Reduction of ML Decoding Complexity for MIMO Sphere Decoding, QOSTBC, and OSTBC

Luay Azzam and Ender Ayanoglu

Center for Pervasive Communications and Computing
Department of Electrical Engineering and Computer Science
University of California, Irvine
email: lazzam@uci.edu, ayanoglu@uci.edu

***Abstract*— In this paper, we discuss three applications of the QR decomposition algorithm to decoding in a number of Multi-Input Multi-Output (MIMO) systems. In the first application, we propose a new structure for MIMO Sphere Decoding (SD). We show that the new approach achieves 80% reduction in the overall complexity compared to conventional SD for a 2×2 system, and almost 50% reduction for the 4×4 and 6×6 cases. In the second application, we propose a low complexity Maximum Likelihood Decoding (MLD) algorithm for quasi-orthogonal space-time block codes (QOSTBCs). We show that for $N = 8$ transmit antennas and 16-QAM modulation scheme, the new approach achieves $> 97\%$ reduction in the overall complexity compared to conventional MLD, and $> 89\%$ reduction compared to the most competitive reported algorithms in the literature. This complexity gain becomes greater when the number of transmit antennas (N) or the constellation size (L) becomes larger. In the third application, we propose a low complexity Maximum Likelihood Decoding (MLD) algorithm for orthogonal space-time block codes (OSTBCs) based on the real-valued lattice representation and QR decomposition. For a system employing the well-known Alamouti OSTBC and 16-QAM modulation scheme, the new approach achieves $> 87\%$ reduction in the overall complexity compared to conventional MLD. Moreover, we show that for square L-QAM constellations, the proposed algorithm reduces the decoding computational complexity from $\mathcal{O}(L^{N/2})$ for conventional MLD to $\mathcal{O}(L)$ for systems employing QOSTBCs and from $\mathcal{O}(L)$ for conventional MLD to $\mathcal{O}(\sqrt{L})$ for those employing OSTBCs without sacrificing the performance.**

I. Introduction

Multiple-input multiple-output (MIMO) communication systems provide high data rates and improved performance without increasing the bandwidth or transmitted power. Consequently, MIMO techniques have become a significant part of most current wireless systems. Maximum Likelihood Decoding (MLD) is the optimum decoding algorithm that is used for MIMO systems [1]. However, MLD complexity increases exponentially with the number of antennas and the constellation order being used for modulation [2]. Therefore, Sphere Decoder (SD) or the Fincke-Pohst algorithm [3], [4] was proposed as an alternative to provide MLD performance with less complexity.

The use of Space-time block codes (STBCs) along with MIMO is very useful to maximize spatial diversity and introduce high capacity gains [5]. Among these codes we focus on orthogonal space-time block codes (OSTBCs) [6] and quasi-orthogonal space-time block codes (QOSTBCs) [7]-[9]. OSTBCs are attractive since they achieve the maximum diversity, the maximum coding gain, and the highest throughput [6] as well as simple MLD. However, full-rate OSTBCs do not exist for systems with more than $N = 2$ transmit antennas [10]. Consequently, QOSTBCs were proposed in which the constraint of orthogonality is relaxed to obtain higher symbol transmission rate [7]-[9]. In general, QOSTBCs do not achieve the full diversity provided by the channel. Therefore, a number of rotation techniques were proposed to provide full diversity [11]-[14].

MLD of QOSTBCs is performed by searching over a subset of the total number of transmitted symbols. More specifically, a joint detection of at least two complex symbols is required for a full-rate system with $N = 4$ antennas, or $3/4$-rate system with $N = 8$. For the full-rate system with $N = 8$, four complex symbols are jointly detected to obtain the ML solution [12], [15]. Unlike OSTBCs, the decoding complexity is no longer linear, but rather, increases exponentially with N, i.e., $\mathcal{O}(L^{N/2})$ where L is the size of the L-QAM constellation [15], [16]. For OSTBCs, MLD is performed simply by decoding each transmitted symbol independently, resulting in linear decoding complexity [10], [17].

The decoding complexity is very critical for practical employment of MIMO systems. Therefore, the development of low complex decoding algorithms while providing optimal performance is always a necessity for wireless communication systems. In this paper, we focus on SD complexity for uncoded systems as well as MLD complexity for OSTBCs and QOSTBCs. We propose a new lattice representation to be used for SD which is verified to reduce the overall decoding complexity while providing optimal performance. This new structure, along with a couple of techniques (adaptive k-best and quantization [18]), enables reduction of the decoding complexity by $> 50\%$ compared to conventional SD. Subsequently, we introduce a novel optimal decoding algorithm for square QAM constellations based on QR decomposition of the real-valued lattice representation and show that the optimal MLD performance for QOSTBCs and OSTBCs is obtained with a substantial reduction in the decoding complexity.

The remainder of this paper is organized as follows: In Section II, we briefly discuss SD algorithm and propose the new lattice representation. In Section III, we define the system model for QOSTBCs and introduce our new decoding algorithm. A brief discussion on diversity is provided. In Section IV, we propose a simplified MLD algorithm for systems employing OSTBCs. Simulation results are included in Section V. Finally, we conclude the paper in Section VI.

II. Sphere Decoding

A. Conventional SD Algorithm

Consider a MIMO system with *N* transmit and *M* receive antennas. The received signal vector is given by

$$y = Hs + v \tag{1}$$

where $y \in \mathbb{C}^M$ is an *M*-dimensional received complex vector, $s \in \mathbb{C}^N$ is an *N*-dimensional transmitted complex vector whose entries have real and imaginary parts that are integers, $H \in \mathbb{C}^{M\text{x}N}$ is the channel matrix, $v \in \mathbb{C}^M$ is the i.i.d. complex additive white Gaussian noise (AWGN) vector with zero-mean and covariance matrix $\sigma^2 I$. Usually, the elements of the vector s are constrained to a finite set Ω where $\Omega \subset \mathbb{Z}^{2N}$, e.g., $\Omega = \{-3,-1,1,3\}^{2N}$ for 16-QAM where $\mathbb{Z}$ and $\mathbb{C}$ denote the sets of integers and complex numbers respectively.

Assuming H is known at the receiver, the ML detection is given by

$$\hat{s} = \arg\min_{s\in\Omega} ||y - Hs||^2. \tag{2}$$

Solving (2) becomes impractical and exhaustive for high transmission rates, and the complexity grows exponentially [19]. Therefore, SD solves this problem by searching for the closest point among all lattice points that lie inside a sphere centered around the received vector y and of radius d [3], [20], [21]. The algorithm runs recursively until all lattice points inside the sphere are found. If no points were found inside the sphere, then we increase the radius and start over again. Now, introducing this radius constraint on (2) changes the problem to

$$\hat{s} = \arg\min_{s\in\Omega} ||y - Hs||^2 < d^2. \tag{3}$$

A frequently used solution for the QAM-modulated complex signal model given in (3) is to decompose the N-dimensional problem into a $2N$-dimensional real-valued problem, which then can be written as

$$\begin{bmatrix} \Re\{y\} \\ \Im\{y\} \end{bmatrix} = \begin{bmatrix} \Re\{H\} & -\Im\{H\} \\ \Im\{H\} & \Re\{H\} \end{bmatrix} \begin{bmatrix} \Re\{s\} \\ \Im\{s\} \end{bmatrix} + \begin{bmatrix} \Re\{v\} \\ \Im\{v\} \end{bmatrix} \tag{4}$$

where $\Re\{y\}$ and $\Im\{y\}$ denote the real and imaginary parts of y respectively. Assuming $N = M$ and introducing the QR decomposition of H, where *R* is an upper triangular matrix, and the matrix *Q* is unitary, (3) can be written as

$$\hat{s} = \arg\min_{s\in\Omega} ||\bar{y} - Rs||^2 < d^2 \tag{5}$$

where $\bar{y} = Q^H y$. Let $R = [r_{i,j}]_{2N\text{x}2N}$ and note that R is upper triangular, then the search can be performed by starting from $i = 2N$ and working backwards. Whenever a valid lattice point is found inside the sphere, the square of the sphere radius d^2 is set to the newly found point weight, thus reducing the search space for finding other candidate solutions.

To this end, it is important to emphasize the fact that the complexity of SD algorithm, although much lower than MLD is still high and can be further reduced as will be shown in this paper.

B. New Lattice Representation

The lattice representation given in (4) imposes a major restriction on the search algorithm. Specifically, the search is executed serially from one layer to another. This can be made clearer by writing the partial metric weight formula as

$$w_l(x^{(l)}) = w_{l+1}(x^{(l+1)}) + |\hat{y}_l - \sum_{k=l}^{2N} r_{l,k} x_k|^2 \tag{6}$$

with $l = 2N, 2N-1, \ldots, 1$, $w_{2N+1}(x^{(2N+1)}) = 0$ and where $\{x_1, x_2, ..., x_N\}$, $\{x_{N+1}, x_{N+2}, ..., x_{2N}\}$ are the real and imaginary parts of $\{s_1, s_2, ..., s_N\}$ respectively

Obviously, SD starts from the last layer ($l = 2N$), working its way up one layer at a time, computing the partial weight metric of one or more points in the lattice space and taking out those points that violate the radius constraint until reaching the first layer ($l = 1$). According to this representation, it is impossible for instance to calculate $\sum_{k=l}^{2N} r_{l,k} x_k$ in (6) for a lattice point that lies at layer $l = 2N-1$ without assigning an estimate for x_{2N}. This approach results in two related drawbacks. First, the decoding of any x_l requires an estimate value for all preceding x_j for $j = l+1, ..., 2N$. Secondly, there is no room for parallel computations since the structure of the tree search is sequential.

Our goal is to relax the lattice structure and make it more flexible for parallelism. Concurrently, we attempt to reduce the number of computations required at each lattice point. This can be done by making the decoding of every two adjacent layers totally independent of each other. To do so, we start by reshaping the channel matrix representation given in (4) to the following form

$$\tilde{H} = \begin{bmatrix} \Re(H_{1,1}) & -\Im(H_{1,1}) & \cdots & \Re(H_{1,N}) & -\Im(H_{1,N}) \\ \Im(H_{1,1}) & \Re(H_{1,1}) & \cdots & \Im(H_{1,N}) & \Re(H_{1,N}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Re(H_{N,1}) & -\Im(H_{N,1}) & \cdots & \Re(H_{N,N}) & -\Im(H_{N,N}) \\ \Im(H_{N,1}) & \Re(H_{N,1}) & \cdots & \Im(H_{N,N}) & \Re(H_{N,N}) \end{bmatrix}$$

where $H_{m,n}$ is the i.i.d. complex path gain from transmit antenna n to receive antenna m. By careful observation of the columns of $\tilde{H}$ starting from the left hand side, and defining each pair of columns as one set, we note that the columns in each set are orthogonal, a property that has a substantial effect on the structure of the problem. Using this channel representation changes the order of detection of the transmitted symbols from

$$\hat{s} = \begin{bmatrix} \Re(s_1) & \cdots & \Re(s_N) & \Im(s_1) & \cdots & \Im(s_N) \end{bmatrix}^T$$

to the following order

$$\hat{s} = \left[\begin{array}{ccccc} \Re(s_1) & \Im(s_1) & \cdots & \Re(s_N) & \Im(s_N) \end{array} \right]^T .$$

This structure becomes advantageous after applying the QR decomposition to $\tilde{H}$. By doing so, and due to that special form of orthogonality among the columns of each set, all the elements $r_{k,k+1}$ for $k = 1, 3, ..., 2N-1$ in the upper triangular matrix R become zero. This allows parallel detection of the real and imaginary parts of every detected symbol and significantly reduces the overall decoding complexity.

In [18], we provide a mathematical proof to show that $r_{k,k+1} = 0$ for $k = 1, 3, ..., 2N-1$. This proof is based on the Gram-Schmidt algorithm. We also propose a number of techniques that can be used on top of the new lattice representation to further reduce the computational complexity of SD.

III. Quasi-Orthogonal Space-Time Coding

A. System Model and Proposed Algorithm

Consider a MIMO system with N transmit and M receive antennas, and an interval of T symbols during which the channel is constant. The received signal is given by

$$Y = \sqrt{\frac{\rho}{N}} C_N H + V \tag{7}$$

where $Y = [y_t^j]_{T\times M}$ is the received signal matrix of size $T\times M$ and whose entry y_t^j is the signal received at antenna j at time t, $t = 1, 2, \ldots, T$, $j = 1, 2, \ldots, M$; $V = [v_t^j]_{T\times M}$ is the noise matrix; and $C_N = [c_t^i]_{T\times N}$ is the transmitted signal matrix whose entry c_t^i is the signal transmitted at antenna i at time t, $t = 1, 2, \ldots, T$, $i = 1, 2, \ldots, N$. $H = [h_{i,j}]_{N\times M}$ is the channel coefficient matrix of size $N\times M$ whose entry $h_{i,j}$ is the channel coefficient from transmit antenna i to receive antenna j. The entries of the matrices H and V are mutually independent, zero-mean, and circularly symmetric complex Gaussian random variables of unit variance; and the parameter ρ is the signal-to-noise-ratio (SNR) per receiving antenna.

Assuming that the channel H is known at the receiver, the ML estimate is obtained at the decoder by performing $\min_C ||Y - \sqrt{\frac{\rho}{N}} CH||_F^2$, where $||.||_F$ is the Frobenius norm. For QOSTBCs, the measure $||Y - \sqrt{\frac{\rho}{N}} CH||_F^2$ can be decoupled into two parts, where each part solves $N/2$ symbols concurrently [16]. A number of techniques were proposed in the literature to reduce the decoding complexity [22], [23]. In this paper, we show that this complexity can still be reduced substantially. In the following, we only consider $M = 1$ for simplicity.

We start by rewriting (7) in matrix form, then we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = C_N \begin{bmatrix} h_{1,1} \\ h_{2,1} \\ \vdots \\ h_{N,1} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}. \tag{8}$$

We decompose the T-dimensional complex problem defined by (8) to a $2T$-dimensional real-valued problem. Applying the new lattice representation discussed in the previous section to (8), we obtain

$$\begin{bmatrix} \Re(y_1) \\ \Im(y_1) \\ \vdots \\ \Re(y_T) \\ \Im(y_T) \end{bmatrix} = \check{H} \begin{bmatrix} \Re(s_1) \\ \Im(s_1) \\ \vdots \\ \Re(s_m) \\ \Im(s_m) \end{bmatrix} + \begin{bmatrix} \Re(v_1) \\ \Im(v_1) \\ \vdots \\ \Re(v_T) \\ \Im(v_T) \end{bmatrix} \tag{9}$$

where $\check{H}$ is the equivalent real-valued channel whose real-valued fading coefficients are defined using the complex fading coefficients $h_{i,j}$ from transmit antenna i to receive antenna j as $h_{2n-1} = \Re(h_{n,1})$, and $h_{2n} = \Im(h_{n,1})$ for $n = 1, 2, \ldots, N$. Now, we specify the complex transmitted symbols of C_N by their real and imaginary parts as $s_m = x_{2m-1} + jx_{2m}$ for $m = 1, 2, \ldots, n_s$ where n_s is the number of transmitted symbols per code block and defined as $n_s = Tr$, where r is the code rate. Then (9) is equivalent to

$$\tilde{y} = \tilde{H}x + \tilde{v}. \tag{10}$$

In this paper, we focus on the full rate QOSTBC using $N = 4$ and defined as

$$C_4 = \begin{bmatrix} s_1 & s_3 & s_4 & s_2 \\ s_3^* & -s_1^* & s_2^* & -s_4^* \\ s_4^* & s_2^* & -s_1^* & -s_3^* \\ s_2 & -s_4 & -s_3 & s_1 \end{bmatrix}. \tag{11}$$

Note that the following derivation is applicable to arbitrary N. The reader is referred to [24] for more details.

Using (11), the equivalent channel matrix $\check{H} = \left[\begin{array}{cccc} \check{h}_1 & \check{h}_2 & \ldots & \check{h}_8 \end{array} \right]$, where $\check{h}_k$ is the k-th column of $\check{H}$, is given by

$$\check{H} = \begin{bmatrix} h_1 & -h_2 & h_7 & -h_8 & h_3 & -h_4 & h_5 & -h_6 \\ h_2 & h_1 & h_8 & h_7 & h_4 & h_3 & h_6 & h_5 \\ -h_3 & -h_4 & h_5 & h_6 & h_1 & h_2 & -h_7 & -h_8 \\ -h_4 & h_3 & h_6 & -h_5 & h_2 & -h_1 & -h_8 & h_7 \\ -h_5 & -h_6 & h_3 & h_4 & -h_7 & -h_8 & h_1 & h_2 \\ -h_6 & h_5 & h_4 & -h_3 & -h_8 & h_7 & h_2 & -h_1 \\ h_7 & -h_8 & h_1 & -h_2 & -h_5 & h_6 & -h_3 & h_4 \\ h_8 & h_7 & h_2 & h_1 & -h_6 & -h_5 & -h_4 & -h_3 \end{bmatrix}.$$

We observe that

$$\begin{aligned} \langle \check{h}_1, \check{h}_i \rangle &= 0, i \neq 3, & \langle \check{h}_2, \check{h}_i \rangle &= 0, i \neq 4 \\ \langle \check{h}_5, \check{h}_i \rangle &= 0, i \neq 7, & \langle \check{h}_6, \check{h}_i \rangle &= 0, i \neq 8. \end{aligned} \tag{12}$$

where $\langle \check{h}_i, \check{h}_j \rangle$ is the inner product of columns $\check{h}_i$ and $\check{h}_j$. Interchanging the columns of $\check{H}$ such that every two columns that are not orthogonal to each other become adjacent makes the subsequent analysis appear in a compact form. Thus, we rewrite $\check{H}$ as $\tilde{H} = \left[\begin{array}{cccccccc} \check{h}_1 & \check{h}_3 & \check{h}_2 & \check{h}_4 & \check{h}_5 & \check{h}_7 & \check{h}_6 & \check{h}_8 \end{array} \right]$. Then, (12) can be rewritten as

$$\begin{aligned} \langle \tilde{h}_1, \tilde{h}_i \rangle &= 0, i \neq 2, & \langle \tilde{h}_5, \tilde{h}_i \rangle &= 0, i \neq 6 \\ \langle \tilde{h}_3, \tilde{h}_i \rangle &= 0, i \neq 4, & \langle \tilde{h}_7, \tilde{h}_i \rangle &= 0, i \neq 8. \end{aligned} \tag{13}$$

Consequently, the vectors $\tilde{y}$, x and $\tilde{v}$ defined in (10) have the same element values with different rows order given by $1, 3, 2, 4, 5, 7, 6, 8$.

Applying QR decomposition to $\check{H}$ in (10), we have

$$\begin{aligned} \tilde{y} &= QRx + \tilde{v} \\ Q^H \tilde{y} &= Rx + Q^H \tilde{v} \\ \bar{y} &= Rx + \bar{v} \end{aligned} \tag{14}$$

where $\bar{v}$ and $\tilde{v}$ have the same statistical properties since Q is unitary and so is Q^H. Due to (13), it is straightforward to observe that R is block diagonal of the form

$$R = \begin{bmatrix} R_{1,2} & 0 & 0 & 0 \\ 0 & R_{3,4} & 0 & 0 \\ 0 & 0 & R_{5,6} & 0 \\ 0 & 0 & 0 & R_{7,8} \end{bmatrix} \tag{15}$$

where

$$R_{i,i+1} = \begin{bmatrix} r_{i,i} & r_{i,i+1} \\ 0 & r_{i+1,i+1} \end{bmatrix} \qquad i = 1, 3, 5, 7. \tag{16}$$

Note that the elements of the upper triangular matrix R are the inner products of columns of $\tilde{H}$ [25]. Note also that the diagonal elements of R are the norm values of nonzero vectors [25], and thus $r_{i,i}$ and $r_{i+1,i+1}$ for $i = 1, 3, 5, 7$ will never be zeros. Using (14), the ML problem is now simpler and rather than minimizing $||Y - CH||^2$, the solution is obtained by minimizing the metric $||\bar{y} - Rx||^2$ in a layered fashion over all different combinations of the vector x. To make this clearer, let the square L-QAM alphabet be given as Ω^2, where $\Omega = \{-\sqrt{L}+1, -\sqrt{L}+3, \ldots, \sqrt{L}-1\}$. Then

$$\hat{x} = \arg \min_{x \in \Omega^8} ||\bar{y} - Rx||^2. \tag{17}$$

Let x_i and x_{i+1} be given as

$$\arg \min_{\substack{x_i \in \Omega \\ x_{i+1} \in \Omega}} [|\bar{y}_{i+1} - r_{i+1,i+1} x_{i+1}|^2 + |\bar{y}_i - r_{i,i} x_i - r_{i,i+1} x_{i+1}|^2]$$

for $i = 1, 3, 5, 7$. Then, the decoded message is

$$\hat{x} = (x_1, x_2, \ldots, x_8)^T.$$

In other words, the ML solution is obtained by jointly decoding two real symbols through a simpler 2×2 real-valued upper triangular equivalent channel matrix. Note that this simplification is obtained through the orthogonality properties in (13) and the QR decomposition in (14), resulting in (15) and (16). This is similar to but simpler than [23] and [26] in that all have joint detection of two real symbols ($N/2$ times in [23] and [26] and n_s times in this work) but with minimizing a norm of size $2N$ in [23] and [26] while minimizing a norm of size 2 in this work. This means that the original complex ML problem is decomposed into $n_s = 4$ parallel real-valued upper triangular problems, each of dimension 2. Writing this in matrix form, the ML solution is obtained by carrying out

$$\arg \min_{x_i \in \Omega, x_{i+1} \in \Omega} \left|\left| \begin{bmatrix} \bar{y}_i \\ \bar{y}_{i+1} \end{bmatrix} - \begin{bmatrix} r_{i,i} & r_{i,i+1} \\ 0 & r_{i+1,i+1} \end{bmatrix} \begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix} \right|\right|^2$$

for $i = 1, 3, 5, 7$.

Obviously, this approach allows finding the ML solution faster, and requires a small number of computational operations compared to conventional ML, and decoding algorithms proposed in [23] and [26].

B. QOSTBC with Full Diversity

In general, QOSTBCs do not achieve the full diversity provided by the channel. In order to achieve full diversity and improve the performance at high SNR, a conventional approach in the literature suggests that half of the symbols in a quasi-orthogonal design are chosen from a signal constellation set $\mathcal{A}$ and the other half are chosen from a rotated constellation set $e^{j\phi}\mathcal{A}$ [11], [12]. Another approach is to apply multi-dimensional rotated constellations which exhibit full diversity and maximum coding gain [13], [14]. However, no proper expressions for the rotation matrix of sizes greater than four exist. Applying the first approach to our algorithm introduces interference among the real symbols. Instead, a two-dimensional rotation of the real symbols $(x_1, x_2, \ldots, x_{2n_s})$ that are not orthogonal to each other is always applied, thus maintaining the orthogonality properties among the channel columns and maximizing the diversity. By this, we overcome the problem of finding proper expressions for the rotation matrix and maintain the orthogonality properties defined in (13).

The two-dimensional rotation matrix is defined in [14], [26] as

$$G = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{18}$$

where the optimal angle is obtained by $\theta = \frac{1}{2}\text{atan}(\frac{1}{2})$ for square QAM constellations [23].

For a detailed proof on how (18) achieves full diversity for arbitrary number of transmit antennas N while maintaining the same complexity of our proposed algorithm, we refer the reader to [24].

IV. Orthogonal Space-Time Block Coding

In this section, we introduce a third application of the QR decomposition algorithm in MIMO decoding. We consider a MIMO system whose input-output relation is represented by (7), where C_N is an orthogonal space time block code. Generally, OSTBCs have a very simple and decoupled MLD algorithm. For an OSTBC of rate $r = K/T$, the squared norm $||Y - C_N H||_F^2$ can be decoupled into K parts, where each part decodes one transmitted complex symbol independently [27].

Consider the OSTBC proposed by Alamouti [17] for $N = 2$ and defined as

$$C_2 = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}. \tag{19}$$

The receiver decodes s_1 and s_2 by decomposing the measure $||Y - C_2 H||_F^2$ into two parts, and minimizes each separately over all possible values of s_1 and s_2 that belong to the constellation used. In general, the complexity of MLD is $\mathcal{O}(L)$ which is linear with the constellation size L [27]. Thus, the decoding algorithm can be implemented using only linear processing at the receiver. In this paper, we show that this complexity can still be reduced substantially. The proposed algorithm reduces the decoding complexity from $\mathcal{O}(L)$ to $\mathcal{O}(\sqrt{L})$ with a substantial reduction in the number of arithmetic operations required.

Analogous to the derivation for the QOSTBC case, we represent the system by its real-valued representation to obtain

$$\begin{bmatrix} \Re(y_1) \\ \Im(y_1) \\ \vdots \\ \Re(y_T) \\ \Im(y_T) \end{bmatrix} = \check{H} \begin{bmatrix} \Re(s_1) \\ \Im(s_1) \\ \vdots \\ \Re(s_m) \\ \Im(s_m) \end{bmatrix} + \begin{bmatrix} \Re(v_1) \\ \Im(v_1) \\ \vdots \\ \Re(v_T) \\ \Im(v_T) \end{bmatrix}. \tag{20}$$

Since C_N is an orthogonal matrix and due to the real-valued representation of the system using (20), we observe that

- All columns of $\check{H} = \begin{bmatrix} \check{h}_1 & \check{h}_2 & \dots & \check{h}_{2K} \end{bmatrix}$ where $\check{h}_i$ is the ith column of $\check{H}$, are orthogonal to each others, or equivalently
$$\langle \check{h}_i, \check{h}_j \rangle = 0, i \neq j \tag{21}$$
- The norm of every column in $\check{H}$ is equal to the norm of any other column in $\check{H}$, i.e.,
$$\text{norm}(\check{h}_i) = \text{norm}(\check{h}_j), \qquad i, j = 1, 2, \dots, 2K. \tag{22}$$

These two properties have a major impact on the complexity reduction of our proposed decoding algorithm.

Applying QR decomposition to (20) produces

$$\bar{y} = Rx + \bar{v} \tag{23}$$

where R is a $2K \times 2K$ diagonal matrix (see [25] for proof), a property which substantially reduces the decoding complexity. In fact, due to (21) and (22), QR can be simplified into two steps. To illustrate this, let $\check{H} = \begin{bmatrix} \check{h}_1 & \check{h}_2 & \dots & \check{h}_{2K} \end{bmatrix}$ where $\check{h}_i$ is the ith column of $\check{H}$. Then, due to (21) R is diagonal. The definition of the diagonal elements in R in QR decomposition is

$$r_{i,i} = \text{norm}(\check{h}_i).$$

Due to (22) the matrices Q and R are computed by
step 1: calculating the diagonal elements of the matrix R

$$\begin{aligned} r_{1,1} &= \text{norm}(\check{h}_1) \\ r_{i,i} &= r_{1,1} \end{aligned}$$

for $i = 2, \dots, 2K$.
step 2: computing the unitary matrix $Q = \begin{bmatrix} q_1 & q_2 & \dots & q_{2k} \end{bmatrix}$ as

$$q_i = \check{h}_i / r_{i,i}$$

for $i = 1, 2, \dots, 2K$.

Using QR algorithm and the above two observations, the MLD problem is now simpler and rather than minimizing $||Y - C_N H||^2$, the solution is obtained by minimizing the metric $||\bar{y} - Rx||^2$ over all different combinations of the vector x. In other words, the MLD solution is found by minimizing

$$\left\| \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_{2K} \end{bmatrix} - \begin{bmatrix} r_{1,1} & 0 & \cdots & 0 \\ 0 & r_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{2K,2K} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{2K} \end{bmatrix} \right\|^2$$

over all combinations of $x \in \Omega^{2K}$. This can be further simplified as

$$\hat{x}_i = \arg \min_{x_i \in \Omega} |\bar{y}_i - r_{i,i} x_i|^2 \tag{24}$$

for $i = 1, 2, \dots, 2K$. Then, the decoded message is

$$\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{2K})^T.$$

This means that the MLD solution is obtained by decoding the real and imaginary parts of each complex transmitted symbol independently through a simple 1×1 real-valued channel matrix, i.e., a real scalar number. To further clarify our proposed decoding algorithm, we provide some representative examples.

Example 1: In this example, we consider the Alamouti OSTBC defined by (19) with $N = K = T = 2$ and $M = 1$. We consider conventional MLD where the transmitted complex symbols $\hat{s}_1$ and $\hat{s}_2$ are decoded by carrying out [10]

$$\hat{s}_1 = \arg \min_{s_1 \in \Omega^2} \left| \left[\sum_{j=1}^{M} \left(y_1^j h_{1,j}^* + (y_2^j)^* h_{2,j} \right) \right] - s_1 \right|^2 + \left(-1 + \sum_{j=1}^{M} \sum_{i=1}^{2} |h_{i,j}|^2 \right) |s_1|^2$$

and,

$$\hat{s}_2 = \arg \min_{s_2 \in \Omega^2} \left| \left[\sum_{j=1}^{M} \left(y_1^j h_{2,j}^* - (y_2^j)^* h_{1,j} \right) \right] - s_2 \right|^2 + \left(-1 + \sum_{j=1}^{M} \sum_{i=1}^{2} |h_{i,j}|^2 \right) |s_2|^2.$$

Obviously, the complexity is linear with the constellation size since each symbol is being decoded separately. However, this decoding complexity can be further simplified as illustrated in the following example.

Example 2: Again, we consider the Alamouti OSTBC with $N = 2$ and $M = 1$. The system model is given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_{1,1} \\ h_{2,1} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{25}$$

Representing (25) in the real domain, we have

$$\begin{bmatrix} \Re(y_1) \\ \Im(y_1) \\ \Re(y_2) \\ \Im(y_2) \end{bmatrix} = \check{H} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \Re(v_1) \\ \Im(v_1) \\ \Re(v_2) \\ \Im(v_2) \end{bmatrix} \tag{26}$$

where $x_1 = \Re(s_1)$, $x_2 = \Im(s_1)$, $x_3 = \Re(s_2)$, $x_4 = \Im(s_2)$ and

$$\check{H} = \begin{bmatrix} h_1 & -h_2 & h_3 & -h_4 \\ h_2 & h_1 & h_4 & h_3 \\ h_3 & h_4 & -h_1 & -h_2 \\ h_4 & -h_3 & -h_2 & h_1 \end{bmatrix}.$$

Note that the columns of the channel matrix $\check{H}$ are orthogonal. Thus, applying QR decomposition to (26) and pre-multiplying both sides with Q^H produce

$$\begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \\ \bar{y}_4 \end{bmatrix} = \begin{bmatrix} r_{1,1} & 0 & 0 & 0 \\ 0 & r_{2,2} & 0 & 0 \\ 0 & 0 & r_{3,3} & 0 \\ 0 & 0 & 0 & r_{4,4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \bar{v}_1 \\ \bar{v}_2 \\ \bar{v}_3 \\ \bar{v}_4 \end{bmatrix}.$$

The MLD finds $\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4$ by performing

$$\hat{x}_i = \arg\min_{x_i \in \Omega} |\bar{y}_i - r_{i,i} x_i|^2 \quad (27)$$

for $i = 1, 2, 3, 4$, and the decoded message is $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)^T$.

Generally, we observe that the proposed algorithm produces $2K$ parallel 1×1 real-valued subsystems which results in a simplified MLD problem that can be solved in a parallel fashion to obtain the optimal solution while substantially reducing the overall decoding complexity.

V. Simulation Results

Figure 1 shows the performance of SD using the new lattice representation proposed in Section II versus conventional SD, for 2×2, 4×4, and 6×6 cases using 16-QAM modulation. For 2×2 the proposed algorithm achieves exactly the same performance as conventional SD, but with much smaller complexity as shown in Figure 2. However, there is < 1 dB performance loss in the proposed 4×4 and 6×6 compared to the conventional. This loss is due to the k-best criteria adoption at a certain level of the tree as well as applying the quantization process at the low levels of the tree (refer to [18] for adaptive k-best and quantization techniques). From Figure 2, it is clear that the proposed algorithm reduces the complexity by 80% for the 2×2 case, and 50% for both the 4×4 and 6×6 systems.

We also provide simulation results for the proposed algorithm in Section III. We denote the algorithm by PR and compare its performance to the performance of optimal conventional ML detection. Figure 3 shows simulation results for the transmission of 4 bits/s/Hz with $N = 4$, $M = 1$ and 16-QAM using the quasi-orthogonal code defined in (11). We compare PR with conventional ML detection with and without rotation. PR achieves full diversity through the rotation matrix G, while ML achieves it by replacing the transmitted symbols s_3, s_4 with $s_3 e^{j\phi}, s_4 e^{j\phi}$ in (11) with $\phi = \pi/4$ (i.e., choosing half of the transmitted symbols from a rotated constellation set $e^{j\phi}\mathcal{A}$ as discussed earlier) [6]. PR achieves the same performance as conventional ML, with a substantial complexity gain.

The complexity is measured in terms of the number of real multiplications required to decode one block of transmitted symbols as a function of the constellation size L. A complex multiplication is equivalent to 4 real multiplications C_M and 2 real additions C_A, while a complex addition is equivalent to 2 real additions. The complexity of the decoding algorithm proposed in [23] is in the order $\mathcal{O}(L^{N/4})$ which is equal to the complexity of the algorithm presented in [26] and will be denoted $\mathcal{C}_{[23],[26]}$ in the sequel. In Table I, we give a comparison between ML, algorithms in [23] and [26], and PR in terms of the number of real multiplications and real additions considering $N = 4$ for different constellation sizes. The number of multiplications and additions shown include the computation of QR and $\bar{y} = Q^H \tilde{y}$.

Apparently, the complexity gain obtained by PR is substantial and exceeds 95% compared to conventional ML. It is important to emphasize the fact that the complexity reduction, as shown in Table I, becomes greater as N or L is larger.

Finally, we consider the proposed decoding algorithm pre-

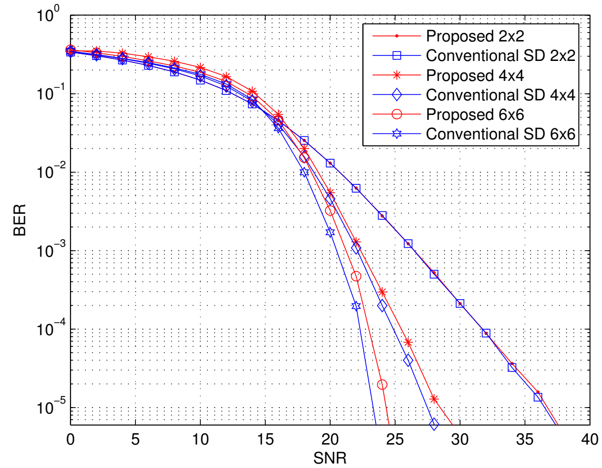

Fig. 1. BER vs SNR for the proposed and conventional SD over a 2×2, 4×4, and 6×6 MIMO flat fading channel using 16-QAM modulation.

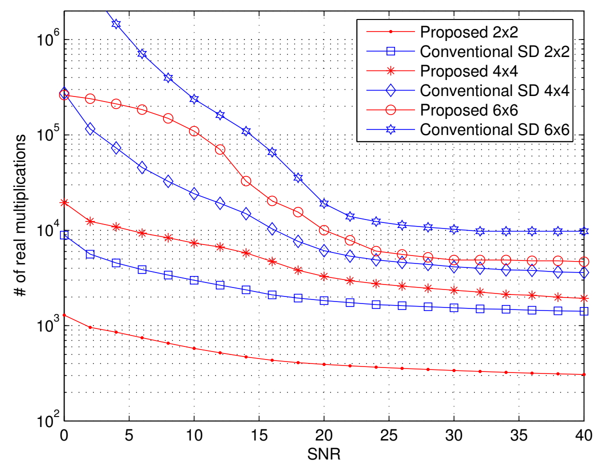

Fig. 2. Number of real multiplications vs SNR for the proposed and conventional SD over a 2×2, 4×4, and 6×6 MIMO flat fading channel using 16-QAM modulation.

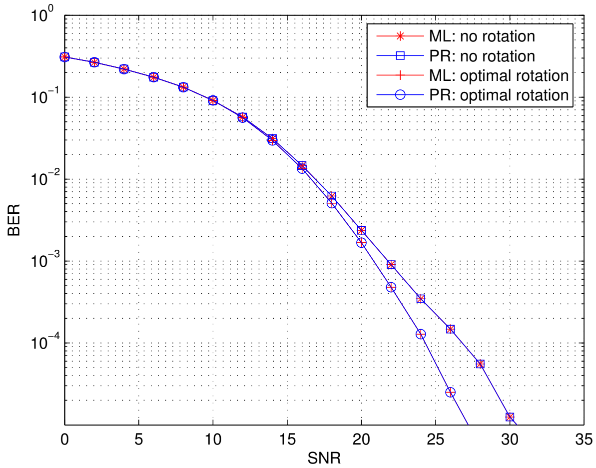

Fig. 3. BER vs SNR for PR and conventional ML for 4×1 system employing QOSTBC and 16-QAM.

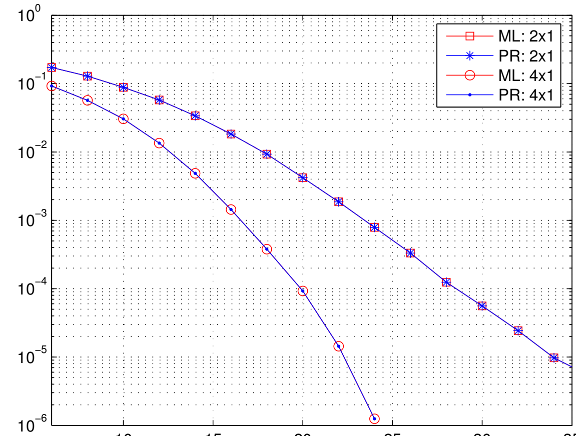

Fig. 4. BER vs SNR for PR and conventional ML for 2×1, and 4×1 systems employing OSTBC and 16-QAM.

TABLE I

OF REAL MULTIPLICATIONS AND REAL ADDITIONS VS L USING QOSTBC WITH $N = 4$

	L	4	16	64	256
C_M	ML	1536	24576	393216	6291456
	[23], [26]	512	2048	8192	32768
	PR	432	1104	3792	14544
C_A	ML	1376	22016	352256	5636096
	[23], [26]	432	1728	6912	27648
	PR	288	720	2448	9360

sented in Section IV for OSTBC . Table II shows a complexity comparison for $N = 2$, $M = 1$ considering the Alamouti OSTBC defined in (19). Moreover, in Table III, we show the same comparison for $N = 4$, $M = 1$ using the OSTBC $\mathcal{G}_4$ defined in [27]. We explained in Section IV that the proposed algorithm provides the optimal ML performance while reducing the complexity significantly. To verify this by simulation, we provide simulation results for OSTBCs employing 2×1, 4×1 and 16-QAM in Figure 4. Apparently, the complexity gain obtained is for free since there is no performance loss.

TABLE II

OF REAL MULTIPLICATIONS AND REAL ADDITIONS VS L USING ALAMOUTI CODE

	L	4	16	64	256
C_M	ML	224	896	3584	14336
	PR	80	112	176	304
C_A	ML	176	704	2816	11264
	PR	31	47	79	143

TABLE III

OF REAL MULTIPLICATIONS AND REAL ADDITIONS VS L USING OSTBC $\mathcal{G}_4$

	L	4	16	64	256
C_M	ML	960	3840	15360	61440
	PR	248	312	440	696
C_A	ML	864	3456	13824	55296
	PR	159	191	255	383

VI. CONCLUSIONS

In this paper, we discussed three applications of the QR decomposition algorithm to decoding in MIMO systems. We proposed a new lattice representation for sphere decoding. This new structure has the major impact of carrying out the decoding of the real and imaginary parts of every transmitted complex symbol independently from each other, thus allowing for a parallel detection. This, in turn, reduces the number of computations required at the receiver and consequently reduces the overall decoding complexity. In the other two applications, on the other hand, an efficient ML decoding algorithm based on QR decomposition of the channel matrix is proposed for quasi-orthogonal space-time block codes and orthogonal space-time block codes. The performance is shown to be optimal while reducing the decoding complexity significantly compared to conventional ML. Furthermore, the proposed algorithm reduces the decoding computational complexity from $\mathcal{O}(L^{N/2})$ for conventional MLD to $\mathcal{O}(L)$ for systems employing QOSTBCs and from $\mathcal{O}(L)$ for conventional MLD to $\mathcal{O}(\sqrt{L})$ for those employing OSTBCs, and consequently the complexity gain becomes grater as the constellation size is larger.

REFERENCES

[1] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere

decoding algorithm," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1566–1577, July 2005.

[2] E. Zimmermann, W. Rave, and G. Fettweis, "On the Complexity of Sphere Decoding," in Proc. International Symp. on Wireless Pers. Multimedia Commun., Abano Terme, Italy, Sep. 2004.

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

[4] E. Viterbo and J. Boutros, "A Universal Lattice Code Decoder for Fading Channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.

[5] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and A. Calderbank, "Space-Time Processing for Broadband Wireless Access," *IEEE Communications Magazine*, vol. 40, pp. 136–142, Sep. 2002.

[6] H. Jafarkhani, "Space-Time Coding: Theory and Practice," 2005.

[7] ——, "A Quasi-Orthogonal Space-Time Block Code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[8] C. Papadias and G. Foschini, "Capacity-Approaching Space-Time Codes for Systems Employing Four Transmitter Antennas," *IEEE Trans. Inf. Theory*, vol. 49, pp. 726–732, Mar. 2003.

[9] L. Xian and H. Liu, "Rate-One Space-Time Block Codes with Full Diversity," *IEEE Trans. Commun.*, vol. 53, pp. 1986–1990, Dec. 2005.

[10] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[11] N. Sharma and C. Papadias, "Improved Quasi-Orthogonal Codes Through Constellation Rotation," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 332–335, Mar. 2003.

[12] O. Tirkkonen, "Optimizing Space-Time Block Codes by Constellation Rotations," *Finish Wireless Communications Workshop (FWWC)*, pp. 59–60, Oct. 2001.

[13] J.Boutros, E. Viterbo, C. Rastello, and J. Belfiore, "Good Lattice Constellation for Both Rayleigh Fading and Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 42, pp. 502–518, Mar. 1996.

[14] J.Boutros and E. Viterbo, "Signal Space Diversity: A Power- and Bandwidth-Efficient Diversity Technique for the Rayleigh Fading Channel," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1453–1467, Jul. 1998.

[15] W. Su and X. Xia, "Signal Constellations for Quasi-Orthogonal Space-Time Block Codes With Full Diversity," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2331–2347, Oct. 2004.

[16] A. Peng, Kim, and S. Yousefi, "Low-Complexity Sphere Decoding Algorithm for Quasi-Orthogonal Space-Time Block Codes," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 377–382, Mar. 2006.

[17] S. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE J. on Selected Areas in Communications*, vol. 16, pp. 1451–1458, Oct. 1998.

[18] L. Azzam and E. Ayanoglu, "Reduced Complexity Sphere Decoding for Square QAM via a New Lattice Representation," *IEEE GLOBECOM*, Nov. 2007.

[19] D. Garrett, L. Davis, S. ten Brink, B. Hochwald, and G. Knagge, "Silicon complexity for maximum likelihood MIMO detection using spherical decoding," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1544–1552, Sep. 2004.

[20] M. O. Damen, H. E. Gamal, and G. Caire, "On Maximum-Likelihood Detection and the Search For the Closest Lattice Point," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2389–2402, Oct. 2003.

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

[22] M. Le, V. Pham, L. Mai, and G. Yoon, "Low-Complexity Maximum-Likelihood Decoder for Four-Transmit-Antenna Quasi-Othogonal Space-Time Block Code," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1817–1821, Nov. 2005.

[23] C. Yuen, Y. Guan, and T. Tjhung, "Quasi-Orthogonal STBC With Minimum Decoding Complexity," *IEEE Trans. Wireless Communications*, vol. 4, no. 5, pp. 2089–2094, Sep. 2005.

[24] L. Azzam and E. Ayanoglu, "Maximum Likelihood Detection of Quasi-Orthogonal Space-Time Block Codes: Analysis and Simplification," *to appear in IEEE ICC*, May 2008.

[25] G. Golub and C. V. Loan, "Matrix Computations," *3rd Edition, The John Hopkins University Press, Baltimore*, 1996.

[26] H. Lee, J.Cho, J. Kim, and I. Lee, "An Efficient Decoding Algorithm for STBC with Multi-dimensional Rotated Constellations," *IEEE ICC*, vol. 12, pp. 5558–5563, 2006.

[27] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results," *IEEE J. on Selected Areas in Communications*, vol. 17, pp. 451–460, Mar. 1999.