg, P Luethi, C Studer, N Felber, W Fichtner, Reduced-complexity MIMO detector with close-to ML error rate performance, in *Proc. of the GLSVLSI*, 2007, pp. 200–203
19. D Wuebben, R Bohnke, V Kuhn, K-D Kammeyer, MMSE-based lattice-reduction for near-ML detection of MIMO systems, in *Proc. of the ITG Workshop on Smart Antennas*, 2004, pp. 106–113
20. S Roger, A Gonzales, V Almenar, AM Vidal, Lattice-reduction-aided K-best MIMO detector based on the channel matrix condition number. 2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP), March 2010, pp. 1–4
21. C-F Liao, Y-H Huang, Cost reduction algorithm for 8x8 lattice reduction-aided K-best MIMO detector, in *Proc. of the IEEE International Conference of Signal Processing, Communication and Computing*, 2012, pp. 186–190
22. X-F Qi, K Holt, A lattice-reduction-aided soft demapper for high-rate coded MIMO-OFDM systems. IEEE Signal Processing Letters **14**(5), 305–308 (2007)
23. M Shabany, PG Gulak, The application of lattice-reduction to the K-best algorithm for near-optimal MIMO detection. IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008. 18–21 May 2008, pp. 316–319
24. JC Marinello, T Abrao, Lattice reduction aided detector for dense MIMO via ant colony optimization, in *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 2839–2844. Shanghai
25. LG Barbero, JS Thompson, A fixed-complexity MIMO detector based on the complex sphere decoder, in *Proc. of the Workshop on Signal Processing Advances for Wireless Communications*, 2006, pp. 1–5
26. E Agrell, T Eriksson, E Vardy, K Zeger, Closest point search in lattices. IEEE Transactions on Information Theory **48**(8), 2201–2214 (2002)
27. K-W Wong, C-Y Tsui, S-K Cheng, W-H Mow, A VLSI architecture of a K-best lattice decoding algorithm for MIMO channels, in *Proc. of the IEEE International symposium on Circuits and Systems*, vol. 3, 2002, pp. 273–276
28. E Viterbo, J Boutros, A universal lattice code decoder for fading channels. IEEE Transactions on Information Theory **45**(5), 1639–1642 (1999)
29. S Aubert, M Mohaisen, From linear equalization to lattice-reduction-aided sphere-detector as an answer to the MIMO detection problematic in spatial multiplexing systems. Vehicular Technologies, 978-953-7619-X-X, INTECH, (2011)
30. BA Lamacchia, Basis reduction algorithms and subset sum problems. Technical report, MSc Thesis, Massachusetts Institute of Technology, 1991
31. AK Lenstra, HW Lenstra, L Lovász, Factoring polynomials with rational coefficients. Mathematische Annalen **261**(4), 515–534 (1982)
32. M Seysen, Simultaneous reduction of a lattice basis and its reciprocal basis. Combinatorica **13**(3), 363–376 (1993)
33. D Wübben, R Böhnke, V Kühm, K-D Kammeyer, Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice-reduction, in *Proc. of the IEEE International Conference on Communications*, vol. 2, 2004, pp. 798–802
34. S Roger, A Gonzalez, V Almenar, AM Vidal, On decreasing the complexity of lattice-reduction-aided K-best MIMO detectors, in *Proc. of the European Signal Processing Conference*, 2009, pp. 2411–2415
35. B Gestner, W Zhang, X Ma, DV Anderson, VLSI implementation of a lattice reduction algorithm for low-complexity equalization, in *Proc. of the IEEE International Conference on Circuits and Systems for Communications*, 2008, pp. 643–647
36. T Cui, C Tellambura, An efficient generalized sphere decoder for rank-deficient MIMO systems. IEEE Communications Letters **9**(5), 423–425 (2005)
37. L Wang, L Xu, S Chen, L Hanzo, MMSE soft-interference-cancellation aided iterative center-shifting K-best sphere detection for MIMO channels, in *the Proc. of the IEEE International Conference on Communications*, 2008, pp. 3819–3823
38. J Jalden, B Ottersten, On the complexity of sphere decoding in digital communications. IEEE Transactions on Signal Processing **53**(4), 1474–1484 (2005)
39. X Wang, Z He, K Niu, W Wu, X Zhang, An improved detection based on lattice reduction in MIMO systems, in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2006, pp. 1–5
40. C Studer, A Burg, H Bolcskei, Soft-output sphere decoding: algorithms and VLSI implementation. IEEE Journal on Selected Areas in Communications **26**(2), 290–300 (2008)
41. W Zhang, M Xiaoli, Approaching optimal performance by lattice-reduction aided soft detectors. 41st Annual Conference on Information Sciences and Systems, 2007. CISS '07. 14–16 March 2007, pp. .818–822
42. M Pohst, On the computation of lattice vectors of minimal length, successive minima and reduced basis with applications. ACM SIGSAM Bull. **15**, 37–44 (1981)