Abstract—For MIMO data detection methods, sphere decoding (SD) algorithms have been shown to achieve similar performance to the maximum-likelihood detection (MLD) algorithm with acceptable computational complexity. Based on different search strategies, SD can be divided into two parts: depth-first search SD and breadth-first search SD. In this work, by combining the advantages of these two search schemes, we propose a novel hybrid sphere decoding to effectively reduce the complexity. Besides, a layer reordered method is adopted to improve the performance. Compared with other SD algorithms, the proposed one has not only good performance but also lower complexity.

Keywords—MIMO systems, sorted QR decomposition, sphere decoding, maximum-likelihood detection

I. INTRODUCTION

Since recent decade, wireless communications based on multiple-input multiple-output (MIMO) system [1] has received widespread attention. For MIMO data detection methods, maximum-likelihood detection (MLD) algorithm has the best performance of all the detection algorithms. However, the involved computational complexity is an exponential function of the signal constellation and the number of transmitter antennas, which is prohibitively high for practical applications. To date, many of data detection methods have been proposed to reach performance close to MLD with fewer operations, and which can be classified into three types. The first type is based on QR decomposition (QRD), like the sorted QR decomposition (SQRD) [2]; the second type is based on successive interference cancellation (SIC) combined with zero-forcing or minimum mean square error (MMSE) methods, like the vertical Bell Labs layered space time algorithm (V-BLAST); the third type are sequential decoders based on MLD, like the sphere decoding (SD) [3]-[8].

The SD algorithms utilize two different search schemes: the depth-first search scheme, like Fincke-Pohst SD (F-P SD) [3][4] and Schnorr-Euchner SD (S-E SD) [5]; and the breadth-first search scheme, like K-Best SD [6]-[8]. Both search schemes have advantages and disadvantages. In this paper, we detail their pros and cons and combine them in an effective way. As a result, a novel hybrid SD (HSD) is obtained, which has lower complexity than K-Best SD algorithms.

For most SD algorithms in literature, S-E SD has not been effectively implemented because its throughput at each layer is non-fixed. However, in our proposed HSD, by integrating K-Best SD, concept of S-E SD can be effectively exploited to reduce redundant operations, while the throughput rate can be maintained fixed at each decoding layer. Besides, the layer reordered method in [2] is adopted in the proposed HSD to achieve better performance. With the use of the layer reordered method before SD, probability of missing the ML solutions at early layers will be reduced. Overall, the performance can be improved.

This paper is organized as follows. Section II describes MIMO system model. In Section III, layer reordered SD algorithms including depth-first search SD and breadth-first search SD are introduced. In Section IV, we detail the proposed HSD algorithm. Simulation results and conclusion will be given in Section V and Section VI, respectively.

II. MIMO SYSTEMS

For the convenience of further discussion, some notations are defined first. Boldface denotes matrices or vectors; $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively; $(\cdot)^+$ denotes pseudo-inverse; $\| (\cdot) \|^2$ is two-norm of $(\cdot)$.

Let’s consider a spatial multiplexing MIMO system with $N_T$ transmit and $N_R \geq N_T$ receive antennas, and assume that channel state information (CSI) is perfectly known at the receiver. The discrete-time baseband received signal at any given time can be written as

$$\mathbf{r} = \mathbf{Hx} + \mathbf{n} \tag{1}$$

where $\mathbf{r} = [r_1,r_2,\cdots,r_{N_R}]^T$, $r_i \in \mathbb{C}$ (the infinite complex field), is the received signal vector; $\mathbf{x} = [x_1,x_2,\cdots,x_{N_T}]^T$, $x_i \in \Omega$, is the transmitted signal vector and $\Omega$ is a finite constellation set;
represents an additive white Gaussian noise (AWGN) at the $N_r$ receive antennas. The components of $\mathbf{n}$ are independent and identically distributed (i.i.d.) and $E\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{N_r}$, and the averaged transmit power of each antenna is normalized to one, i.e., $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_{N_r}$. $\mathbf{H}=[h_{ij}] \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, and $h_{ij}$ represents complex channel gain from the $j$-th transmit antenna to the $i$-th receive antenna.

III. LAYER REORDERED SPHERE DECODING

First of all, a kind of layer reordered method, SQRD [2], is introduced in this section. By combining the method with our proposed algorithm, we can effectively improve the decoding performance.

Performance of QR decomposition can be improved if the channel matrix $\mathbf{H}$ is preprocessed before QR decomposition. Columns of $\mathbf{H}$ can be reordered according to the norm of each column, so the signals with higher signal-to-noise ratio (SNR) are detected first. This can be done by multiplying $\mathbf{H}$ by a permutation matrix $\mathbf{P}$, i.e., $\mathbf{P}\mathbf{H} = \mathbf{Q}$. Applying SQRD to (1), one has

$$\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{\eta}$$

where $\mathbf{y} = \mathbf{Q}^H \mathbf{P}\mathbf{r}$ and $\mathbf{\eta} = \mathbf{Q}^H \mathbf{P}\mathbf{n}$. If we consider noise effect, $\mathbf{H}$ and $\mathbf{r}$ can be extended as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H} \\ \sigma_n \mathbf{I}_{N_r} \end{bmatrix}, \mathbf{r} = \begin{bmatrix} \mathbf{r} \\ 0_{N_t \times 1} \end{bmatrix}$$

where $\mathbf{H}$ is a $(N_r + N_t) \times N_t$ extended matrix and $\mathbf{r}$ is a $(N_r + N_t) \times 1$ extended vector. By replacing $\mathbf{H}$ and $\mathbf{r}$ in ZF-SQRD with (3), one can obtain the so called MMSE-SQRD method which achieves minimum mean square error (MMSE).

After the layer reordered process, SD with two different kinds of search methods is introduced below. The first is the depth-first search scheme including K-Best SD. For this reason, the K-Best SD is investigated. The detector visits all the children of those stored parents before it proceeds to the next layer. Instead of expanding every node at each layer, it only keeps $K$ candidates with smaller PEDs. The descriptions of the K-Best SD are as following:

(a-1) Let $n = N_r$, calculate the PED $P_{n,1}$ according to (4), and (5). Next, sort PEDs and select $K$ candidates with the smallest PEDs among $\Omega$ lattice points.

(a-2) $n = n - 1$.

(a-3) Calculate the branch cost $b_k^i(\mathbf{x})$, where $k \ 1 \leq k \leq K$ represents the candidate’s index according to the values of PEDs from last layer. Then, accumulate branch costs and get $P_k^i(\mathbf{x})$. Finally, sort PEDs and select $K$ candidates with the smallest PEDs among $K\Omega$ lattice points.

(a-4) Go to step (a-2) until $n = 1$.

The $K$ value determines the trade-off between computational complexity and performance. If $K$ is too large, many node weights will have to be computed and it may cost a lot redundant computations. On the contrary, if $K$ is too small, the ML solutions will most likely be missed. Therefore, there are many proposed ideas to improved K-Best SD [6]-[8].

IV. THE PROPOSED HYBRID SPHERE DECODING (HSD)

In section III, the mentioned two search schemes have their
own advantages and disadvantages. The searching criterion of S-E SD is depth-first searching according to the node with the smallest PED, therefore the nodes it finds at the first leaf node are most likely the confirmed nodes at the end of loop. But, if SNRs of the transmitted signals at former layers are too small to make the PED reliable, the final confirmed nodes may be altered, and that is the reason why SQRD is considered in the proposed method. If we assume that S-E and K-Best SD find the same nodes finally, the former usually searches these nodes faster than the latter. However, in S-E SD these nodes can not be immediately affirmed that they have the smallest PEDs until the loop ends, so to verify the results by comparing with other nodes’ PEDs is needed, and that is why its throughput is not fixed.

For K-Best SD, fixed throughput is its greatest advantage; nevertheless, it wastes too many operations at later layers. Many proposed schemes focus on this issue. For instance, [6] selects the different number of the held nodes at each layer, but there is no regular way to determine the $K$ value at each layer, therefore [6] is difficult for hardware implementation. [8] also uses the same concept like [6] and proposes an algorithm to determine the $K$ value, but parameters in the algorithm are based on the simulation results. [7] adds a estimated sphere radius to eliminate the unwanted nodes with larger PEDs.

The above mentioned methods need some stochastic assumptions or simulations to prune the nodes in K-Best SD, so a novel ideal is proposed without any uncertain factor. Based on the characteristics of S-E and K-Best SD, we integrate and utilize the advantages of both S-E and K-Best SD to achieve a more efficient algorithm with the layer reordered method, SQRD.

The concept of HSD is to search the transmitted signals speedily with the lowest operations subject to a fixed throughput. To begin with, how do S-E SD and K-Best SD effectively combine together? In order not to miss the ML solutions, we set a parameter $D$ to do K-Best SD at the first $D$ layers. As described in section III (B), at each layer, $K$ nodes and their corresponding PEDs defined in (4) and (5) are stored,

\[
P^k_D(x) = \sum_{i=N_{K}}^{D} b^k_i(x)
\]

where $1 \leq k \leq K$. When $D = N_{K}$, HSD is equivalent to K-Best SD. On the contrary, if $D < N_{K}$, only the preceding $D$ layers perform K-Best SD to save the possible ML solutions, and the following $N_{K} - D$ layers perform S-E SD with $K$ candidates constraint. The operational description of HSD at the later $N_{K} - D$ layers is summarized as follows.

\(\text{(b-1)}\) Let $k = 1$.

\(\text{(b-2)}\) From the given node with the PED $P^1_D(x)$, execute $N_{K} - D$ layers S-E SD with K-Best constraint (i.e., select $K$ candidates with the smallest PEDs at each layer) until it searches the leaf nodes, and the minimum PED is saved as the reference sphere radius $r^2 = P^1_D(x)$.

\(\text{(b-3)}\) $k = k + 1$ and $K' = K - 1$.

\(\text{(b-4)}\) If $P^k_D(x) < r^2$, execute $N_{K} - D$ layers S-E SD with K-Best constraint (i.e., select $K'$ candidates with the smallest PEDs at each layer). If $P^k_D(x) < r^2$, the reference sphere radius is replaced with $P^1_D(x)$.

\(\text{(b-5)}\) Go to (b-3) until $k = K$.

Fig. 1 shows an example of HSD for a $4 \times 4$ 4QAM MIMO system with $K = 3$. In Fig.1 (a), nodes $\{1,2,3\}$ are saved after K-Best SD, which is just performed once due to $D = 1$. Next, S-E SD is performed based on these three nodes. First, based on the first node, the nodes $\{4,10,16\}$ with the minimum PED 0.8 are saved. The steps are illustrated in (i)–(iii). For the second and the third nodes $\{2,3\}$, if their PEDs are smaller than the reference sphere radius 0.8, S-E SD is executed as well. At the end of loop,
the nodes with minimum PED, \{1,4,10,16\}, are the detected transmitted signals. In Fig.1 (b)–(d), with different \(D\) values, the steps are similar. When \(D = 4\) in Fig.1 (d), HSD is obviously equivalent to layer reordered K-Best SD [6].

In most practical cases, the nodes found at first in HSD are the same as the detected transmitted signals in K-Best SD, such as nodes \{1,4,10,16\} in Fig. 1. If these nodes are close to ML solution, their accumulated PED may be almost equal to 0 rather than 0.8. Under this situation, when \(D = 1\), only steps (i)–(iii) in Fig.1 (a) are done. So HSD can rapidly detect the transmitted signals; that is, it costs fewer operations than K-Best SD.

Contrasted with S-E SD (when \(D = 0\), HSD is equivalent to S-E SD because there is no \(K\) value constraint), HSD searches a small quantity of nodes, so its performance will be affected certainly. To resolve this problem, applying layer reordered method to the search process is an effective solution. By using SQRD to suppress error rate resulted from the earlier \(D\) layers, it can be assumed that error propagation is small enough to be ignored at the later \(N_f - D\) layers. Besides, with more detected signals, there is less uncertainty for the remaining undetected signals. In other words, the remaining undetected signals suffer less interantenna interference.

V. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

In our simulations, a packet structure similar to that of the IEEE 802.11n standard is assumed: \(N_f = N_s = 4\) and the modulation scheme is 64-QAM. Multipath Rayleigh fading channels in the simulations are generated by the modified Jakes' model [9], and the indoor channel model in [10] is applied. Besides, MMSE-SQRD is used in simulation for layer reordered method.

A. BER Performance

Fig. 2 shows the BER performance of HSD with distinct \(D\) values under 64-QAM modulation respectively. LR K-Best SD in [6] is one case of HSD; when \(D = 4\) in HSD, they are equivalent. Moreover, when \(K = 1\), HSD is equivalent to SQRD. In Fig. 2, it is obvious that no matter what \(D\) is, the performance overlaps other curves in the same \(K\) value. Contrasting with computational complexity, the simulation outcomes conform to our expectation that S-E SD typically searches the ML solution faster than K-Best SD.

B. Computational Complexity

As mentioned before, what HSD wants to achieve is to avoid redundant operations in K-Best SD. Fig. 3 and Fig. 4 show the number of searched nodes with \(K = 4\) and \(K = 16\), respectively, which are based on the following formula in the simulations

\[
\text{# of searched nodes} = \Omega + (D - 1)(K\Omega) + (N_f - D)\Omega + m\Omega
\]

where \(m\) is the number of nodes whose PEDs are smaller than the reference sphere radius \(r\). The first and the second terms in (7) are referred to K-Best SD at the first \(D\) layers; the third term is referred to the first searched leaf node in HSD; and the fourth term is referred to conditioned searching. For instance, when \(D = 4\), \(K = 16\), and \(\Omega = 64\) (64QAM) in our simulations, the number of searched nodes for one received signal \(r\) in (1) is equal to 3136; when \(D = 1\), \(K = 16\), and \(\Omega = 64\) in our simulations, the number of searched nodes for one received signal \(r\) in (1) is equal to 256 + 64\(m\), where 256 are the basic number of the searched nodes. If PED of the first reference sphere radius is almost equal to 0 (i.e. the ML solutions are detected at the beginning), the value \(m\) may equal 0. Assumed \(m = 0\), compared with K-Best SD, the normalized number of searched nodes is equal to 8.16\%, and this value is very close to our simulation result, the value 8.67\% in Table II, which will be discussed later. That is, when SNR is high (SNR>26dB in our simulations), \(m \equiv 0\).

In Fig. 3 and Fig. 4, HSD with smaller \(D\) values always has less complexity than that with larger ones at the same SNR, because S-E SD in HSD does not execute exhaustive search, which has a constraint on K-Best nodes. Especially when \(K\) is getting larger, the scenario is more conspicuous. In addition, Table I and Table II list the normalized searched nodes relative to the number of searched nodes with \(D = 4\) at SNR=14, 20, and 26. Contrasted with the performance in Fig. 2, it is obvious that under the same performance HSD can save the most so-called redundant operations than K-Best SD. Compared with LR K-Best SD [3], the complexity reduction of HSD with \(K = 4\) and \(K = 6\) at SNR=14 are 27.78\% and 80.84\%, respectively. As SNR increases, the amount of reduced complexity obviously also
grows accordingly. When SNR=26, it can achieve 67.33% and 91.33% reduction in the computation complexity, when K=4 and K=6, respectively.

Another important issue in sphere decoding, particularly S-E SD, is the storage cost. Although concept of S-E SD is used in HSD, it is restricted by the K value. Therefore, the maximum storage of HSD is equal to that of K-Best SD. Examples in Fig.1 illustrate the issue in detail. No matter what the value D is, the storage at each layer is no more than K.

![Fig. 3. The number of searched nodes of HSD with K=4 and D=1~4.](image)

![Fig. 4. The number of searched nodes of HSD with K=16 and D=1~4.](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNR=14 BER =1.5×10^{-1}</td>
</tr>
<tr>
<td>LR K-Best SD [3]</td>
<td>1</td>
</tr>
<tr>
<td>HSD with D=3</td>
<td>89.43%</td>
</tr>
<tr>
<td>HSD with D=2</td>
<td>77.44%</td>
</tr>
<tr>
<td>HSD with D=1</td>
<td>72.22%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNR=14 BER =1.5×10^{-1}</td>
</tr>
<tr>
<td>LR K-Best SD [3]</td>
<td>1</td>
</tr>
<tr>
<td>HSD with D=3</td>
<td>75.06%</td>
</tr>
<tr>
<td>HSD with D=2</td>
<td>50.43%</td>
</tr>
<tr>
<td>HSD with D=1</td>
<td>19.16%</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

In this paper, an efficient hybrid sphere decoding algorithm for MIMO systems is proposed. In the proposed method, concept of depth-first search scheme is adopted instead of deriving sphere radius with some assumptions like in most proposed methods. Without any stochastic assumption, the omitted searched nodes are exactly redundant. By evaluating the proposed method in terms of computational complexity and BER performance, it shows that the proposed method operates well in the targeted channel environments and its computational complexity is much less than those of the conventional methods.

REFERENCES


