An Efficient Tree Search for Reduced Complexity Sphere Decoding

Luay Azzam and Ender Ayanoglu

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

***Abstract*— The complexity of sphere decoding (SD) has been widely studied due to the importance of this algorithm in obtaining the optimal Maximum Likelihood (ML) performance with lower complexity. In this paper, we propose a proper tree search traversal technique that reduces the overall SD computational complexity without sacrificing the performance. We exploit the similarity among the complex symbols in a square QAM lattice representation and rewrite the squared norm ML metric in a simpler form allowing significant reduction of the number of operations required to decode the transmitted symbols. We also show that this approach achieves $> 45\%$ complexity gain for systems employing 4-QAM, and that this gain becomes bigger as the constellation size is larger.**

I. Introduction

The use of multiple antennas at both the transmitter and the receiver provides high capacity gains without increasing the bandwidth or transmitted power. Thus, multiple-input multiple-output (MIMO) systems have attracted much attention and serious research interests. Improving the performance is the main challenge of receiver design. Therefore, a number of decoding algorithms with different complexity-performance tradeoffs can be used. For instance, linear detection methods such as a zero-forcing (ZF) or minimum mean squared error (MMSE) have linear complexity but provide suboptimal performance. Ordered successive interference cancelation decoders, on the other hand, such as the vertical Bell Laboratories layered space-time (V-BLAST) algorithm can also be used which provide suboptimal performance with higher complexity compared to ZF and MMSE [1]. The optimum detection method is the Maximum Likelihood (ML) detection. However, in MIMO systems, ML algorithm has an exponential complexity with the constellation size and the number of antennas [2]. Therefore, sphere decoding (SD) [3] was proposed as an alternative for ML that provides optimal performance with reduced computational complexity [4], [5].

Different techniques have been proposed in the literature to reduce the complexity of SD. Among these, the increased radius search (IRS) [6] and the improved increasing radius search (IIRS) [7] are noteworthy. These two techniques attempt tackling this complexity by making a proper choice of the sphere radius. Other methods such as the application of the K-best lattice decoder [8] or a combination of the SD and the K-best decoder [9] were used where the complexity reduction came at the cost of a BER performance degradation.

In this paper, we improve the SD complexity efficiency by reducing the number of computations required to obtain the ML optimal solution. This complexity reduction is accomplished by exploiting the similarity between the complex symbols in a square QAM lattice representation. The proposed technique is generic and can be used with depth-first and breadth-first sphere decoders. Moreover, we show that the conventional ML metric can be rewritten in a simpler form which can be also used for reduced complexity ML decoding. In our previous work presented in [10], we proposed a new lattice representation that enables decoding the real and imaginary parts of each complex symbol independently. In this work, we use that same lattice representation and show that a complexity gain of at least 45% is obtained without sacrificing the performance.

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. 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

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

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 where $\mathbb{Z}$ and $\mathbb{C}$ denote the sets of integers and complex numbers respectively.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

Assuming H is known at the receiver, the ML detection is carried out as

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

Solving (2) is well known to be NP-hard since a full search over the entire lattice space is performed [11]. Consequently, SD was proposed where (2) is being solved by searching only those lattice points that lie inside a sphere of radius d centered around the received vector y.

A frequently used solution for the QAM-modulated complex signal model given in (1) is to decompose the N-dimensional complex-valued problem into a $2N$-dimensional real-valued problem, which then can be written as

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

where $\Re\{y\}$ and $\Im\{y\}$ denote the real and imaginary parts of y.

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, (1) can be written as

$$\begin{aligned} y &= QRs + v \\ Q^H y &= Rs + Q^H v \\ \bar{y} &= Rs + \bar{v} \end{aligned} \quad (3)$$

where $\bar{v}$ and v have the same statistical properties since Q is unitary and so is Q^H. Then, SD solves

$$\hat{s} = \arg \min_{s \in \Omega^{2N}} ||\bar{y} - Rs||^2 < d^2 \quad (4)$$

The SD algorithm can be viewed as a pruning algorithm on a tree of depth $2N$, having branches that correspond to the elements drawn by the set Ω. This is shown in Figure 1 for a 16-QAM constellation. Note that pruned branches are shown as dashed.

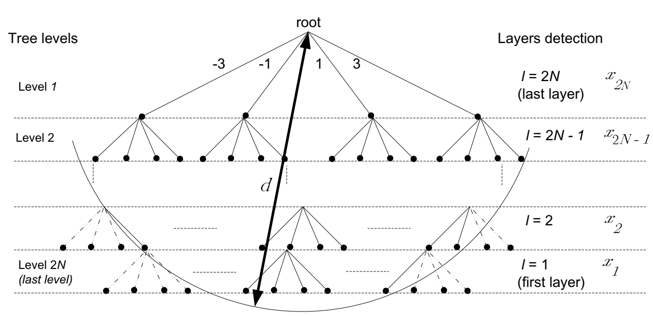

Fig. 1. Tree search example for a 16-QAM showing sphere radius, tree levels and detection layers.

Sphere decoding starts the search process from the root of the tree and works its way down along the branches until (4) is violated. This occurs when the total weight of a node exceeds the square of the sphere radius d^2. At this point, the corresponding branch is pruned and any path that passes through that node is declared as an improbable way to a candidate solution. Then, the algorithm backtracks and proceeds down a different branch. 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 path from the root to the leaf that is inside the sphere with the lowest weight is chosen to be the estimated solution $\hat{s}$.

III. Proposed Sphere Decoder

The complexity of SD is measured in terms of the number of operations required per visited node multiplied by the number of visited nodes through out the algorithm search [11]. This complexity can be reduced by either reducing the number of nodes to be visited or the number of operations to be carried out at each node or both. Making a good choice of the sphere radius to start the algorithm execution with, and thus reducing the number of visited nodes, has been widely studied in [6], [7] and the references therein. In this paper, we attempt to reduce the SD complexity by reducing the number of operations required at each node. We start by writing the node weight as [10]

$$\begin{aligned} w_l(x^{(l)}) &= w_p + w_{pw} \\ &= w_p + |\bar{y}_l - \sum_{k=l}^{2N} r_{l,k} x_k|^2 \end{aligned} \quad (5)$$

with $l = 2N, 2N-1, \ldots, 1$, w_p is the weight of the node's parent, w_{pw} is the partial weight of the node, 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.

This means that the total weight of a node is merely the weight of its parent summed with its partial weight represented by the second term in (5). This squared term can be simplified by exploiting the similarity between the symbols in a square QAM lattice structure. To make this clearer, let us consider a 16-QAM modulation scheme. Every node in the tree has 4 branches; each represents one of the real values given by $\Omega = \{-3, -1, 1, 3\}$. For a specific l, the computation of $|\bar{y}_l - \sum_{k=l}^{2N} r_{l,k} x_k|^2$ for $x_k = -3$ and $x_k = 3$ is very similar. Basically, all the terms are common except for the minus sign inside the summation. The same applies for $x_k = -1$ and $x_k = 1$. This similarity makes it possible to divide the set $\Omega = \{-3, -1, 1, 3\}$ into two smaller sets namely $\Omega_1 = \{-3, -1\}$ and $\Omega_2 = \{1, 3\}$, and enforce the algorithm to compute the weight of the nodes that correspond only to one set. These weights can then be reused to find the weight of the nodes in the other set. We provide the following example to explain this.

Example: Consider a MIMO system having $N = M = 2$ and employing 4-QAM modulation scheme. Then, SD constructs a tree with $2N = 4$ levels where the branches coming out from each node represent the real values in the set $\Omega = \{-1, 1\}$. This tree is shown in Figure 2. Now, using the real-valued lattice representation developed in [10], and applying the QR decomposition to the channel matrix, the

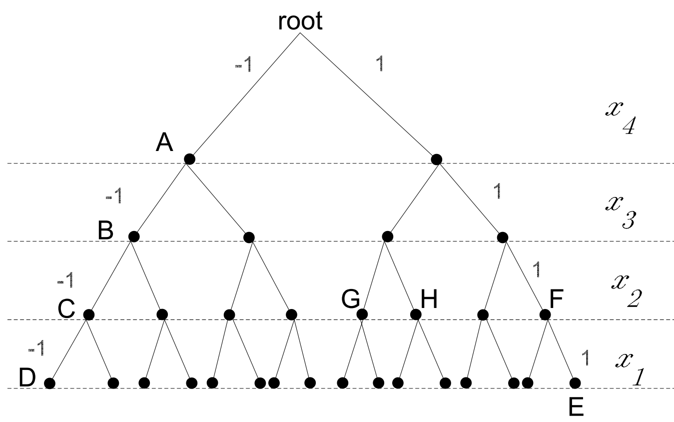

Fig. 2. Tree structure for 2×2 system employing 4-QAM.

input-output relation is given by

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

Proposed SD starts by dividing the set Ω into two smaller sets $\Omega_1 = \{-1\}$ and $\Omega_2 = \{1\}$. Then using (5), it calculates the weights of nodes A, B, C, and D which correspond to $x_k \in \Omega_1$ for $k = 1, 2, 3, 4$. For e.g., D represents the solution $(x_1, x_2, x_3, x_4) = (-1, -1, -1, -1)$. Now these basic nodes act as other parents for the nodes that are on the same tree level. For example, the node E has now two parents which are the original parent F and the new parent D. For level k in the tree, we denote all branches in that level which share the same property with their new parents in having the same value of x_k, by $\tilde{x}_k$, and denote the others which have a different values than their parent nodes by $\check{x}_k$. For e.g., G represents $\tilde{x}_k = -1$, whereas H represents $\check{x}_k = 1$. By setting up the above definitions, (5) is no longer used and the weight of any node in the tree can be found using a less computational expression given by

$$w_l(x^{(l)}) = w_p + w_{np} + 4\left(\bar{y}_l - \sum_{\tilde{x}_k \in \Omega_1} r_{l,k}|\tilde{x}_k|\right)\left(\sum_{\check{x}_k \in \Omega_2} r_{l,k}|\check{x}_k|\right) \quad (6)$$

where w_{np} is the weight of the new parent. Note that it is straight forward to obtain (6) by expanding the squared term in (5) and taking out the common terms between any node weight and its new parent weight.

Dealing only with the absolute values in the above expression reduces the number of required computations significantly. For 4-QAM, we have $|\tilde{x}_k| = |\check{x}_k|$, thus (6) can be rewritten as

$$w_l(x^{(l)}) = w_p + w_{np} + 4\left(\bar{y}_l - \sum_{\tilde{x}_k \in \Omega_1} r_{l,k}\right)\left(\sum_{\check{x}_k \in \Omega_2} r_{l,k}\right). \quad (7)$$

Then the partial weight of any node, say E and H can be simply found as

$$w_E = w_D + 4\bar{y}_1(r_{1,1} + r_{1,3} + r_{1,4})$$

and,

$$w_H = w_C + 4(\bar{y}_2 - r_{2,3})(r_{2,2} + r_{2,4}).$$

Calculating the partial weight of node E using the conventional SD requires 25 real multiplications and 24 real additions, whereas these numbers reduce to 4 real multiplications and 3 real additions using the proposed technique. These savings even becomes bigger as the constellation size is larger. This is because the number of tree nodes increases exponentially with the constellation size. For e.g., the number of nodes for a 4×4 system and 16-QAM is 87380. This number increases to $19 \cdot 10^6$ for a 4×4 system and 64-QAM.

To this end, it is worth mentioning here that in the proposed technique, we basically rewrite the conventional ML metric in a simpler form. As a result, this new form can be used for conventional ML decoding as well. It is also important to emphasize the fact that the proposed SD provides the same performance as the conventional SD with much lower computational complexity. This complexity gain becomes bigger as the constellation size and the number of antennas are larger.

IV. Simulation Results

We have considered 2×2, 4×4 systems using 4-QAM, 16-QAM and 64-QAM modulation schemes. Since the multiplications are the most expensive operations in terms of machine cycles compared to additions, the complexity is measured in terms of the number of real multiplications required to decode the transmitted complex symbols. We use the real-valued lattice representation presented in [10] for the conventional and proposed SD. This means that the complexity gain shown in all the figures below is on top of the gain obtained in [10]. We denote the conventional SD by Conv and the proposed SD by PR. The radius d should be chosen properly so that it is not 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 [12] for the radius, which is $d^2 = 2\sigma^2 N$, where N is the problem dimension and σ^2 is the noise variance.

Figure 3 shows the complexity curves for both algorithms using 4-QAM. For 2×2 system, the complexity gain is 45%. This gain increases up to reach 60% for the 4×4 case.

Similarly, Figures 4 and 5 show complexity comparison using the same configuration with 16-QAM, and 64-QAM respectively. Again, the proposed SD achieves high complexity reduction compared to conventional SD. The gain ranges from

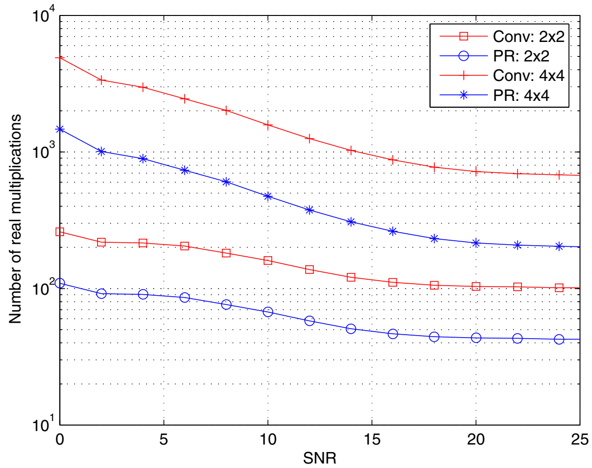

Fig. 3. # of real multiplications vs SNR for the proposed and conventional SD over a 2×2, and 4×4 MIMO flat fading channel using 4-QAM.

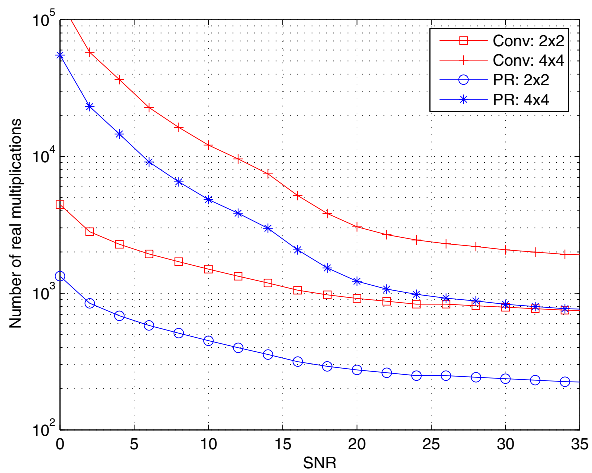

Fig. 4. # of real multiplications vs SNR for the proposed and conventional SD over a 2×2, and 4×4 MIMO flat fading channel using 16-QAM.

50% for 2×2 with 16-QAM, up to 65% for 4×4 and 64-QAM.

Finally, we should emphasize that these complexity gains are for free. This is because the proposed algorithm provides the same performance results as the conventional SD.

V. Conclusions

A simple and general tree search technique for sphere decoding is proposed in this paper. The performance of the proposed SD is the same as that of conventional SD. However, a significant complexity reduction compared to conventional SD is achieved. This complexity reduction is accomplished by exploiting the similarity between the complex symbols in a square QAM constellation, allowing for rewriting the conventional ML metric in a simpler form. This new form is also applicable to performing ML decoding with reduced

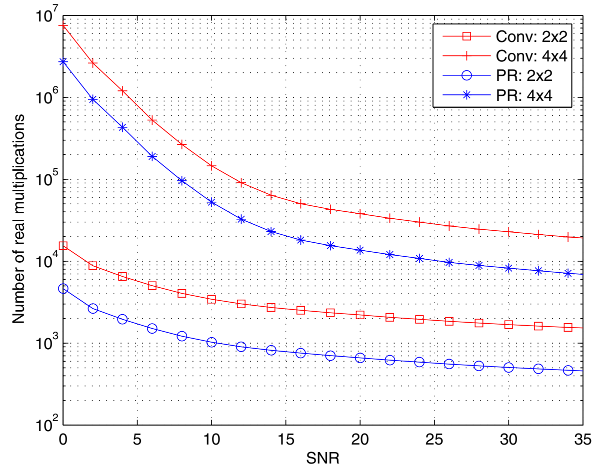

Fig. 5. # of real multiplications vs SNR for the proposed and conventional SD over a 2×2, and 4×4 MIMO flat fading channel using 64-QAM.

computational complexity. The complexity gain ranges between 45% and 65% depending on the number of antennas and constellation size being used. Simulation results are provided for 2×2 and 4×4 systems employing 4-QAM, 16-QAM, and 64-QAM. We show that the gain achieved by using the proposed SD becomes bigger 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] J. Jalden 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, Jul. 1999.

[7] W. Zhao and G. Giannakis, "Sphere Decoding Algorithms With Improved Radius Search," *IEEE Trans. Commun.*, vol. 53, pp. 1104–1109, Jul. 2005.

[8] 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.

[9] J. Tang, A. Tewfik, and K.K.Parhi, "Reduced Complexity Sphere Decoding and Application to interfering IEEE 802.15.3a piconets," *IEEE ICC*, vol. 5, pp. 2864–2868, June 2004.

[10] L. Azzam and E. Ayanoglu, "Reduced Complexity Sphere Decoding for Square QAM via a New Lattice Representation," *IEEE GLOBECOM*, pp. 4242–4246, Nov. 2007.

[11] B. Hassibi and H. Vikalo, "On The Expected Complexity of Integer Least-Squares Problems," *IEEE ICASSP*, pp. 1497–1500, 2002.

[12] 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.*, Jul. 2003.