

Reduced Complexity Sphere Decoding for Square QAM via a New Lattice Representation

Luay Azzam and Ender Ayanoglu

Department of Electrical Engineering and Computer Science
University of California, Irvine
email: lazzam@uci.edu, ayanoglu@uci.edu

Abstract—Sphere decoding (SD) is a low complexity maximum likelihood (ML) detection algorithm, which has been adapted for different linear channels in digital communications. The complexity of SD has been shown to be exponential in some cases, and polynomial in others and under certain assumptions. The sphere radius and the number of nodes visited throughout the tree traversal search are the decisive factors for the complexity of the algorithm. The radius problem has been addressed and treated widely in the literature. In this paper, we propose a new structure for SD, which drastically reduces the overall complexity. The complexity is measured in terms of the floating point operations per second (FLOPS) and the number of nodes visited throughout the algorithm's tree search. This reduction in complexity is due to the ability of decoding the real and imaginary parts of every jointly detected symbol independently of each other, making use of the new lattice representation. We further show by simulations that the new approach achieves 80% reduction in the overall complexity compared to the conventional SD for a 2x2 system, and almost 50% reduction for the 4x4 and 6x6 cases, thus relaxing the requirements for hardware implementation.

I. INTRODUCTION

Minimizing the bit error rate (BER) and thus improving the performance is the main challenge of receiver design for multiple-input multiple-output (MIMO) systems. However, the performance improvements usually come at the cost of increased complexity in the receiver design. Assuming that the receiver has perfect knowledge of the channel H , different algorithms have been implemented to separate the data streams corresponding to N transmit antennas [1]. Among these algorithms, Maximum Likelihood detection (ML) is the optimum one. However, in MIMO systems, the ML problem becomes exponential in the number of possible constellation points making the algorithm unsuitable for practical purposes [2]. Sphere decoding, or the Fincke-Pohst algorithm [3], on the other hand, reduces the computational complexity for the class of computationally hard combinatorial problems that arise in ML detection problems [4]-[5].

Complexity reduction techniques for SD have been proposed in the literature. Among these techniques, the increased radius search (IRS) [6] and the improved increasing radius search (IIRS) [7] suggested improving SD complexity efficiency by making a good choice of the sphere radius, trying to reduce the number of candidates in the search space.

In this paper we improve the SD complexity efficiency by

reducing the number of FLOPS required by the SD algorithm keeping in mind the importance of choosing a radius. The radius d should not be too small to result in an empty sphere and thus restarting the search, and at the same time, it should not be too large to increase the number of lattice points to be searched. We use the formula presented in [8] for the radius, which is $d^2 = 2\sigma^2 N$, where N is the problem dimension and σ^2 is the noise variance. The reduction of the number of FLOPS is accomplished by introducing a new and proper lattice representation, as well as incorporating quantization at a certain level of the SD search. It is also important to mention that searching the lattice points using this new formulation can be performed in parallel, since the new proposed structure in this paper enables decoding the real and imaginary parts of each symbol independently and at the same time.

The remainder of this paper is organized as follows: In Section II, a problem definition is introduced and a brief review of the conventional SD algorithm is presented. In Section III, we propose the new lattice representation and perform the mathematical derivations for complexity reduction. Performance and complexity comparisons for different number of antennas or modulation schemes are included in Section IV. Finally, we conclude the paper in Section V.

II. PROBLEM DEFINITION AND THE CONVENTIONAL SPHERE DECODER

Consider a MIMO system with N transmit and M receive antennas. The received signal at each instant of time is given by

$$y = Hs + v \quad (1)$$

where $y \in \mathbb{C}^M$, $H \in \mathbb{C}^{M \times N}$ is the channel matrix, $s \in \mathbb{C}^N$ is an N dimensional transmitted complex vector whose entries have real and imaginary parts that are integers, $v \in \mathbb{C}^M$ 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 (quadrature amplitude modulation) 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 carried out as

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

Solving (2) becomes impractical for large constellations and large N . Therefore, instead of searching the whole space defined by all combinations drawn by the set Ω , SD solves this problem by searching only over those lattice points or combinations that lie inside a sphere centered around the received vector y and of radius d . Introducing this constraint on (2) will change the problem to

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

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

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

where $\Re\{y\}$ and $\Im\{y\}$ denote the real and imaginary parts of y , respectively [1], [4], [8]-[9]. Assuming $N = M$ in the sequel, 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

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

where $\bar{y} = Q^H y$. Let $R = [r_{i,j}]_{2N \times 2N}$ and note that R is upper triangular. Now to solve (5), the SD algorithm constructs a tree as shown in Figure 1, where the branches coming out

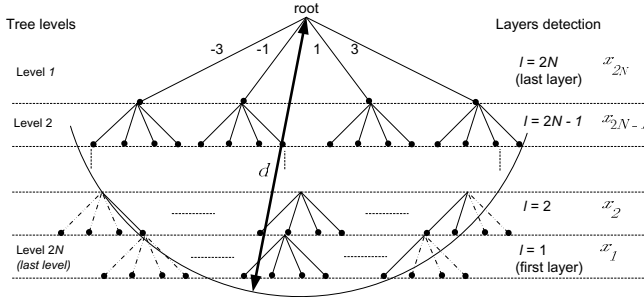


Fig. 1. Tree search example for a 16-QAM showing sphere radius, tree levels and detection layers.

of each node correspond to the elements drawn by the set Ω . It then executes the decoding process starting with the last layer ($l = 2N$) which matches the first level in the tree, calculating the partial metric $\|\Im\{\bar{y}_N\} - r_{2N,2N}\Im\{s_N\}\|^2$, and working its way up in a similar way to the successive interference cancellation technique [10], until decoding the first layer by calculating the corresponding partial metric $\|\Re\{\bar{y}_1\} - r_{1,1}\Re\{s_1\} + \dots + \Re\{\bar{y}_N\} - r_{1,N}\Re\{s_N\} + \Im\{\bar{y}_1\} - r_{1,N+1}\Im\{s_1\} + \dots + \Im\{\bar{y}_N\} - r_{1,2N}\Im\{s_N\}\|^2$. The summation of all partial metrics along the path of a node starting from the root constitutes the weight of that node. If that weight exceeds the square of the sphere radius d^2 , the algorithm prunes the corresponding branch, declaring it as an improbable way to a candidate solution. In other words, all nodes that lead to a solution that is outside the sphere are pruned at some level of

the tree. These branches are shown as dashed lines in Figure 1. Whenever a valid lattice point at the bottom level of the tree is found within 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. Finally, the leaf with the lowest weight will be the survivor one, and the path along the tree from the root to that leaf represents the estimated solution \hat{s} .

To this end, it is important to emphasize the fact that the complexity of this algorithm, although it is much lower than ML detection, is exponential with N at low SNR, and is directly related to the choice of the radius d , as well as the number of floating point operations taking place at every tree node inside the sphere.

III. NEW LATTICE REPRESENTATION

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

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

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

Obviously, SD algorithm starts from the upper level in the tree ($l = 2N$), traversing down one level at a time, and computing the weight for one or more nodes (depending on the search strategy adopted, i.e., depth-first, breadth-first, or other reported techniques in the literature) until finding a candidate solution at the bottom level of the tree ($l = 1$). According to this representation, it is impossible, for instance, to calculate $\sum_{k=l}^{2N} r_{l,k}x_k$ for a node in level ($l = 2N - 1$) 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 $j = l + 1, \dots, 2N$. Secondly, there is no room for parallel computations since the structure of the tree search is sequential.

The main contribution in this paper is that we relax the tree search structure making it more flexible for parallelism, and at the same time reducing the number of computations required at each node by making the decoding of every two adjacent levels in the tree totally independent of each other.

We start by reshaping the channel matrix representation given in (4) in the following form:

$$\tilde{H} = \begin{bmatrix} \Re(H_{1,1}) & -\Im(H_{1,1}) & \cdots & \Re(H_{1,N}) & -\Im(H_{1,N}) \\ \Im(H_{1,1}) & \Re(H_{1,1}) & \cdots & \Im(H_{1,N}) & \Re(H_{1,N}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Re(H_{N,1}) & -\Im(H_{N,1}) & \cdots & \Re(H_{N,N}) & -\Im(H_{N,N}) \\ \Im(H_{N,1}) & \Re(H_{N,1}) & \cdots & \Im(H_{N,N}) & \Re(H_{N,N}) \end{bmatrix} \quad (7)$$

where $H_{m,n}$ is the i.i.d. complex path gain from transmit antenna n to receive antenna m . By looking attentively at the columns of \tilde{H} starting from the left hand side, and defining each pair of columns as one set, we observe 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 received symbols to the following form

$$\hat{y} = [\Re(\hat{y}_1) \quad \Im(\hat{y}_1) \quad \cdots \quad \Re(\hat{y}_N) \quad \Im(\hat{y}_N)]^T. \quad (8)$$

This means that the first and second levels of the search tree (see Figure 1) correspond to the real and imaginary parts of s_N , unlike conventional SD, where those levels correspond to the imaginary part of s_N and s_{N-1} respectively. The new 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 $k = 1, 3, \dots, 2N-1$ in the upper triangular matrix R become zero. The locations of these zeros are very important since they introduce orthogonality between the real and imaginary parts of every detected symbol.

In the following, we will prove that the QR decomposition of \tilde{H} introduces the aforementioned zeros. There are several methods for computing the QR decomposition, we will do so by means of the Gram-Schmidt algorithm.

Proof: Let

$$\tilde{H} = [\tilde{\mathbf{h}}_1 \quad \tilde{\mathbf{h}}_2 \quad \cdots \quad \tilde{\mathbf{h}}_{2N}] \quad (9)$$

where $\tilde{\mathbf{h}}_k$ is the k th column of \tilde{H} . Recalling the Gram-Schmidt algorithm, we define

$$\mathbf{u}_1 = \tilde{\mathbf{h}}_1$$

and then,

$$\mathbf{u}_k = \tilde{\mathbf{h}}_k - \sum_{j=1}^{k-1} \phi_{\mathbf{u}_j} \tilde{\mathbf{h}}_j \quad \text{for } k = 2, 3, \dots, 2N.$$

where $\phi_{\mathbf{u}_j} \tilde{\mathbf{h}}_k$ is the projection of vector $\tilde{\mathbf{h}}_k$ onto \mathbf{u}_j defined by

$$\phi_{\mathbf{u}_j} \tilde{\mathbf{h}}_k = \frac{\langle \tilde{\mathbf{h}}_k, \tilde{\mathbf{u}}_j \rangle}{\langle \tilde{\mathbf{u}}_j, \tilde{\mathbf{u}}_j \rangle} \tilde{\mathbf{u}}_j \quad (10)$$

and $\mathbf{e}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$ for $k = 1, 2, \dots, 2N$. Rearranging the equations

$$\tilde{\mathbf{h}}_1 = \mathbf{e}_1 \|\mathbf{u}_1\|$$

$$\tilde{\mathbf{h}}_2 = \phi_{\mathbf{u}_1} \tilde{\mathbf{h}}_2 + \mathbf{e}_2 \|\mathbf{u}_2\|$$

$$\tilde{\mathbf{h}}_3 = \phi_{\mathbf{u}_1} \tilde{\mathbf{h}}_3 + \phi_{\mathbf{u}_2} \tilde{\mathbf{h}}_3 + \mathbf{e}_3 \|\mathbf{u}_3\|$$

\vdots

$$\tilde{\mathbf{h}}_k = \sum_{j=1}^{k-1} \phi_{\mathbf{u}_j} \tilde{\mathbf{h}}_k + \mathbf{e}_k \|\mathbf{u}_k\|.$$

Now, writing $Q = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3 \quad \cdots \quad \mathbf{e}_n]$, we have these equations in the matrix form as

$$Q \begin{bmatrix} \|\mathbf{u}_1\| & \langle \mathbf{e}_1, \tilde{\mathbf{h}}_2 \rangle & \langle \mathbf{e}_1, \tilde{\mathbf{h}}_3 \rangle & \cdots & \langle \mathbf{e}_1, \tilde{\mathbf{h}}_n \rangle \\ 0 & \|\mathbf{u}_2\| & \langle \mathbf{e}_2, \tilde{\mathbf{h}}_3 \rangle & \cdots & \langle \mathbf{e}_2, \tilde{\mathbf{h}}_n \rangle \\ 0 & 0 & \|\mathbf{u}_3\| & \cdots & \langle \mathbf{e}_3, \tilde{\mathbf{h}}_n \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \|\mathbf{u}_n\| \end{bmatrix}. \quad (11)$$

Obviously, the matrix Q is unitary, and the matrix on the right is the upper triangular R matrix of the QR decomposition.

Now our task is to show that the terms $\langle \mathbf{e}_k, \tilde{\mathbf{h}}_{k+1} \rangle$ are zero for $k = 1, 3, \dots, 2N-1$. Three observations conclude the proof.

First, since $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_{k+1}$ are orthogonal for $k = 1, 3, \dots, 2N-1$, then $\phi_{\mathbf{u}_k} \tilde{\mathbf{h}}_{k+1} = \phi_{\mathbf{u}_{k+1}} \tilde{\mathbf{h}}_k = 0$ for same k .

Second, the projection of \mathbf{u}_m for $m = 1, 3, \dots, k-2$ on the columns $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_{k+1}$ is equal to the projection of \mathbf{u}_{m+1} on the columns $\tilde{\mathbf{h}}_{k+1}$ and $-\tilde{\mathbf{h}}_k$ respectively. To formalize this:

$$\begin{aligned} \langle \mathbf{u}_m, \tilde{\mathbf{h}}_k \rangle &= \langle \mathbf{u}_{m+1}, \tilde{\mathbf{h}}_{k+1} \rangle \\ \langle \mathbf{u}_m, \tilde{\mathbf{h}}_{k+1} \rangle &= -\langle \mathbf{u}_{m+1}, \tilde{\mathbf{h}}_k \rangle \end{aligned} \quad (12)$$

for $k = 1, 3, \dots, 2N-1$ and $m = 1, 3, \dots, k-2$. This property becomes obvious by using the first observation and revisiting the special structure of (7).

Third, making use of the first two observations, and noting that $\|\tilde{\mathbf{h}}_k\| = \|\tilde{\mathbf{h}}_{k+1}\|$ for $k = 1, 3, \dots, 2N-1$, it can be easily shown that $\|\mathbf{u}_k\| = \|\mathbf{u}_{k+1}\|$ for the same k .

Then,

$$\begin{aligned} \langle \mathbf{e}_k, \tilde{\mathbf{h}}_{k+1} \rangle &= \left\langle \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}, \tilde{\mathbf{h}}_{k+1} \right\rangle \\ &= \frac{1}{\|\mathbf{u}_k\|} \left\langle \tilde{\mathbf{h}}_k - \sum_{j=1}^{k-1} \phi_{\mathbf{u}_j} \tilde{\mathbf{h}}_j, \tilde{\mathbf{h}}_{k+1} \right\rangle \\ &= \frac{1}{\|\mathbf{u}_k\|} \left(\langle \tilde{\mathbf{h}}_k, \tilde{\mathbf{h}}_{k+1} \rangle - \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_1 \rangle \langle \mathbf{u}_1, \tilde{\mathbf{h}}_{k+1} \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} - \right. \\ &\quad \left. \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_2 \rangle \langle \mathbf{u}_2, \tilde{\mathbf{h}}_{k+1} \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} - \cdots - \right. \\ &\quad \left. \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_{k-2} \rangle \langle \mathbf{u}_{k-2}, \tilde{\mathbf{h}}_{k+1} \rangle}{\langle \mathbf{u}_{k-2}, \mathbf{u}_{k-2} \rangle} - \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_{k-1} \rangle \langle \mathbf{u}_{k-1}, \tilde{\mathbf{h}}_{k+1} \rangle}{\langle \mathbf{u}_{k-1}, \mathbf{u}_{k-1} \rangle} \right) \end{aligned} \quad (13)$$

Now, applying the above observations to (13), we get

$$\begin{aligned} \langle \mathbf{e}_k, \tilde{\mathbf{h}}_{k+1} \rangle &= \frac{1}{\|\mathbf{u}_k\|} \left(0 - \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_1 \rangle \langle \mathbf{u}_1, \tilde{\mathbf{h}}_{k+1} \rangle}{\|\mathbf{u}_1\|^2} - \right. \\ &\quad \left. - \frac{\langle \mathbf{u}_1, \tilde{\mathbf{h}}_{k+1} \rangle \langle \tilde{\mathbf{h}}_k, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} - \cdots - \right. \\ &\quad \left. \frac{\langle \tilde{\mathbf{h}}_k, \mathbf{u}_{k-2} \rangle \langle \mathbf{u}_{k-2}, \tilde{\mathbf{h}}_{k+1} \rangle}{\|\mathbf{u}_{k-2}\|^2} - \frac{-\langle \mathbf{u}_{k-2}, \tilde{\mathbf{h}}_{k+1} \rangle \langle \tilde{\mathbf{h}}_k, \mathbf{u}_{k-2} \rangle}{\|\mathbf{u}_{k-2}\|^2} \right) \\ &= 0. \end{aligned}$$

This concludes the proof. \blacksquare

In this context, the SD algorithm executes in the following way. First, the partial metric weight $|\hat{y}_{2N} - r_{2N,2N} \hat{x}_{2N}|^2$ for the μ nodes in the first level of the tree is computed, where μ is the number of elements in Ω . This metric is then checked against the specified sphere radius d^2 . If the weight at any node is greater than the sphere radius, then the corresponding branch is pruned. Otherwise, the metric value is saved for the next step. At the same time, another set of μ partial metric computations of the form $|\hat{y}_{2N-1} - r_{2N-1,2N-1} \hat{x}_{2N-1}|^2$ take place at the second level, since these two levels are independent as proved above. These metrics are checked against d^2 in a similar way to that done in the above level. The weights of the survivor nodes from both levels are summed up and the summation is checked against the sphere constraint, ending up

with a set of survivor \hat{s}_N symbols. Secondly, the estimation of the remaining \hat{x}_{2N-2} or \hat{s}_{N-1} symbols is done by quantization to the nearest constellation element in Ω . In other words, the values of $\hat{x}_{2N-2}, \dots, \hat{x}_1$ are calculated recursively for each combination of survived $\hat{x}_{2N}, \hat{x}_{2N-1}$, and the total weight given by $\|\hat{y} - R\hat{s}\|^2$ is determined at the bottom level of the tree for those leaves that obey the radius constraint. Finally, the leaf with the minimum weight is chosen to be the decoded message (\hat{s}). This can be formalized as

Step1:

for $i = 1$ to μ

$$\hat{x}_{2N} = \Omega(\mu)$$

$$\hat{x}_{2N-1} = \Omega(\mu)$$

if $|\hat{y}_{2N} - r_{2N,2N}\hat{x}_{2N}|^2 < d^2 \rightarrow$ add to survivor set 1

else \rightarrow prune branch

if $|\hat{y}_{2N-1} - r_{2N-1,2N-1}\hat{x}_{2N-1}|^2 < d^2 \rightarrow$ add to survivor set 2

else \rightarrow prune branch

next i

save all combinations of $\hat{x}_{2N}, \hat{x}_{2N-1}$ whose weight summations comply to the radius constraint. Denote the number of survivors at the end of this step by $\{\lambda\}$.

Step2: (for every combination in λ , calculate $\hat{x}_{2N-2}, \dots, \hat{x}_1$ recursively as shown below)

for $l = 2N - 2$ to 1 , step -2

set $v=l/2$, and calculate

$$e_l = \sum_{k=l+1}^{2N} r_{l,k}\hat{x}_k, \quad e_{l-1} = \sum_{k=l+1}^{2N} r_{l-1,k}\hat{x}_k$$

$$\hat{x}_l = \mathcal{U}\left(\frac{\Im(\hat{y}_v) - e_l}{r_{l,l}}\right), \quad \hat{x}_{l-1} = \mathcal{U}\left(\frac{\Re(\hat{y}_v) - e_{l-1}}{r_{l-1,l-1}}\right)$$

$$w_l(\hat{x}^{(l)}) = w_{l+1}(\hat{x}^{(l+1)}) + |\Im(\hat{y}_v) - e_l - \hat{x}_l r_{l,l}|^2$$

$$w_{l-1}(\hat{x}^{(l-1)}) = w_l(\hat{x}^{(l)}) + |\Re(\hat{y}_v) - e_{l-1} - \hat{x}_{l-1} r_{l-1,l-1}|^2$$

next l

where $\mathcal{U}(\cdot)$ quantizes the value (\cdot) to the closest element in the set Ω . The output of the above two steps is a set of candidate solutions $\hat{x}_{2N}, \dots, \hat{x}_1$ with corresponding weights.

Step3:

choose that set of $\hat{x}_{2N}, \dots, \hat{x}_1$ which has the lowest weight to be the detected message.

The above algorithm's complexity is linear with the number of antennas, and the performance is the same as conventional SD for MIMO systems for $M = N = 2$. However, there is a performance loss for systems with $N \geq 3$ (e.g., there is a 4 dB loss compared to conventional SD at a BER of 10^{-5} for a 4x4 system). This is mainly due to the use of quantization which takes place at all tree levels except the first two, and makes the estimation of \hat{x} s loose as we further traverse down in the tree. Thus, we introduce some rules in the middle levels of the tree when $N \geq 3$, while still using the above steps at the very first and very last two levels in the tree, in order to obtain near optimal performance (less than 1 dB loss), sticking with a complexity that is very much small compared to the

conventional SD. A brief discussion on how to specify these rules are proposed in Section IV.

IV. SIMULATION RESULTS

We have considered 2x2, 4x4, and 6x6 cases using 16-QAM and 64-QAM modulation schemes. As mentioned in the previous section, we introduce some rules in the middle levels of the tree when $N \geq 3$, which are similar to the well-known k -best rules [11] but with different values for k at different levels of the tree. Therefore, in our simulations for the 4x4 and 6x6 cases, we executed the algorithm in the following way:

For the 4x4 system, the first two levels of the tree which correspond to the imaginary and real parts of the symbol \hat{s}_4 are treated the same way as explained in Step 1 of the algorithm. For each survivor \hat{s}_4 , the weight for all different μ^2 possibilities of \hat{s}_3 is calculated, and those weights w_l that violate the radius constraint d^2 are dismissed. The best 8 survivors, or in other words, those 8 \hat{s}_3 's that have lowest weights are kept for next steps while for the others the corresponding paths are pruned. In the third two levels, the same procedure performed in the previous step is applied and the best 8 \hat{s}_2 's are kept. Finally, a quantization process followed by an estimation of the transmitted message is carried out exactly the same way as in step 2 and 3 explained in the previous section. On the other hand, the 6x6 case has a similar approach but with different parameters. The first two levels are treated similarly as explained in step 1. For the 16-QAM (64-QAM) case, the best 16, 8, and 4 (32,32, and 16) survivors of \hat{s}_6 , \hat{s}_5 , and \hat{s}_4 respectively are kept in the middle levels until reaching the last four levels which are then processed by quantization, in order to obtain \hat{s}_2 and \hat{s}_1 .

Figure 2 shows the performance of the proposed algorithm versus conventional SD, for 2x2, 4x4, and 6x6 cases using 16-QAM modulation. For 2x2 the proposed algorithm achieves exactly the same performance as conventional SD, but with much smaller complexity as shown in Figure 3. However, there is < 1 dB performance loss in the proposed 4x4 and 6x6 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 as mentioned above. From Figure 3, it is clear that the proposed algorithm reduces the complexity by 80% for the 2x2 case, and 50% for both the 4x4 and 6x6 systems. Figures 4 and 5 show the performance and complexity curves for the 2x2, 4x4, and 6x6 cases, for 64-QAM modulation. Again, the performance is close to the conventional for the 2x2 case, and has less than 1 dB degradation loss for the 4x4 and 6x6 cases. The difference in the complexity for the proposed and conventional SD are within the same range as in the 16-QAM modulation case.

V. CONCLUSIONS

A simple and general lattice representation in the context of sphere decoding was proposed in this paper. The performance of the proposed structure was shown to be the same as the conventional SD for 2x2 systems while close to optimal (< 1

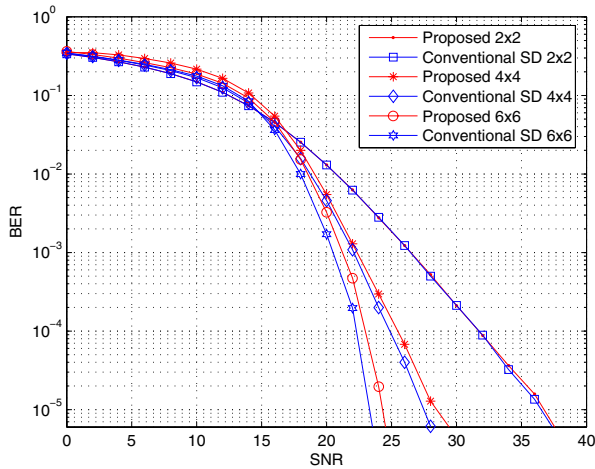


Fig. 2. BER vs SNR for the proposed and conventional SD over a 2x2, 4x4, and 6x6 MIMO flat fading channel using 16-QAM modulation.

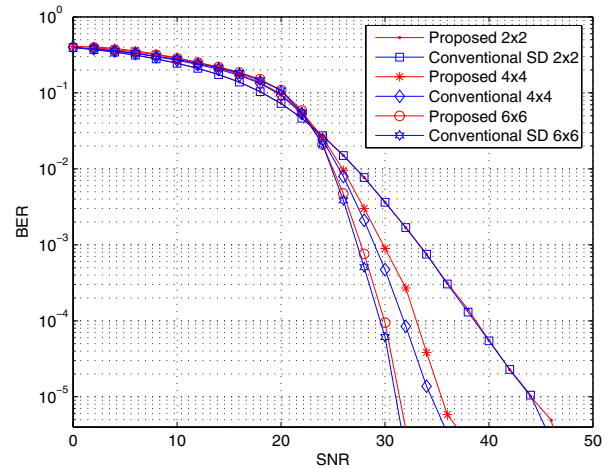


Fig. 4. BER vs SNR for the proposed and conventional SD over a 2x2, 4x4, and 6x6 MIMO flat fading channel using 64-QAM modulation.

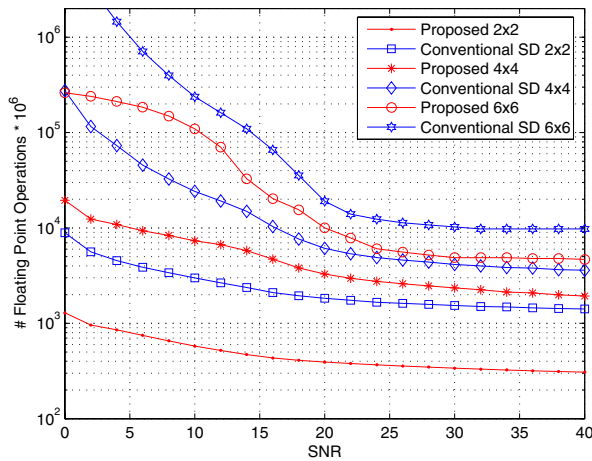


Fig. 3. Total number of floating point operations vs SNR for the proposed and conventional SD over a 2x2, 4x4 and 6x6 MIMO flat fading channel using 16-QAM modulation.

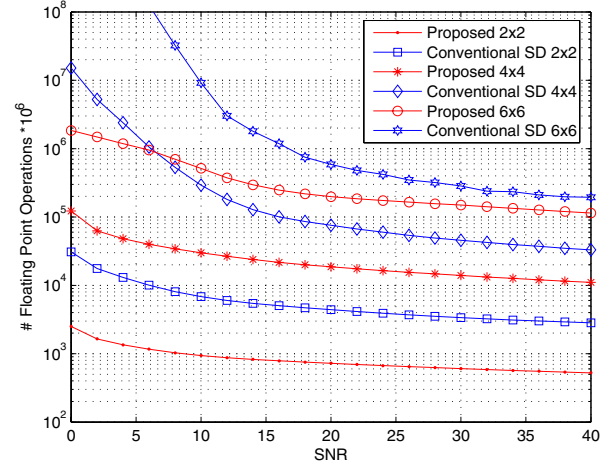


Fig. 5. Total number of floating point operations vs SNR for the proposed and conventional SD over a 2x2, 4x4, and 6x6 MIMO flat fading channel using 64-QAM modulation.

dB loss) in the 4x4 and 6x6 cases. A complexity reduction of 80% was attainable for the 2x2 case, and 50% for the 4x4 and 6x6 cases, compared to their correspondence for the conventional SD.

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] J. Jaldn and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474–1484, April 2005.
- [5] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806–2818, August 2005.
- [6] 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.
- [7] W. Zhao and G. Giannakis, "Sphere decoding algorithms with improved radius search," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1104–1109, July 2005.
- [8] G. Rekaya and J. Belfiore, "On the complexity of ML lattice decoders for decoding linear full-rate space-time codes," in *Proceedings 2003 IEEE International Symposium on Information Theory*, July 2003.
- [9] A. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," *IEEE ICC*, vol. 1, pp. 460–464, 2002.
- [10] B. Hassibi, "A fast square-root implementation for blast," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 2, Asilomar, USA, Nov. 2000, pp. 1255–1259.
- [11] K. Wong, C. Tsui, R.-K. Cheng, and W. Mow, "A VLSI architecture of a K-best lattice decoding algorithm for MIMO channels," in *Proc. IEEE ISCAS02*, vol. 3, pp. 273–276, 2002.