Probability-Distribution-Based Node Pruning for Sphere Decoding

Tao Cui, *Student Member, IEEE*, Shuangshuang Han, *Student Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*—Node pruning strategies based on probability distributions are developed for maximum-likelihood (ML) detection for spatial-multiplexing multiple-input–multiple-output (MIMO) systems. Uniform pruning, geometric pruning, threshold pruning, hybrid pruning, and depth-dependent pruning are thus developed in detail. By considering the symbol error probability in the high signal-to-noise ratio (SNR) region, the desirable diversity order of uniform pruning and the threshold level for threshold pruning are derived. Simulation results show that threshold pruning saves complexity compared with popular sphere decoder (SD) algorithms, such as the K-best SD, the fixed-complexity SD (FSD), and the probabilistic tree pruning SD (PTP-SD), particularly for high SNRs and large-antenna MIMO systems. Furthermore, our proposed node pruning strategies may also be applied to other systems, including coded MIMO systems and relay networks.**

***Index Terms*—Maximum likelihood (ML), multiple-input–multiple-output (MIMO), sphere decoder (SD), statistical pruning, wireless communications.**

I. Introduction

Although multiple-input–multiple-output (MIMO) systems over rich-scattering wireless channels promise enormous capacity improvements without bandwidth or signal power increases [1], their realization depends on the availability of low-complexity high-performance signal detection algorithms. These requirements are met by the sphere decoder (SD) [2], with basic versions such as the Fincke–Pohst SD (FP-SD) and the more efficient Schnorr–Euchner SD (SE-SD)[1] [4], [5]. Hardware implementation details of the SD can be found in [6]. Although it avoids the exponential complexity of exhaustive search maximum-likelihood (ML) detection [7], the SD's average complexity is exponential in the number of transmit antennas [8]–[22]. In [9], an increasing radius algorithm (IRA) chooses a smaller radius for the lower layers in the search tree and increases the radius gradually for higher layers. This process amounts to pruning more nodes at lower levels and less at higher layers. Thus, excessive pruning may need to restart the IRA several times, which brings additional complexity. In addition, the IRA cannot attain different diversity orders[2] and achieve a flexible performance–complexity tradeoff. To extend the IRA, Shim and Kang [10] propose probabilistic tree pruning sphere decoding (PTP-SD), which prunes more nodes by adding a probabilistic noise constraint on top of the sphere constraint. Then, Shim and Kang [11] extend the PTP-SD and provide further improvement of the computational complexity with minimal extra cost and negligible performance penalty. Additional pruning methods are also proposed in [12] and [13]. References [14] and [16] combine the PTP-SD and fixed-complexity SD (FSD) [17] to preserve the advantages of branch pruning using an adaptively updated PTP-SD threshold. To prune more nodes, a new probabilistic sorting rule is developed by exploiting the properties of the path metric to yield more effective sorting [18]. The K-best SD [19], which prunes all but K-best nodes in each layer, traverses the search space in a breadth-first manner.

All those SD algorithms [9]–[14], [16]–[22] implement different node pruning strategies, with different performance–complexity tradeoffs. A general framework for such pruning is desirable. For this purpose, we propose and develop a statistical pruning SD (SPSD). Our main contributions in this paper are summarized as follows.

- Our key idea is that each essential node, e.g., node i, is assigned with probability $f(i)$ that indicates the likelihood of being pruned. For example, $f(i) = 0$ means that the ith node is retained, and $f(i) = 1$ means that ith node is eliminated. For other cases, given the probability distribution $f(i)$, our algorithm randomly generates a pruning decision for node i based on $f(i)$. We can choose $f(i)$ based on experimental results or common statistical distributions. For example, $f(i)$ may be set small for nodes in upper layers of the search tree, which means more such nodes are retained, and the ML solution is more likely to be found. In addition, $f(i)$ may be varied to achieve different performance–complexity tradeoffs. Performance and complexity are measured by the symbol error rate (SER) and the number of nodes visited. The SER in the high-SNR region is however closely related to the diversity

Manuscript received October 25, 2011; revised November 16, 2012; accepted December 2, 2012. Date of publication December 10, 2012; date of current version May 8, 2013. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by the China Scholarship Council. This paper was presented in part at the IEEE International Conference on Communications 2007, Scottish Exhibition & Conference Centre, Glasgow, U.K., June 2007. The review of this paper was coordinated by Prof. J. Chun.

T. Cui is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: taocui@caltech.edu).

S. Han and C. Tellambura are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton AB T6G 2V4, Canada (e-mail: shan7@ece.ualberta.ca; chintha@ece.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2012.2233226

[1]We consider the SE-SD [3] only. The extension of our proposed algorithms to the Fincke-Phost SD [2] is straightforward.

[2]If the error probability decays proportion to SNR^{-d}, d is called the diversity order.

order. Thus, a flexible tradeoff between diversity order and complexity reduction is achievable. Many existing SDs (such as [5], [9], and [19]) can be cast as special cases of the proposed approach.

- Based on several classical probability distributions, we propose the following node pruning rules: 1) uniform pruning, where all the child nodes of a node, except the first one, are pruned independently with equal probability; 2) geometric pruning, where the child nodes are pruned dependently where the pruning probability agrees with the geometric distribution; 3) threshold pruning, where the child nodes are pruned if its cost exceed a threshold; 4) hybrid pruning, where the threshold pruning or geometric pruning rule is combined with uniform pruning; and 5) depth-dependent pruning, where the pruning probability depends on the search depth. Three cases of these rules are given in this paper.
- The performance of the proposed SPSD is also analyzed in this paper. The upper bounds for the frame error rate (FER) of uniform and threshold pruning rules are derived. We also show that both the uniform and threshold pruning rules can achieve a desired diversity order by specifically setting the pruning probability for the former and the threshold for the latter according to different SNRs. Furthermore, pruning probability of the uniform rule is analyzed when full diversity order is needed. It is also shown that the FER upper bound in the full diversity case could be affected by the predesignated SNR loss and the achievable diversity orders or SNR gains could be controlled by the choice of pruning probability and the threshold. For example, for uniform pruning, to reduce complexity, a large pruning probability should be chosen based on the achievable diversity order, and *vice versa*. This also happens for the threshold rule, with a smaller desired diversity order; the threshold could be chosen to be a smaller value, resulting in lower complexity.
- The performance and complexity of all the proposed rules and those of the existing SDs, such as PTP-SD [10], FSD [17], and K-best SD [19], as a function of the number of transmit antennas and receive antennas are compared by simulations. The simulations show the advantages of our approach to large MIMO systems at high SNRs. It is noteworthy that our proposed threshold rule obtains significant complexity savings than these SD algorithms.

This paper is organized as follows. Section II describes an uncoded MIMO system and reviews the basic SD. The SPSD is developed in Section III. Five pruning rules are proposed in Section IV. Performance and complexity analyses of the proposed SPSD are given in Section V. The simulation results are presented in Section VI, followed by the conclusions in Section VII.

Notation: Bold symbols denote matrices or vectors. $(\cdot)^T$ denotes transpose. $\mathbb{R}$ denotes the real number set. $\Re\{x\}$ and $\Im\{x\}$ denote the real part and the imaginary part of x, respectively. $\|(\cdot)\|^2$ is the two-norm square of $(\cdot)$. A circularly complex Gaussian variable with mean μ and variance σ^2 is denoted by $z \sim \mathcal{CN}(\mu, \sigma^2)$. A^c denotes the complement event of A.

II. System Model and Conventional Sphere Decoder

We consider a spatial-multiplexing MIMO system with N transmit antennas and M receive antennas. A rich-scattering memoryless (flat fading) channel is assumed [7]. The transmitter selects complex symbols from a finite constellation $\mathcal{Q}$. The received signals may then be written as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$, $x_i \in \mathcal{Q}$ is the transmitted signal vector, $\mathbf{r} = [r_1, \ldots, r_M]^T$, $r_i \in \mathcal{C}$ is the received signal vector, $\mathbf{H} = [h_{i,j}] \in \mathcal{C}^{M \times N}$ is the channel matrix, and $\mathbf{w} = [w_1, \ldots, w_M]^T$, $w_i \in \mathcal{C}$ is an additive white Gaussian noise (AWGN) vector. The elements of $\mathbf{H}$ are independent identically distributed (i.i.d.) complex Gaussian, i.e., $h_{i,j} \sim \mathcal{CN}(0, 1)$. The components of $\mathbf{w}$ are i.i.d., and $w_i \sim \mathcal{CN}(0, \sigma^2)$. The channel $\mathbf{H}$ is assumed to be perfectly known to the receiver, which can be estimated by standard pilot-based channel estimation methods [23]. The number of receive antennas exceeds the number of transmit antennas $N \le M$ (if $N > M$, the rank-deficient system can be converted into a full-rank system with $N = M$ using the method in [24]). The channel matrix is factorized as $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is unitary, and $\mathbf{R}$ is an $N \times N$ upper triangular matrix. Received signal $\mathbf{r}$ is multiplied by $\mathbf{Q}^H$ as a part of the preprocessing stage. The ML detector is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{Q}^N} \|\mathbf{y} - \mathbf{R}\mathbf{x}\|^2 \tag{2}$$

where $\mathbf{y} = \mathbf{Q}^H\mathbf{r}$. Note that (2) forms the basis for the SD. In the following, without loss of generality, formulation (2) and $\mathcal{Q}$ are considered real, e.g., pulse amplitude modulation (PAM), which may be represented as $\mathcal{Q} = \{2q - |\mathcal{Q}| + 1 | q = 0, 1 \ldots |\mathcal{Q}| - 1\}/E_0$, where E_0 is the parameter to make the average energy of the constellation to be 1. If $\mathcal{Q}$ is complex, such as quadratic-amplitude modulation (QAM), then it can be decoupled to two real PAM constellations, and (2) still holds, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$ and $\mathbf{R} \in \mathbb{R}^{n \times n}$, where $n = 2N$. The reader is referred to [25] for additional details on how this equivalent real model is derived from the complex model.

A. Conventional SD

Here, the real-system formulation (2) is used to briefly explain the basic SD. Define $\mathbf{p}(\mathbf{x}) = \mathbf{y} - \mathbf{R}\mathbf{x}$. The basic SD operates by discarding $\mathbf{x}$ that do not satisfy $\|\mathbf{p}(\mathbf{x})\|^2 = \sum_{i=1}^n |p_i(\mathbf{x})|^2 \le d^2$. Since $\mathbf{R}$ is upper triangular, $p_i(\mathbf{x}) = y_i - \sum_{j=i}^n r_{i,j}x_j$ is a function of $x_i, \ldots, x_n$ only, where $r_{i,j}$ is the (i, j)th entry of $\mathbf{R}$. Since $|p_i(\mathbf{x})|^2$, $i = 1, \ldots, n$ is positive, for given $x_{k+1}, \ldots, x_n$, a necessary condition for valid $\mathbf{x}$ is that $|p_k(\mathbf{x})|^2 \le d^2 - \sum_{j=k+1}^n |p_j(\mathbf{x})|^2$.

In the SE enumeration [3], the admissible points are searched in a zigzag order from the midpoint $x_{k,\mathrm{mid}} = \lceil 1/r_{k,k}(y_k - \sum_{j=k+1}^n r_{k,j}x_j) \rfloor$, where $\lceil x \rfloor$ is the nearest integer around x. The spanning order is $x_{k,\mathrm{mid}}, x_{k,\mathrm{mid}}+1, x_{k,\mathrm{mid}}-1, x_{k,\mathrm{mid}}+2, \ldots$, when $y_k - d_k - r_{k,k}x_{k,\mathrm{mid}} \ge 0$ ($d_k^2 = d^2 - \sum_{j=k+1}^n |p_j(\mathbf{x})|^2 = d_{k+1}^2 - |p_{k+1}(x_{k+1}, \ldots, x_n)|^2$), and $x_{k,\mathrm{mid}}, x_{k,\mathrm{mid}} - 1, x_{k,\mathrm{mid}} + 1, x_{k,\mathrm{mid}} - 2, \ldots$, if otherwise. This method is more efficient than the Phost enumeration [5].

III. Statistical Pruning

Before describing specific pruning rules, we describe a generic detector. For the search tree with n layers, this detector is given in Algorithm 1,[3] where k $(k = n, n-1, \ldots 1)$ denotes the current layer in the search tree, $\mathbf{g}$ is a vector that includes the pruning probabilities that is designed to use the pruning rules for the nodes in the kth layer and d_k^2 is the current partial upper bound obtained by the radius minus the current accumulated partial cost. The algorithm is invoked as SPSD-decode $(n, \mathbf{g}, d_n^2, \mathbf{y}, \mathbf{R})$, where d_n is the initial radius, and $\mathbf{y}$ is the received signal. The initial radius d_n can be selected based on the noise level [26] for the original FP-SD, whereas it also can be typically set as $d_n^2 = +\infty$ for the conventional SE-SD [5].

Algorithm 1: Statistical Pruning Sphere Decoder **SPSD-decode** $(k, \mathbf{g}, d_k^2, \mathbf{y}, \mathbf{R})$

Input: k, $\mathbf{g}$, d_k^2, $\mathbf{y}$, $\mathbf{R}$
Output: $\mathbf{x}_{\min}$
1 Generate all the l_0 successors $\mathcal{A}$ in the kth layer satisfying $(y_k - r_{k,k}a_i)^2 \le d_k^2$ by eliminating nonessential nodes;
2 Let $[\sim, \text{temp}] = \text{sort}(\mathbf{c})$, where $\mathbf{c} = [c_1, c_2, \ldots, c_i, \ldots, c_{l_0}]$ and $c_i = (y_k - r_{k,k}a_i)^2$, and then $A = A(\text{temp})$;
3 **for** $i \leftarrow 1$ **to** $|\mathcal{A}|$ do
4 $\quad p = \text{rand}(1)$;
5 $\quad$**if** $p \le g(i)$ **then**
6 $\quad\quad$discard the ith node;
7 $\quad$**else**
8 $\quad\quad$keep the ith node in $\mathcal{A}$;
9 $\quad$**end**
10 **end**
11 $l = length(\mathcal{A})$;
12 **for** $i \leftarrow 1$ **to** l (*every element in* $\mathcal{A}$) do
13 $\quad \tilde{x}_k = a_i$;
14 $\quad$**if** a_i *is not a leaf node* then
15 $\quad\quad$Let $y_{k-1} = y_{k-1} - \sum_{j=k}^{n} r_{k-1,j}\tilde{x}_j$;
16 $\quad\quad$**SPSD-decode** $(k-1, \mathbf{g}, d_k^2 - c_i, \mathbf{y}, \mathbf{R})$;
17 $\quad$**else if** a_i *is a leaf node* $(k = 1)$ and *its cost is smaller than the current best cost* then
18 $\quad\quad$The best solution $\mathbf{x}_{\min} = \tilde{\mathbf{x}}$;
19 $\quad\quad$Update $d_n^2 = \|\mathbf{y} - \mathbf{R}\tilde{\mathbf{x}}\|^2$ and all d_i^2, $i = n-1, \ldots, 1$;
20 **end**

Note that Algorithm 1 is built on the top of the SE-SD [3] along with additional pruning of essential nodes based on heuristic rules. The full pruning rule $\mathbf{g}$ may be dependent on the search layer. This allows further flexibility to implement, e.g., more aggressive/conservative pruning strategies at different search layers. Recall that the pruning of nonessential nodes (i.e., $c_i > c_*$) does preserve the optimality of the algorithm. The pruning of essential nodes by the probabilistic rules, however, is the main concern of this paper. Specifically, considering $|\mathcal{Q}|$ nodes that are children of a node, their pruning probabilities may be given as

$$g(i) = \begin{cases} 1 & \text{if } c_i > c_* \\ f(i) & i = 1, \ldots, t \end{cases} \tag{3}$$

where $i \in \{1, 2, \ldots |\mathcal{Q}|\}$, c_* and c_i are the current best cost and the partial cost of the current node, respectively. t is the number of nodes whose partial cost is below c_*.

Note that, by pruning probability, we always refer to the second item $f(i)$ in (3), which defines the probability of pruning for nodes with $c_i \le c_*$ (essential nodes). This simple but critical difference with the SD makes the SPSD terminate sooner than the latter, hopefully with the ML solution. Set $\mathcal{A}$ is sorted in line 2 of Algorithm 1 because smaller cost nodes are more likely to give high-quality solutions. Experimentally, we know that pruning at different layers of the search tree will affect the performance–complexity tradeoff differently. Thus, in Algorithm 1, heuristic rules may vary for different layers. Thus, the pruning rule can be strong in the first few layers since the bound d_k^2 itself is not tight enough to identify nonessential nodes and can be weak in the last few layers when bound d_k^2 is tight.

A search algorithm is *complete* if it is guaranteed to return at least one valid solution. That is, at least the Babai point [4] or the decision-feedback equalization point [7] is guaranteed. To ensure this condition, in Algorithm 1, at least one child node is always kept. Thus, the pruning probability assigned to the child node with minimum cost is always zero.

To clarify these ideas, consider a simple example where a node has four essential child nodes. Assume that the child nodes are sorted in increasing cost and assigned their pruning probabilities as [0, 0.2, 0.5, 0.8], i.e., the first child node is never pruned, and others have more chance of being pruned because they are less likely to lead to the ML solution. Similarly, with different probability distributions, existing SDs can be cast as special cases of our framework. Some examples are as follows:

1) The SE-SD [5]: This case arises when the pruning probability $f(i)$ for all the child nodes is 0, i.e., SE-SD does not use statistical pruning.
2) The IRA [9]: The nodes with the costs smaller than the radius are retained, i.e., their pruning probabilities are 0. While the pruning probabilities are 1 for other nodes, which are pruned for sure. This is an example of uniform pruning rule.
3) The K-best SD [19]: At each layer of the search tree, the pruning probability of the best K nodes is 0, whereas the pruning probability of all the remaining $(K+1, K+2, \ldots)$ nodes is one.

In the next section, several specific pruning rules are also proposed.

IV. Pruning Rules

Five specific heuristic pruning rules are developed here. Uniform, geometric, threshold, hybrid and depth-dependent pruning rules are developed.

[3] A complete Matlab-like algorithm SD description can be found in [4].

A. Pruning Probability Distribution Basics

To keep our statistical framework simple, we initially define the pruning probability of the kth layer to be $f(i,k)$, $i = 1, \dots t$ and $k = 1, 2, \dots, n$. The value of $f(i,k)$ can be chosen to execute a strong or weak pruning regime and is not dependent on the layer number in the first several depth-independent rules, denoted by $f(i)$. However, $f(i,k)$ can also be chosen to vary with the layer. In the following, we only consider nodes that do not exceed the current best cost c_*. The set of such nodes is $\mathcal{A}$, and its size is t [see (3)].

The probability that node a_i is pruned is $f(i)$, and $f(i)$ is a nondecreasing function in i with $f(1) = 0$ and $0 \leq f(i) \leq 1$. As previously mentioned, boundary condition $f(1) = 0$ ensures the completeness of the SPSD. $f(i)$ is chosen as a nondecreasing function in i, because, intuitively, a child with a smaller cost is more likely to lead to the optimal solution. Based on the different probability distributions, several pruning rules are proposed next.

B. Pruning Rules

In the following, several specific pruning rules are given.

1) Uniform Pruning Rule. $f(i) = 1 - p$, for $2 \leq i \leq t$, and $f(i) = 0$, for $i = 1$: All child nodes, except the first one, are pruned with equal probability and independently. This rule is rational when *a priori* information is not available as to which child leads to the optimal solution and which ones should be pruned.

2) Geometric Pruning Rule. $f(i) = 1 - p^{i-1}$: Because all the child nodes are ordered in an increasing cost, the pruning probability could be defined to be an increasing function of i, i.e., geometric distribution. In this case, the child nodes are pruned dependently.

Since $f(i), i \geq 2$ in the geometric pruning rule is greater than that in the uniform pruning rule, the former is stronger than the latter. In both pruning rules, the strength of pruning is controlled through p.

These two pruning rules are defined with two well-known classical probability distributions. However, we may design the pruning probability distribution considering the cost c_i of node a_i. For example, a child node whose cost is significantly larger than its parent may not lead to an optimal solution. This idea leads to the following threshold pruning rule.

3) Threshold Pruning Rule: In the conventional SD, the current node is pruned if its cost exceeds the current best cost. A variation of this idea is proposed here. Thus, to further prune nodes, a threshold δ_k is applied at the kth layer. As previously mentioned, a_1 is never pruned $(f(1) = 0)$ for the completeness of the SPSD. For $i = 2, \dots, t$, if the cost c_i of child node a_i is larger than δ_k, a_i is pruned, i.e., $f(i) = 1$ when $c_i > \delta_k$.

Here, threshold δ_k is associated with the kth layer. Since $\mathcal{A}$ is in a nondecreasing order of cost, if node a_i is pruned, all the children a_j for $j \geq i$ are pruned. The strength of pruning varies inversely with δ_k. For example, if $\delta_k = +\infty$, it simply reduces to the SD.

4) Hybrid Pruning Rule: The threshold rule can also be combined with the uniform rule or the geometric rule to possibly take advantage of both the cost information and probabilistic pruning. Thus, the hybrid pruning rule could be constructed as $f(i) = 1 - p^{\sqrt{c_i/c_1} - 1}$ or $f(i) = \min\{(c_i - c_1)(1-p)/c_1, 1\}$, where c_1 is the minimum cost in the kth layer. For these two examples, if the cost c_i of the child node a_i is less than or equal to c_1 (threshold rule), a_i is not pruned. Otherwise, the nodes are pruned by $f(i)$ (uniform or geometric rule).

5) Depth-Dependent Pruning Rule: In the search process, if the pruning probability $f(i)$ at early search layers is too high, the probability of discarding the ML solution increases. To keep the ML solution until the bottom search layer, $f(i)$ may be defined depending on different tree layers (depth-dependent pruning rule) denoted by $f(i,k)$, $k = 1, 2, \dots n$. In the following, three cases are given.

Case I: $f(i,k) = 1 - \exp(-k)$, $2 \leq i \leq t$.
The children at each layer are pruned by probability $f(i,k)$, which is a nonincreasing function in k. For large k, the pruning probability decreases, which helps to avoid discarding the ML solution. Once the node at the kth layer is pruned, all the children of this node are pruned.

Case II: $f(i,k) = (n-k)/2(n-1), 1 \leq k \leq n$ where $2 \leq i \leq t$. The pruning probability increases linear with the layer. For the first layer, the pruning probability is zero, and for the bottom layer, it is 1/2.

Case III: $f(i,k) = \begin{cases} 0 & n_1 + 1 \leq k \leq n \\ p_0 & 1 \leq k \leq n_1 \end{cases}$, where $1 \leq n_1 \leq n$ and $2 \leq i \leq t$. The nodes at early search layers $(n_1 \leq k \leq n)$ are all kept and expanded; and the remaining nodes are pruned by probability p_0 $(1 \leq k \leq n_1 - 1)$. That is, full enumeration at the beginning of the search process is used to improve the probability of finding the ML solution. However, in the latter search layers, nodes are pruned with probability p_0 to reduce the complexity. Note that, if $n_1 = 0$, this rule is the conventional SD; if $n_1 = n$, uniform pruning becomes one special case with $1 - p_k = p_0$.

In these three cases, the first node at each layer is never pruned $(f(1,k) = 0)$ for the detection completeness. Because the pruning probability in Case II is larger than Case I for each layer, the former is stronger than the latter. Only these three cases are given in this paper; there are many other cases for depth-dependent pruning rules, which are not further discussed here.

Remarks:

- Our idea of probability-distribution-based node pruning can be used with other tree search algorithms, e.g., best-first search, breadth-first search, and iterative deepening [27].
- More importantly, the same idea can also be applied to the complex-valued SDs, such as [17], to achieve a flexible performance–complexity tradeoff.
- The IRA in [9] is a special case of Algorithm 1. The IRA chooses a smaller radius for the lower layers of the search tree (see details in [9]). However, if the IRA cannot find a point, the radius is increased, and the search resumes. Our threshold pruning rule at least obtains one point as the solution, and the threshold for each layer is different with the IRA.

- The threshold pruning rule can be readily incorporated into the SD algorithm by replacing $d_k^2 = d_{k+1}^2 - |p_{k+1}(x_{k+1}, \ldots, x_n)|^2$ in Section II-A with $d_k^2 = \min\{d_{k+1}^2 - |p_{k+1}(x_{k+1}, \ldots, x_n)|^2, \delta_k\}$. If d_k^2 returns a null set, we keep $a_1 = \lceil \rho_k / r_{k,k} \rfloor$ in $\mathcal{A}$. In fact, δ_k can be considered as a local bound, as opposed to the global bound d^2 in the SD.
- The K-best SD is a special case of the depth-dependent pruning rule (Case III). When $n_1 = 0$, $p_0 = 0$ for $2 \leq i \leq K$, and $p_0 = 1$ for $K+1 \leq i \leq t$; the K-best SD is obtained.

V. Performance Analysis and Complexity Measurement

A. Performance Analysis

Here, the performance of uniform and threshold rules is analyzed. The parameters p and δ_k are determined to achieve different diversity orders and performance gains. To make the analysis tractable, detection ordering is ignored. The radius is assumed infinite, and the effect of decreasing the radius as in Algorithm 1 is ignored. The results here can be considered as upper bounds for those cases with column reordering.

Proposition I: The upper bound on the FER of uniform pruning is

$$P_f \leq (1-p) \sum_{i=1}^{n} \frac{|\mathcal{Q}|}{\left(1 + \frac{d_{\min}^2}{4\sigma^2}\right)^i} + \left(\frac{|\mathcal{Q}|}{1 + \frac{d_{\min}^2}{4\sigma^2}}\right)^n \quad (4)$$

where $d_{\min}$ is the minimum Euclidean distance of $\mathcal{Q}$, and σ^2 is the variance of the noise. For the proof, see Appendix A.

In the high-SNR region, the SER P_s can be approximated by P_f, $P_s \approx P_f/n$, where a frame error is caused by a single symbol error with high probability.

From (4), if p is fixed for all SNRs and $p \neq 1$ (uniform pruning), the first term dominates P_f. As (4) is only an upper bound, it suggests that the diversity order of uniform pruning is at least one. The simulation results indicate that the diversity order of the uniform rule is indeed at least one. Since geometric pruning is stronger than uniform pruning, the diversity order in geometric pruning is also at least one for fixed p. Equation (4) also indicates that, to achieve a diversity order n with uniform pruning, $1-p$ must at least decrease as fast as $1/(1 + d_{\min}^2/4\sigma^2)^{n-1}$. Therefore, p must vary according to the SNR or σ^2. We thus choose

$$p = 1 - \xi \left(\frac{1}{1 + d_{\min}^2/4\sigma^2}\right)^{K_0 - 1} \quad (5)$$

where ξ is a constant. Substituting (5) into (4), we have

$$P_f \leq \xi \sum_{i=1}^{n} \frac{|\mathcal{Q}|}{\left(1 + \frac{d_{\min}^2}{4\sigma^2}\right)^{K_0+i-1}} + \left(\frac{|\mathcal{Q}|}{1 + \frac{d_{\min}^2}{4\sigma^2}}\right)^n. \quad (6)$$

If $K_0 < n$, the first term dominates P_f in the high-SNR region, and the other terms can be neglected. Therefore, uniform pruning achieves at least a diversity order K_0. From (6), we can see that ξ controls the SNR gain of statistical pruning.

To achieve the full diversity order n, one can choose

$$p = 1 - \beta P_{\mathrm{ML}} \left(\frac{|\mathcal{Q}|}{1 + d_{\min}^2/4\sigma^2}\right)^{-1} \quad (7)$$

where P_{ML} is the FER of the ML detector. Substituting (7) into (4), we have

$$P_f \leq \beta P_{\mathrm{ML}} \sum_{i=1}^{n} \frac{1}{\left(1 + \frac{d_{\min}^2}{4\sigma^2}\right)^{i-1}} + \left(\frac{|\mathcal{Q}|}{1 + \frac{d_{\min}^2}{4\sigma^2}}\right)^n \quad (8)$$

where β controls the SNR loss incurred by the statistical pruning compared with the ML detector.

From [28], when SNR becomes high, the asymptotic form of P_{ML} can be expressed as

$$P_{\mathrm{ML}} = \alpha(n, \mathcal{Q}) \left(\frac{1}{2\gamma}\right)^n \binom{2n-1}{n-1} \quad (9)$$

where γ denotes SNR. $\alpha(n, \mathcal{Q})$ is a coefficient that depends on n and the constellation. Let $\{\mathbf{d}_j\}$ denote the set of vectors with $x_k \in Q$ as their lth element, and $\{\mathbf{d}_i\}$ denote the set of vectors that differ in their lth element from $\mathbf{d}_j$. α is given by [28]

$$\alpha = \frac{1}{|Q|^n} \sum_{x_k \in Q} \sum_i \sum_j \left(\frac{\|\mathbf{d}_i - \mathbf{d}_j\|^2}{2E_s}\right)^{-n} \quad (10)$$

where E_s is the average symbol energy of Q. Since (9) scales as γ^n, from (8), statistical pruning by using (7) can still achieve diversity order n. Note that performance analysis here only consider a fixed pruning probability for all x_k, $k = 1, \ldots, n$. We may also assign a different pruning probability for different x_k's.

Proposition II: The FER of threshold rule is bounded as

$$P_f \leq \sum_{i=1}^{n} \int_{\frac{\delta_i}{\sigma^2}}^{+\infty} f_i(x)dx + \left(\frac{|\mathcal{Q}|}{1 + \frac{d_{\min}^2}{4\sigma^2}}\right)^n \quad (11)$$

where $f_i(x)$ is the probability density function (pdf) of the chi-square distribution $\chi^2(2(n-i+1))$ [29]. For the proof, see Appendix B.

To achieve a diversity order of at least K_0, δ_i may be chosen such that

$$\int_{\frac{\delta_i}{\sigma^2}}^{+\infty} f_i(x)dx = \frac{\xi}{(1 + d_{\min}^2/4\sigma^2)^{K_0}} \quad (12)$$

where ξ is a constant that controls the SNR gain. Since $f_i(x)$ is known, (12) can be solved numerically. At each SNR, (12) needs to be solved only once. We simply set $\delta_i = 0$ for $i \leq n - K_0 + 1$. In this case, the upper bound on P_f is also given by (11).

With the same δ, it can be easily verified that

$$\int_{\frac{\delta}{\sigma^2}}^{+\infty} f_i(x)dx < \int_{\frac{\delta}{\sigma^2}}^{+\infty} f_j(x)dx \quad (13)$$

for $i > j$. Therefore, with the same δ, the first term in (11) always dominates P_f. A simplified rule to achieve a diversity order of at least K_0 can be obtained by setting δ as the solution of

$$1 - \gamma\left(1, \frac{\delta}{2\sigma^2}\right) = \frac{\xi}{\left(1 + d_{\min}^2/4\sigma^2\right)^{K_0}}. \tag{14}$$

Similarly, we simply set $\delta_i = \delta$ for $i > n - K_0 + 1$ and $\delta_i = 0$ for $i \leq n - K_0 + 1$. Using (11), it can be readily verified that this choice of δ_i achieves a diversity order of K_0. Interestingly, the cost threshold δ only depends on the SNR and K_0 but not on i.

Remarks:

- The upper bound in (4) may not be tight. The p given in (5) and (7) may achieve better performance than that suggested by (4). This also holds for δ_i in (12) and (14).
- The simulation results (see Section VI) show that the performance difference between uniform and geometric rule is small when using the same p defined in (5) and (7). For the same p, uniform pruning only has an SNR gain over geometric pruning, although the latter is stronger than the former, as remarked upon in Section IV. The value of p given by (5) and (7) along with geometric pruning achieves the same diversity order. However, the diversity order analysis for this case seems intractable.

VI. Simulation Results

Here, the SPSD is simulated for an uncoded MIMO system over a flat Rayleigh fading channel. The modulation formats 4-QAM and 16-QAM are used. Both performance and complexity of any SD are compared. The complexity is measured by the average number of nodes visited. The computational complexity of the preprocessing stage is not counted. The ML curve, which is the optimal performance, is obtained with the conventional SE-SD. The initial radius of the SPSD is chosen to be infinite and is updated whenever it reaches a leaf node (Algorithm 1). In hybrid pruning, $f(i) = 1 - p^{\sqrt{c_i/c_1} - 1}$, where c_1 is the minimum cost in the kth layer. For the depth-dependent rule, only the result of Case I is given.

A. Comparison of Different Pruning Rules

Here, 4-QAM, and eight transmit antennas and eight receive antennas are used in Figs. 1 and 2, where ξ is set to be 0.8 [ξ is the constant in (5)]; the achievable diversity order K_0 is set to be 2 and 4 for uniform pruning, geometric pruning, and hybrid pruning; ξ is chosen to be 1; and K_0's are chosen to be 4 or 8 for threshold pruning.

Fig. 1 shows the SER performance of the SPSD with different statistical pruning rules. As shown, Case I of the depth-dependent pruning Rule achieves near-ML performance, which means the pruning probabilities for this rule are small for all the layers. For other rules, our derivation of achievable diversity order K_0 in (5) and (14) is validated here. At the desirable diversity order K_0 of 2, geometric pruning achieves a diversity order of 2. Likewise, with the achievable diversity order of 4, the uniform, threshold, and hybrid pruning rules could achieve a diversity order of 4. All these rules achieve the desirable diversity order corresponding to the value of K_0, which prove that the results for the uniform and threshold rules are also applicable for the geometric and hybrid rules. Another interesting observation is that, by setting a greater desirable diversity order K_0, the threshold rule performs closer to the optimal ML detection. For example, at an SER of 10^{-3}, threshold rule with diversity order K_0 of 8 attains 3-dB gains than the case with a diversity order of 4.

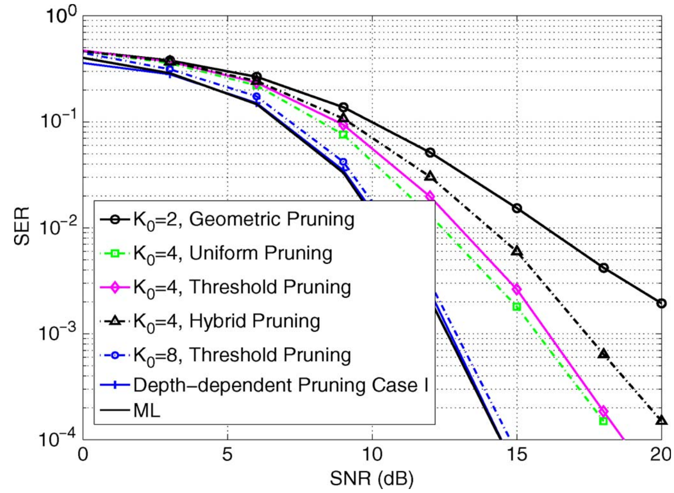


Fig. 1. Performance comparison for different statistical pruning rules for an 8×8 4-QAM MIMO system. The ML curve is given by the SE-SD.

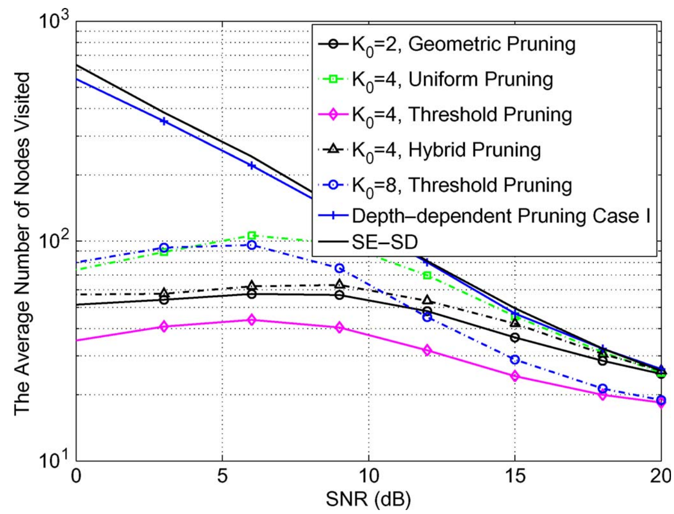


Fig. 2. Complexity comparison of different statistical pruning rules for an 8×8 4-QAM MIMO system.

Fig. 2 compares the complexity of different pruning rules with that of the optimal SE-SD. As expected, the SE-SD has the highest complexity compared with all the proposed rules. The only exception is the depth-dependent pruning rule, which has almost the same complexity as the SE-SD, but it does achieve the near ML performance (see Fig. 1). An immediate observation is that the achievable diversity order K_0 has significant effect on complexity. With a smaller desired diversity order K_0, the complexity is lower. This shows that lower complexity could be achieved by sacrificing the desirable diversity order or SER performance. All the rules excluding the depth-dependent

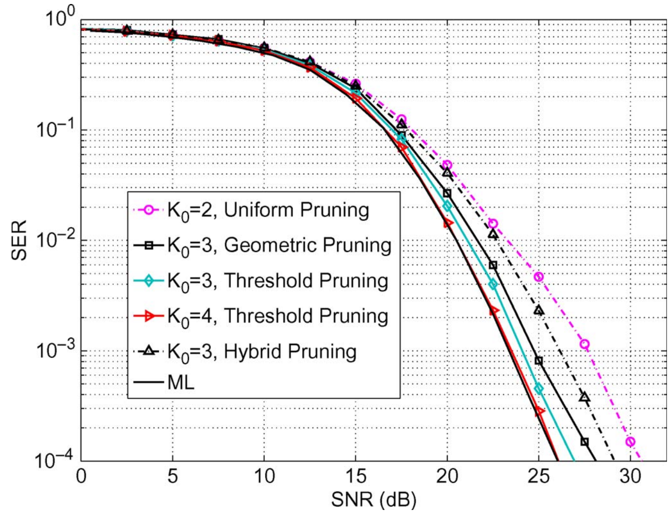


Fig. 3. Performance comparison of different statistical pruning rules for a 4×4 16-QAM MIMO system. The ML curve is obtained by the SE-SD.

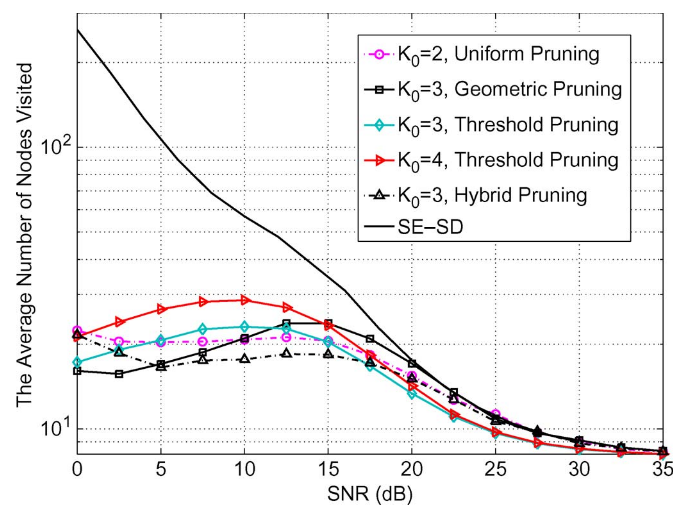


Fig. 4. Complexity comparison of different statistical pruning rules for a 4×4 16-QAM MIMO system.

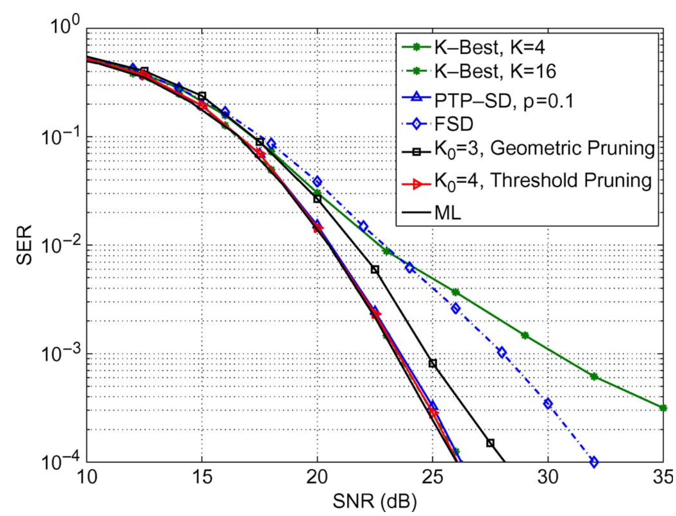


Fig. 5. Performance comparison of different detectors for a 4×4 16-QAM MIMO system. The ML curve is obtained by the SE-SD.

rule obtain more complexity savings in the low-SNR region, whereas their complexity saving over SE-SD reduces when SNR increases. For example, at an SNR of 0 dB, the threshold pruning (diversity order of 8) obtains about 86% of complexity saving than the SE-SD. This number reduces to 50% at 10 dB. Furthermore, the threshold pruning obtains complexity saving for very high SNRs such as 20 dB. The reason is that the SE-SD visits several unnecessary nodes at the early stages of the search process. However, for threshold pruning, the local bound prevents visiting those nodes, particularly when the cost threshold δ_i is chosen to be 0 for $i \leq n - K_0 + 1$, which means only a single node is visited at layers $1, \ldots, n - K_0 + 1$.

The performance of the SPSD for different MIMO systems, i.e., 16-QAM 4×4 MIMO system, is next assessed in Figs. 3 and 4, where the parameter setting is the same as 8×8 4-QAM MIMO system, except $K_0 = 2, 3, 4$. The SER of the SPSD for different statistical pruning rules is given in Fig. 3. As discussed in Fig. 1, by varying achievable diversity K_0, different diversity orders are achieved. For example, the threshold rule achieves full diversity order when the desirable diversity order is set to be 4.

Fig. 4 shows the average number of nodes visited with different pruning rules. Again, similar trends, as shown in Fig. 2, are observed. All the rules save complexity compared with the SE-SD for low SNRs, but the complexity saving reduces with increasing SNR. However, in the high-SNR region, threshold pruning attains lower complexity than the SE-SD.

In Figs. 1–4, achieving full diversity order, threshold pruning obtains the lowest complexity compared with other rules. Therefore, for near-optimal performance with significant complexity savings, threshold pruning is the best choice.

B. Comparison With Other Detectors

It is interesting to compare our SPSD with other existing detectors that use node pruning. Thus, we consider PTP-SD [10], the FSD, and the K-best SD [19]. For the PTP-SD, p' is set to be 0.1, where p' is the pruning probability; for the FSD, the case without channel ordering is used in this paper and the distribution of nodes kept in each layer is [1, 1, 1, 16], whereas K is chosen to be 4 and 16 for the K-best SD (mode 1 in [19] without channel ordering is used for fair comparison). Figs. 5 and 6 compare different MIMO detectors for a 4×4 16-QAM MIMO system. Only geometric and threshold rules are shown because uniform pruning performs close to geometric pruning, and threshold pruning performs better than hybrid pruning.

The SER performance comparison is shown in Fig. 5. Our proposed threshold rule with desirable full diversity order of 4, the PTP-SD, and the K-best SD $(K = 16)$ achieve the near ML performance. However, although the FSD has a fixed complexity by full enumeration in the first layers and pruning all but the first node with the minimum cost in the following layers, at an SER of 10^{-4}, it has 6-dB performance loss compared with our threshold rule. To be fair, this gap is due to not using channel matrix reordering. Similar with the FSD, the K-best SD also obtains fixed complexity achieved; however, it requires large K to achieve full diversity order [19]. Thus, the case $(K = 4)$ only achieves a diversity order of 1.

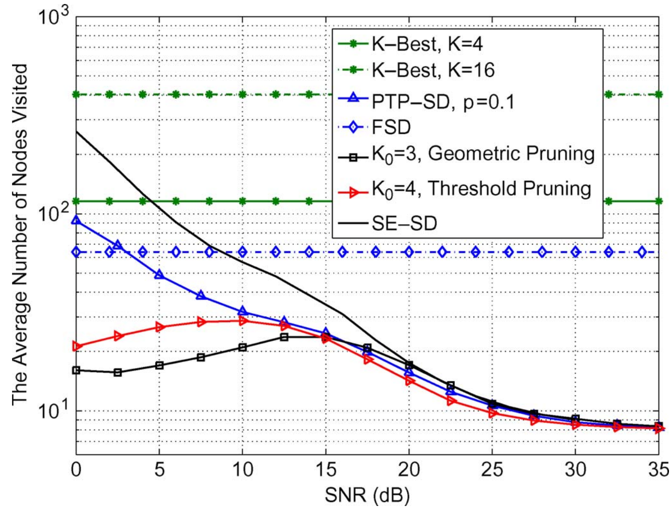


Fig. 6. Complexity comparison of different detectors for a 4 × 4 16-QAM MIMO system.

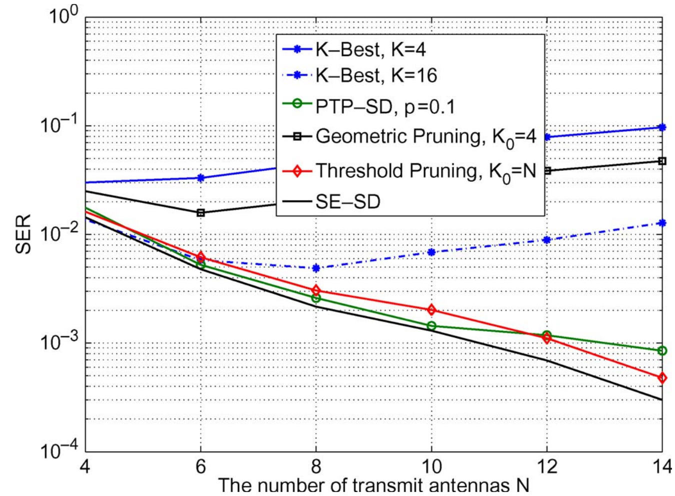


Fig. 7. Performance comparison of different detectors for a 16-QAM MIMO system and different numbers of transmit and receive antennas N. SNR $= 20$ dB.

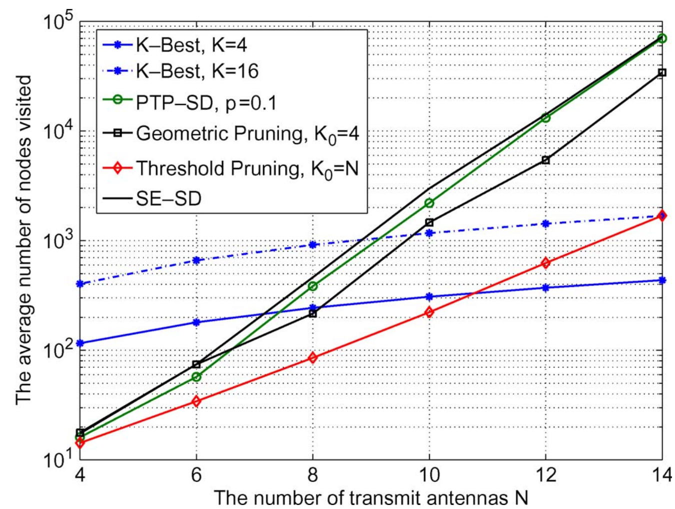


Fig. 8. Complexity comparison of different detectors for a 16-QAM MIMO system and different numbers of transmit and receive antennas N. SNR $= 20$ dB.

Fig. 6 shows the average number of nodes visited by different detectors. As previously mentioned, the choice of smaller desired diversity order K_0 leads to lower complexity. For example, the geometric rule $(K_0 = 3)$ has lower complexity than that of the threshold rule $(K_0 = 4)$. Moreover, the threshold rule has lower complexity than the PTP-SD in the low-SNR region; for example, at an SNR of 0 dB, the former obtains 78% complexity savings than the latter. The reason is that the threshold rule prunes more nodes than the PTP-SD. Another observation is both geometric and threshold rules have significantly lower complexity than K-best SD, which performs a breadth-first search and always prunes all but K-best nodes at each layer. Even so, the complexity of threshold rule is only 4.5% of the K-best SD $(K = 16)$ on average, whereas it also obtains 72% complexity savings than the FSD as well. To summarize, with near optimal SER performance, threshold pruning achieves the lowest complexity compared with PTP-SD, FSD, and K-best SD.

To show the advantages of our approach to large MIMO at high SNRs, a performance and complexity comparison as a function of the number of transmit antennas and receive antennas (16-QAM) is shown in Figs. 7 and 8, where N is the number of transmit or receive antennas. The SNR is fixed at 20 dB.

In Fig. 7, the proposed threshold rule with full diversity order N and the PTP-SD always achieves the near ML performance for different number of transmit antennas. However, the geometric rule with the fixed achievable diversity order of 4 does not reach the optimal performance for large MIMO systems. Due to the same reason, K-best SD with $K = 4$ and $K = 16$ could not also achieve near ML performance. This means that, with increasing the number of antennas N, the achievable diversity order K_0 for the geometric rule and the K for the K-best SD should be larger.

The complexity comparison with the same setup in Fig. 7 is given in Fig. 8. The complexity of the PTP-SD is almost the same of the SE-SD, which grows exponentially with N. Thus, PTP-SD does not achieve complexity savings than the SE-SD for large MIMO systems and high SNRs. However, the complexity of threshold pruning obtains significant complexity savings than the above two SDs. Further, the complexity savings increase with the number of transmit antennas. For example, for $N = 10$ and $N = 14$, the threshold rule obtains one and two orders of magnitude of complexity savings compared with SE-SD and PTP-SD. Although the K-best SD has less complexity when the number of antennas is large, it could not achieve the near ML performance. Therefore, for high SNRs, the threshold pruning rule significantly reduces the complexity with the near-optimal performance, particularly for the MIMO systems with a large number of transmit antennas.

VII. Conclusion

Probability-distribution-based SPSD has been proposed, and five specific pruning rules have been developed: uniform pruning, geometric pruning, threshold pruning, hybrid pruning, and depth-dependent pruning. We analyzed the SER performance of uniform and threshold pruning rules, and derived the pruning probability and the threshold for achievable diversity order K_0.

All these rules achieve lower complexity than the conventional SD in the low-SNR region. In particular, the threshold pruning rule obtains the most significant complexity saving while achieving the full diversity order. For example, in low SNR, 80% and 95% complexity savings are possible over the PTP-SD and K-best SD, while also achieving a slightly better SER performance. Moreover, complexity savings are also obtained for high SNRs (e.g., 20 dB), which also increases with the number of transmit antennas.

For future work, four directions are suggested.

- The probability-distribution-based node pruning that has been introduced in this paper simply uses several classical probability distributions, which are not contingent upon specific channel knowledge. We can expect even better performance by choosing the probability distribution based on specific channel knowledge and adapting to different conditions. Thus, there is much room to develop other pruning probability distributions.
- The probabilistic node pruning idea can also be applied to the soft detection of coded MIMO systems including an outer encoder [30]. For this case, soft information on the coded bits and extrinsic information exchange must be developed.
- The proposed algorithms could be combined with lattice-reduction-aided detection methods (e.g., in [31] and [32]) to achieve more performance improvements or complexity savings.
- Finally, our algorithm could be used elsewhere in wireless communications (e.g., code-division multiple access [33], MIMO relay networks [34], and multiuser networks).

Appendix A

To derive the FER upper bound in uniform rule, let $\mathbf{x}^{(1)} = [x_1^{(1)},\ldots,x_n^{(1)}]^T$ denote the transmitted vector and $\hat{\mathbf{x}} = [\hat{x}_1,\ldots,\hat{x}_n]^T$ denote the vector returned by the SPSD. We have $P_f = \Pr(\hat{\mathbf{x}} \neq \mathbf{x}^{(1)})$. Denote A as the event that $\mathbf{x}^{(1)}$ is visited. By using the total probability theorem [29], the FER can be expressed as

$$\begin{aligned} P_f &= \Pr\left(\hat{\mathbf{x}} \neq \mathbf{x}^{(1)}|A^c\right)\Pr(A^c) + \Pr\left(\hat{\mathbf{x}} \neq \mathbf{x}^{(1)}|A\right)\Pr(A) \\ &= \Pr(A^c) + \Pr\left(\hat{\mathbf{x}} \neq \mathbf{x}^{(1)}|A\right)\Pr(A) \end{aligned} \tag{15}$$

where $\Pr(\hat{\mathbf{x}} \neq \mathbf{x}^{(1)}|A^c) = 1$. We first derive $\Pr(A^c)$ [or $1 - \Pr(A)$] and then analyze the second term of (15). Let $\tilde{\mathbf{x}} = [\tilde{x}_1,\ldots,\tilde{x}_n]^T$ be the temporary value for $\mathbf{x} = [x_1,\ldots,x_n]^T$ during the statistical pruning search, as in Algorithm I, which corresponds to a leaf node in the search tree. Denote A_i as the event that $x_i^{(1)}$ is visited. Note that $\Pr(A) = \Pr(\tilde{\mathbf{x}} = \mathbf{x}^{(1)})$ is the probability that the leaf node corresponding to $\mathbf{x}^{(1)}$ is visited and is given by

$$\begin{aligned} \Pr(A) &= \Pr\left(\tilde{\mathbf{x}} = \mathbf{x}^{(1)}|\tilde{x}_n = x_n^{(1)}\right)\Pr\left(\tilde{x}_n = x_n^{(1)}\right) \\ &\quad + \Pr\left(\tilde{\mathbf{x}} = \mathbf{x}^{(1)}|\tilde{x}_n \neq x_n^{(1)}\right)\Pr\left(\tilde{x}_n \neq x_n^{(1)}\right) \\ &= \Pr\left(\tilde{\mathbf{x}} = \mathbf{x}^{(1)}|\tilde{x}_n = x_n^{(1)}\right)\Pr(A_n) \end{aligned} \tag{16}$$

where $\Pr(\tilde{x}_n = x_n^{(1)}) = \Pr(A_n)$, and $\Pr(\tilde{\mathbf{x}} = \mathbf{x}^{(1)}|\tilde{x}_n \neq x_n^{(1)}) = 0$. By a similar argument, (16) can be expanded as

$$\Pr(A) = \Pr(A_n)\prod_{i=1}^{n-1}\Pr\left(A_i|\tilde{x}_{i+1} = x_{i+1}^{(1)},\ldots,\tilde{x}_n = x_n^{(1)}\right). \tag{17}$$

Let B_i denote the event that $x_i^{(1)}$ is not the first element of $\mathcal{A}$ in Algorithm I. We have

$$\begin{aligned} \Pr\left(A_n^c\right) &= \Pr\left(A_n^c|B_n\right)\Pr(B_n) + \Pr\left(A_n^c|B_n^c\right)\Pr\left(B_n^c\right) \\ &= (1-p)\Pr(B_n) \end{aligned} \tag{18}$$

where $\Pr(A_n^c|B_n) = 1 - p$, and $\Pr(A_n^c|B_n^c) = 0$. The union bound for $\Pr(B_n)$ is given by

$$\Pr(B_n) \leq \mathop{E}_{r_{n,n}} \mathop{E}_{x_n^{(1)}} \left[\sum_{x_n^{(2)} \neq x_n^{(1)}} \Pr\left(\left|y_n - r_{n,n}x_n^{(2)}\right|^2 < \left|y_n - r_{n,n}x_n^{(1)}\right|^2 \left| x_n^{(1)}, r_{n,n}\right.\right)\right] \tag{19}$$

where $x_n^{(2)}$ is the nearest neighbor of $x_n^{(1)}$. From [29], the squared norm of the entries of upper triangular matrix $\mathbf{R}$ have a χ^2 distribution with different degrees of freedom without column reordering; specifically, $|r_{i,i}|^2 \sim \chi^2(2(n-i+1))$, for $i = 1,\ldots,n$, and $|r_{i,j}|^2 \sim \chi^2(2)$, for $j > i$, where $\chi^2(k)$ denotes the chi-squared distribution with k degrees of freedom. We can obtain

$$\begin{aligned} &\Pr\left(\left|y_n - r_{n,n}x_n^{(2)}\right|^2 < \left|y_n - r_{n,n}x_n^{(1)}\right|^2 \left| x_n^{(1)}, r_{n,n}\right.\right) \\ &\qquad = Q\left(\sqrt{\left|r_{n,n}\left(x_n^{(2)} - x_n^{(1)}\right)\right|^2 / 2\sigma^2}\right). \end{aligned} \tag{20}$$

where $Q(\cdot)$ is the Q-function. Using the Chernoff bound for the Q-function, $\Pr(B_n)$ can be bounded as

$$\begin{aligned} \Pr(B_n) &\leq \mathop{E}_{r_{n,n}} \mathop{E}_{x_n^{(1)}} \left[\sum_{x_n^{(2)} \neq x_1^{(1)}} \exp\left(\frac{-r_{n,n}^2\left|x_n^{(2)} - x_n^{(1)}\right|^2}{4\sigma^2}\right)\right] \\ &= \mathop{E}_{x_n^{(1)}} \sum_{x_n^{(2)} \neq x_1^{(1)}} \frac{1}{1 + \frac{\left|x_n^{(2)} - x_n^{(1)}\right|^2}{4\sigma^2}} \leq \frac{|\mathcal{Q}|}{1 + d_{\min}^2/4\sigma^2}, \end{aligned} \tag{21}$$

where $d_{\min}$ is the minimum Euclidean distance of $\mathcal{Q}$, and the equality comes from the moment-generating function of $r_{n,n}$, $M_{r_{n,n}}(t) = E\{e^{tr_{n,n}}\}$. Therefore, $\Pr(A_n^c)$ can be bounded as

$$\Pr(A_n^c) \leq (1-p)\frac{|\mathcal{Q}|}{1 + d_{\min}^2/4\sigma^2}. \tag{22}$$

Similarly, the conditional probability is bounded as

$$\begin{aligned} &\Pr\left(A_i^c|\tilde{x}_{i+1} = x_{i+1}^{(1)},\ldots,\tilde{x}_n = x_n^{(1)}\right) \\ &\leq (1-p)\frac{|\mathcal{Q}|}{\left(1 + d_{\min}^2/4\sigma^2\right)^{n-i+1}}, \quad i = 1,\ldots,n-1. \end{aligned} \tag{23}$$

Finally, an upper bound on $\Pr(A^c)$ is obtained as

$$\begin{aligned}\Pr(A^c) &= 1-\Pr(A)\\ &= 1-\Pr(A_n)\prod_{i=1}^{n-1}\Pr\left(A_i|\tilde{x}_{i+1}=x_{i+1}^{(1)},\ldots,\tilde{x}_n=x_n^{(1)}\right)\\ &= 1-(1-\Pr(A_n^c))\\ &\quad\times\prod_{i=1}^{n-1}\left(1-\Pr\left(A_i^c|\tilde{x}_{i+1}=x_{i+1}^{(1)},\ldots,\tilde{x}_n=x_n^{(1)}\right)\right). \quad (24)\end{aligned}$$

In the high-SNR region, $\Pr(A_n^c)$ and $\Pr(A_i^c|\tilde{x}_{i+1}=x_{i+1}^{(1)},\ldots,\tilde{x}_n=x_n^{(1)})$, $i=1,\ldots,n-1$ are small. $\Pr(A^c)$ can be well approximated as

$$\begin{aligned}\Pr(A^c) &\approx \Pr(A_n^c)+\sum_{i=1}^{n-1}\Pr\left(A_i^c|\tilde{x}_{i+1}=x_{i+1}^{(1)},\ldots,\tilde{x}_n=x_n^{(1)}\right)\\ &\le (1-p)\sum_{i=1}^{n}\frac{|\mathcal{Q}|}{(1+d_{\min}^2/4\sigma^2)^i}. \quad (25)\end{aligned}$$

We then bound $\Pr(\hat{\mathbf{x}}\neq\mathbf{x}^{(1)}|A)$ in (15) in the following. Denote the set of all the visited leaf nodes by $\mathcal{I}$, which is the candidate set for the output of the statistical pruning detection. Since some leaf nodes may be pruned, $|\mathcal{I}|\le|\mathcal{Q}|^n$. In the case of A, $\mathbf{x}^{(1)}\in\mathcal{I}$. The union bound for $\Pr(\hat{\mathbf{x}}\neq\mathbf{x}^{(1)}|A)$ is given by

$$\begin{aligned}\Pr\left(\hat{\mathbf{x}}\neq\mathbf{x}^{(1)}|A\right) &\le \frac{1}{|\mathcal{Q}|^n}\\ &\times\sum_{\mathbf{x}^{(1)}\in\mathcal{Q}^n}\ \sum_{\mathbf{x}^{(2)}\in\mathcal{I},\mathbf{x}^{(2)}\neq\mathbf{x}^{(1)}}\Pr\left(\left\|\mathbf{y}-\mathbf{R}\mathbf{x}^{(2)}\right\|^2\le\left\|\mathbf{y}-\mathbf{R}\mathbf{x}^{(1)}\right\|^2\right). \quad (26)\end{aligned}$$

By using the Chernoff bound for the Q-function to the summand in (26), it can be readily obtained that

$$\begin{aligned}\Pr\left(\hat{\mathbf{x}}\neq\mathbf{x}^{(1)}|A\right) &\le \frac{1}{|\mathcal{Q}|^n}\sum_{\mathbf{x}^{(1)}\in\mathcal{Q}^n}\ \sum_{\mathbf{x}^{(2)}\in\mathcal{I},\mathbf{x}^{(2)}\neq\mathbf{x}^{(1)}}\frac{1}{\left(1+\frac{d_{\min}^2}{4\sigma^2}\right)^n}\\ &\le |\mathcal{I}|\frac{1}{\left(1+\frac{d_{\min}^2}{4\sigma^2}\right)^n}\le\left(\frac{|\mathcal{Q}|}{1+\frac{d_{\min}^2}{4\sigma^2}}\right)^n. \quad (27)\end{aligned}$$

Combining (27) and (25), the FER can be bounded as

$$\begin{aligned}P_f &= \Pr(A^c)+\Pr\left(\hat{\mathbf{x}}\neq\mathbf{x}^{(1)}|A\right)\Pr(A)\\ &\le (1-p)\sum_{i=1}^{n}\frac{|\mathcal{Q}|}{\left(1+\frac{d_{\min}^2}{4\sigma^2}\right)^i}+\left(\frac{|\mathcal{Q}|}{1+\frac{d_{\min}^2}{4\sigma^2}}\right)^n. \quad (28)\end{aligned}$$

APPENDIX B

For the threshold rule, the approach is similar to the analysis of the uniform rule. All the events are defined the same as before. For the threshold pruning rule, the union bound for $P(A_n^c)$ is given by

$$\begin{aligned}P(A_n^c) &\le \underset{r_{n,n}}{E}\ \underset{x_n^{(1)}}{E}\left[\Pr\left(|n_n|^2>\delta_n\right)\right]\\ &= \int_{\frac{\delta_n}{\sigma^2}}^{+\infty} f_n(x)dx = 1-\gamma\left(1,\frac{\delta_n}{2\sigma^2}\right) \quad (29)\end{aligned}$$

where δ_n controls the strength of pruning as in Pruning Rule 2, $f_n(x)$ is the pdf of $\chi^2(2)$, and $\gamma(\alpha,x)$ is the incomplete gamma function. Similarly, we can obtain

$$\Pr\left(A_i^c|\tilde{x}_{i+1}=x_{i+1}^{(1)},\ldots,\tilde{x}_n=x_n^{(1)}\right)\le\int_{\frac{\delta_i}{\sigma^2}}^{+\infty} f_i(x)dx \quad (30)$$

where $f_i(x)$ is the pdf of $\chi^2(2(n-i+1))$. The FER is upper bounded as

$$P_f\le\sum_{i=1}^{n}\int_{\frac{\delta_i}{\sigma^2}}^{+\infty} f_i(x)dx+\left(\frac{|\mathcal{Q}|}{1+\frac{d_{\min}^2}{4\sigma^2}}\right)^n. \quad (31)$$

REFERENCES

[1] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–16, Jan. 1999.

[2] E. Viterbo and J. Bouros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.

[3] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, no. 1–3, pp. 181–199, Aug. 1994.

[4] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[5] M. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[6] A. Burg, M. Borgmann, C. Simon, M. Wenk, M. Zellweger, and W. Fichtner, "Performance tradeoffs in the VLSI implementation of the sphere decoding algorithm," in *Proc. Inst. Elect. Eng.—3G Mobile Commun. Technol. Conf.*, London, U.K., Oct. 2004, pp. 93–97.

[7] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. New York: Cambridge Univ. Press, 2007.

[8] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[9] R. Gowaikar and B. Hassibi, "Statistical pruning for near-maximum likelihood decoding," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2661–2675, Jun. 2007.

[10] B. Shim and I. Kang, "Sphere decoding with a probabilistic tree pruning," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4867–4878, Oct. 2008.

[11] B. Shim and I. Kang, "On further reduction of complexity in tree pruning based sphere search," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 417–422, Feb. 2010.

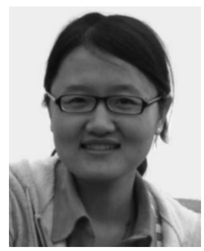
[12] A. Ghaderipoor and C. Tellambura, "A statistical pruning strategy for Schnorr–Euchner sphere decoding," *IEEE Commun. Lett.*, vol. 12, no. 2, pp. 121–123, Feb. 2008.

[13] J. Ahn, H.-N. Lee, and K. Kim, "Schnorr–Euchner sphere decoder with statistical pruning for MIMO systems," in *Proc. 6th ISWCS*, Sep. 2009, pp. 619–623.

[14] S. Lei, Q. Tu, D. Yang, and J. Chen, "Probabilistic tree pruning for fixed-complexity sphere decoder in MIMO systems," in *Proc. Int. Conf. WCSP*, Oct. 2010, pp. 1–6.
[15] S. Han and C. Tellambura, "A complexity-efficient sphere decoder for MIMO systems," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.
[16] S. Lei, X. Zhang, C. Xiong, and D. Yang, "An efficient statistical pruning algorithm for fixed-complexity sphere decoder," *IEICE Trans. Commun.*, vol. E94.B, no. 3, pp. 834–837, 2011.
[17] L. Barbero and J. Thompson, "Fixing the complexity of the sphere decoder for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2131–2142, Jun. 2008.
[18] R. Y. Chang and W.-H. Chung, "Efficient tree-search MIMO detection with probabilistic node ordering," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.
[19] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, Mar. 2006.
[20] J. Maurer, G. Matz, and D. Seethaler, "On the diversity-complexity tradeoff in mimo spatial multiplexing systems," in *Proc. 40th ACSSC*, Nov. 2006, pp. 2077–2081.
[21] J. Maurer, J. Jalden, D. Seethaler, and G. Matz, "Achieving a continuous diversity-complexity tradeoff in wireless MIMO systems via pre-equalized sphere-decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 986–999, Dec. 2009.
[22] S. Han, T. Cui, and C. Tellambura, "Improved K-best sphere detection for uncoded and coded MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 472–475, Oct. 2012.
[23] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2882–2890, Nov. 2003.
[24] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 423–425, May 2005.
[25] M. O. Damen, E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.
[26] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.
[27] W. Zhang, *State Space Search: Algorithms, Complexity and Applications*. New York: Springer-Verlag, 1999.
[28] X. Zhu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 187–191, Feb. 2002.
[29] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill Educ., Dec. 2001.
[30] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
[31] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2701–2710, Jul. 2009.
[32] C.-E. Chen and W.-H. Sheen, "A new lattice reduction algorithm for LR-aided MIMO linear detection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2417–2422, Aug. 2011.
[33] H. Dai and H. Poor, "Iterative space-time processing for multiuser detection in multipath CDMA channels," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2116–2127, Sep. 2002.
[34] S. Han and C. Tellambura, "Complexity-efficient detection for MIMO relay networks," in *Proc. 12th Can. Workshop Inf. Theory*, May 2011, pp. 126–129.

Tao Cui (S'04) received the M.Sc. degree from the University of Alberta, Edmonton, AB, Canada, in 2005 and the M.S. and Ph.D. degrees from California Institute of Technology, Pasadena, in 2006 and 2009, respectively.

His research interests include the interactions between networking theory, communication theory, and information theory.



Shuangshuang Han (S'11) received the B.Eng. degree in communication engineering and the M.Eng. degree in communication and information systems from Shandong University, Jinan, China, in 2006 and 2009, respectively. She is currently working towards the Ph.D. degree in electrical and computer engineering with the University of Alberta, Edmonton, AB, Canada.

Her research interests include signal detection, cooperative communications, and performance analysis of communication systems.

Ms. Han has been awarded a scholarship from the China Scholarship Council for 2009 to 2013.

Chintha Tellambura (F'11) received the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

From 1993 to 1994, he was a Postdoctoral Research Fellow with the University of Victoria and from 1995 to 1996, with the University of Bradford. From 1997 to 2002, he was with Monash University, Caulfield East, Australia. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. His research interests include wireless physical-layer issues, multiple-antenna systems, orthogonal frequency-division multiplexing, relay networks, cognitive radio, and interference in wireless networks.

Dr. Tellambura was the Chair of the Communication Theory Symposium at the 2005 IEEE Global Communications Conference. He was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He received the University of Alberta Killam Professor and McCalla Professor Awards and the Communication Theory Best Paper Award at the 2012 IEEE International Communications Conference.