Maximum Likelihood and Soft Input Soft Output MIMO Detection at a Reduced Complexity

Kai-Jiun Yang and Shang-Ho Tsai, *Senior Member, IEEE*

***Abstract*—This paper proposes an efficient pruning algorithm for MIMO detector, which not only achieves the same performance of the maximum likelihood (ML) decoder, but also massively reduces the number of visiting nodes. The proposed algorithm applies correct sorting order so that it achieves the same performance of the ML decoder. Thanks to the use of the correct ordering, both child and sibling nodes can be pruned efficiently. As a result, the number of visiting nodes is significantly reduced. To acquire properly sorted order, we propose fast sorting algorithms that generate the exact sorted sequence. Furthermore, the proposed schemes can be applied in SISO MIMO detector. The advantages of the proposed sphere decoder become more pronounced in high-order modulation schemes.**

***Index Terms*—ML detector, order sorting, sphere decoding.**

I. Introduction

Maximum likelihood (ML) decoding has been widely researched in multi-input and multi-output (MIMO) systems [1]–[9] to gain excellent performance. The ML decoder searches the entire signal space exhaustively and selects the best solution, such that the complexity grows exponentially as the number of symbols or the data streams increases. To overcome such an issue, sphere decoding (SD) have been proposed for reducing the searching space [3]–[8].

The depth-first [3] sphere decoder sets a sphere constraint (SC), and the searching space is limited within the child nodes that comply with the SC. For fast convergence, an initial SC is suggested which is usually determined by zero-forcing (ZF) [5], minimum-mean-square-error(MMSE) [6], or stochastic channel SNR [7], and the SC is updated whenever a leaf node is visited. As to the breadth-first strategy, e.g., the K-best SD, sets a finite K as its searching range in each layer [8]. Such an architecture is regular and suitable for hardware or parallel implementation. However, the breadth-first SD has BER performance loss compared to that of the ML decoder due to the fixed number of search nodes. Additionally, K is heuristically determined. Therefore, both of the strategies require re-evaluation when either the modulation scheme or the number of streams changes.

Schnorr-Euchner (SE) enumeration [2] sorts the candidates in ascending order based on the Euclidean distances to efficiently converge the SC. However, the sorting complexity increases exponentially as the dimension of the modulation scheme increases. Tabular enumerations [3] that use predefined orders resolve the complexity issue while reducing the performance gap. Lately, compressive sensing detectors [4] focusing on sparse radio-frequency front-end implementations was proposed for signals with sparsity; complexity can be drastically reduced for massive MIMO systems.

The conventional sorting methods suffer from performance degradation. That is, they may not be the correct ordering for all possible combinations of the received vector due to added noise. In this work, we propose methods that achieve the same performance of the ML decoder thanks to the use of correct ordering. The complexity for obtaining the correct ordering has been significantly reduced using the proposed off-line sorting algorithm. As a result, the residual complexity for on-line sorting becomes an easy task. It is worthy pointing out that due to the use of the correct ordering, the pruning can be applied for child as well as sibling nodes, which can again significantly decrease the number of visiting nodes.

The contribution becomes more pronounced when large constellation size such as 256-QAM is applied, or the soft input soft output (SISO) MIMO detection [9]–[11] that requires several decoding iterations.

II. System Model and Background Review

Considering a MIMO system with M_T transmit and receive antennas with channel response $\mathbf{H}$, the transmitted signal vector is $\mathbf{s}$, and the received data vector is $\mathbf{y}$. The relationship between the received and the transmitted signal vectors can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathcal{C}^{M_T \times 1}$ is the noise vector whose elements are independent and identical distributed (i.i.d.) complex Gaussian. The maximum likelihood (ML) detector searches the entire solution space and obtains the optimal solution

$$\mathbf{s}_{ML} = \arg \min_{\mathbf{s} \in \Lambda^{M_T}} \|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 \tag{2}$$

where Λ denotes the modulation scheme with the number of symbols being M. The complexity can be reduced by applying the QR-decomposition $\mathbf{H} = \mathbf{Q}\mathbf{R}$ such that $\hat{\mathbf{y}} = \mathbf{Q}^H \mathbf{y}$. The gist of sphere decoding is to restrict the search within a hypersphere

$$\|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 = \sum_{i=1}^{M_T} \left| \hat{y}_i - \sum_{j=i}^{M_T} R_{ij} s_j \right|^2 < r. \tag{3}$$

The predefined r is the so-called sphere constraint (SC). Since $\mathbf{R}$ is an upper triangular matrix, the decoding starts from the last row and makes a tentative decision for s_{M_T} by minimizing $|\hat{y}_{M_T} - R_{M_T M_T} s_{M_T}|^2$. The computed s_{M_T} is backward substituted to the second last row and the remaining decoding for $s_{M_T-2}, \ldots, s_1$ continues in the same fashion.

A sphere decoder uses partial Euclidean distance (PED) to eliminate the redundant nodes that do not satisfy the SC in (3).

Manuscript received May 18, 2018; revised August 27, 2018; accepted September 24, 2018. Date of publication October 4, 2018; date of current version December 14, 2018. This research was supported by the Ministry of Science and Technology (MOST), Taiwan under Grant MOST 106-2221-E-009-043 and MOST 107-2221-E-009-065 and also by MediaTek Inc. The review of this paper was coordinated by Dr. N.-D. Đào. *(Corresponding author: Shang-Ho Tsai.)*

Kai-Jiun Yang is with the Industrial Technology Research Institute, Zhudong 31057, Taiwan and also with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: kaijiuny.ece98g@g2.nctu.edu.tw).

Shang-Ho Tsai is with the Department of Electrical Engineering,National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: shanghot@alumni.usc.edu).

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.2018.2874044

The PED in layer i is

$$T_i(\mathbf{s}_i) = T_{i+1}(\mathbf{s}_{i+1}) + |e_i(\mathbf{s}_i)|^2$$
$$\text{with } e_i(\mathbf{s}_i) = \hat{y}_i - \sum_{j=i}^{M_T} R_{ij} s_j, \tag{4}$$

where $\mathbf{s}_i = [s_i \ s_{i+1} \ \cdots \ s_{M_T}]^T$, $i = M_T, M_T - 1, \ldots, 1$, and $T_{M_T+1}(\mathbf{s}_{M_T+1})$ is assumed to be 0. When the leaf layer is reached, the PED of current candidate $\mathbf{s}$ equals to the Euclidean distance between the received vector and $\mathbf{s}$,

$$T_1(\mathbf{s}) = \|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 = \sum_{i=1}^{M_T} \left| \hat{y}_i - \sum_{j=i}^{M_T} R_{ij} s_j \right|^2 < r. \tag{5}$$

When a branch violates the SC, *i.e.,* $T_{i+1}(\mathbf{s}_{i+1}) > r$, the traverse switches to the other sibling nodes without evaluating the following child nodes for avoiding redundant computations. As the leaf layer is reached and the accumulated PED is less than the SC, the SC is updated by this accumulated PED.

The soft-input soft-output MIMO detector takes channel observation $\mathbf{y}$ along with *a priori* information $\mathrm{P}[\mathbf{s}]$ to generate intrinsic log-likelihood ratio (LLR) given by [1]

$$L_{i,b} \triangleq \log\left(\frac{\mathrm{P}[x_{i,b} = +1|\mathbf{y}, \mathbf{H}]}{\mathrm{P}[x_{i,b} = -1|\mathbf{y}, \mathbf{H}]}\right). \tag{6}$$

The channel decoder after the detector refines the LLRs and feeds back to the detector as *a priori* information, and such an iteration stops as the predefined iteration rounds are reached. The $x_{i,b} = \{+1, -1\}$ in (6) corresponds to the b^{th} bit of the symbol from the i^{th} antenna in $\mathbf{s}$. Let $\chi_{i,b}^{(+1)}$ and $\chi_{i,b}^{(-1)}$ be the symbol vector sets. To reduce the complexity, the SISO MIMO detector computes intrinsic max-log LLRs

$$L_{i,b}^D \triangleq \min_{\mathbf{s} \in \chi_{i,b}^{(-1)}} \{d(\mathbf{s})\} - \min_{\mathbf{s} \in \chi_{i,b}^{(+1)}} \{d(\mathbf{s})\}, \tag{7}$$

to approximate $L_{i,b}$ in (6), where the path metric of the candidate $\mathbf{s}$ is

$$d(\mathbf{s}) \triangleq \frac{1}{N_0}\|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 - \log \mathrm{P}[\mathbf{s}]. \tag{8}$$

$\mathrm{P}[\mathbf{s}] = \prod_{i=1}^{M_T} \mathrm{P}[s_i|\mathbf{s}^{(i+1)}]$ is a recursive product with $\mathbf{s}^{(i)} \triangleq \{s_i, \ldots, s_{M_T}\}$. Hence, the path metric in (8) can be written as

$$d(\mathbf{s}) = \sum_{i=1}^{M_T} \left(\frac{1}{N_0} \left| \tilde{y}_i - \sum_{j=i}^{M_T} R_{i,j} s_j \right|^2 - \log \mathrm{P}_\mathrm{c}\left[\mathbf{s}^{(i)}\right] \right), \tag{9}$$

where $\mathrm{P}[s_i|\mathbf{s}^{(i+1)}]$ is simplified as $\mathrm{P}_\mathrm{c}[\mathbf{s}^{(i)}]$. Together $d(\mathbf{s})$ can be evaluated recursively with the partial distances (PDs)

$$d_i = d_{i+1} + e_i, \quad i = M_T, \ldots, 1, \tag{10}$$

with the initialization $d_{M_T+1} = 0$ and the distance increment

$$e_i = \frac{1}{N_0} \left| \tilde{y}_i - \sum_{j=i}^{M_T} R_{i,j} s_j \right|^2 - \log \mathrm{P}_\mathrm{c}\left[\mathbf{s}^{(i)}\right]. \tag{11}$$

Additionally, the prior information can be approximated as [9]

$$-\log \mathrm{P}_\mathrm{c}[\mathbf{s}^{(i)}] \approx \sum_{b=1}^{Q} \frac{1}{2}\left(|L_{i,b}^A| - x_{i,b} L_{i,b}^A\right). \tag{12}$$

If the MAP solution for (8) is

$$\mathbf{s}^{\mathrm{MAP}} = \arg \min_{\mathbf{s} \in \mathcal{O}^{M_T}} \frac{1}{N_0}\|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 - \log \mathrm{P}[\mathbf{s}], \tag{13}$$

one of the minima in (7) shall be

$$\lambda^{\mathrm{MAP}} \triangleq \frac{1}{N_0}\|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}^{\mathrm{MAP}}\|^2 - \log \mathrm{P}[\mathbf{s}^{\mathrm{MAP}}], \tag{14}$$

while the other term is

$$\lambda_{i,b}^{\overline{\mathrm{MAP}}} \triangleq \min_{\mathbf{s} \in \chi_{i,b}^{\overline{\mathrm{MAP}}}} \left\{ \frac{1}{N_0}\|\hat{\mathbf{y}} - \mathbf{R}\mathbf{s}\|^2 - \log \mathrm{P}[\mathbf{s}] \right\}, \tag{15}$$

and $\chi_{i,b}^{\overline{\mathrm{MAP}}}$ is referred as the bit-wise counter-hypothesis of the MAP solution. Finally, the SISO MIMO detector generates the extrinsic LLRs as outputs

$$L_{i,b}^E = \begin{cases} \Lambda_{i,b}^{\overline{\mathrm{MAP}}} - \lambda^{\mathrm{MAP}}, & x_{i,b}^{\mathrm{MAP}} = +1 \\ \lambda^{\mathrm{MAP}} - \Lambda_{i,b}^{\overline{\mathrm{MAP}}}, & x_{i,b}^{\mathrm{MAP}} = -1. \end{cases} \tag{16}$$

where the extrinsic metrics is defined as

$$\Lambda_{i,b}^{\overline{\mathrm{MAP}}} = \begin{cases} \lambda_{i,b}^{\overline{\mathrm{MAP}}} - L_{i,b}^A, & x_{i,b}^{\mathrm{MAP}} = +1 \\ \lambda_{i,b}^{\overline{\mathrm{MAP}}} + L_{i,b}^A, & x_{i,b}^{\mathrm{MAP}} = -1. \end{cases} \tag{17}$$

The transform function $f(\cdot)$ that converts a intrinsic metric λ with the corresponding L^A to an extrinsic metric Λ is defined as [9]

$$\Lambda = f(\lambda, L^A, x) \triangleq \begin{cases} \lambda - L^A, & x = +1 \\ \lambda + L^A, & x = -1. \end{cases} \tag{18}$$

When the leaf with the symbol set $\mathbf{x}$ along the way is reached, λ^{MAP} and $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$ are updated based on the comparison between $d(\mathbf{x})$ and λ^{MAP}. If $d(\mathbf{x}) < \lambda^{\mathrm{MAP}}$, a tighter sphere constraint is obtained via updating λ^{MAP} by $d(\mathbf{x})$, and $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$ is renewed to $f(\lambda^{\mathrm{MAP}}, L_{i,b}^A, x_{i,b}^{\overline{\mathrm{MAP}}})$. Otherwise, should the $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$ is greater than $f(d(\mathbf{x}), L_{i,b}^A, x_{i,b}^{\mathrm{MAP}})$, it is updated as such. Without further constraint, all the siblings of the branch at the leaf layer must be visited to secure the minimal $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$.

To maintain the stability of the iteration, $L_{\max}$ is applied for the output extrinsic LLR clipping

$$\lambda_{i,b}^{\overline{\mathrm{MAP}}} \longleftarrow \min\{\lambda_{i,b}^{\overline{\mathrm{MAP}}}, \lambda^{\mathrm{MAP}} + L_{\max}\}, \quad \forall i, b. \tag{19}$$

Additionally, the tree-pruning criterion [9] is applied to eliminate redundant traverse.

III. Proposed Maximum Likelihood Detector

We propose to terminate the evaluation of the sibling nodes conclusively with orderly sorted enumeration, which leads to a significant reduction in the number of visiting nodes. Moreover,

we propose a low-complexity scheme to obtain such an exact enumeration.

From (4), an estimation of s_i in the i-th layer that can be obtained by equalizing the received symbol $\hat{s}_i$

$$\hat{s}_i = \frac{\hat{y}_i - \sum_{j=i+1}^{M_T} R_{ij} s_j}{R_{ii}}. \tag{20}$$

From (3) to (5), we obtain the following expression:

$$T_i(\mathbf{s}_i) = T_{i+1}(\mathbf{s}_{i+1}) + \left| \hat{y}_i - \sum_{j=i}^{M_T} R_{ij} s_j \right|^2 < T_1(\mathbf{s}) < r. \tag{21}$$

From (21), it yields

$$\left| \hat{y}_i - \sum_{j=i+1}^{M_T} R_{ij} s_j - R_{ii} s_i \right|^2 < r - T_{i+1}(\mathbf{s}_{i+1}). \tag{22}$$

Dividing the two sides of (22) by $|R_{ii}|^2$ leads to

$$|\hat{s}_i - s_i|^2 < \frac{r - T_{i+1}(\mathbf{s}_{i+1})}{|R_{ii}|^2} \equiv \alpha_i, \tag{23}$$

where $\hat{s}_i$ is defined in (20), and α_i is positive.

The candidates in the ith layer are enumerated according to the Euclidean distance in ascending order given by

$$|\hat{s}_i - s_{i_1}| < |\hat{s}_i - s_{i_2}| < \cdots < |\hat{s}_i - s_{i_M}|, \tag{24}$$

where $s_{i_l} \in \Lambda$ is the symbol with the lth smallest Euclidean distance. Then if a symbol s_{i_k} violates the constraint in (23), *i.e.,* $|\hat{s}_i - s_{i_k}|^2 > \alpha_i$, one shall terminate the evaluations of all the other sibling nodes $s_{i_{k+1}}$, $s_{i_{k+2}}$, $\cdots$ s_{i_M}, because those points also violate the constraint due to the enumeration in (24). The proposed pruning is even more powerful in the leaf layer. Since there is no child node thereafter, the first candidate in the enumeration is of the minimal accumulated PED among all the siblings, and the SC can be directly updated by the accumulated PED of the first candidate if it is smaller than the SC. Using this pruning method in the other layers also leads to a remarkable reduction, and the BER performance is the same as that of the ML decoder.

Figure 1 shows an example of 16-QAM. The number of symbols has been extended from 16 to 49 to accommodate the corner cases as annotated by the left bottom and right top dot boxes. For M-QAM modulation, the generalized number is $(2\sqrt{M}-1)^2$. The off-line predefine order of the symbols is arranged in ascending Euclidean distances from the symbols to the center O. When the received symbol is identified in the target M-QAM, the boundary can be determined, and excessive symbols are trimmed. Once we solve the ambiguity on-line, the precise order is determined.

Identifying Ambiguity Between Symbols: Off-line Calculation. The evaluation order of two ambiguous symbols $\tilde{l}_{(i)}$ and $\tilde{l}_{(j)}$ can be determined by their perpendicular bisector as shown in Fig. 1. Note that the brace (i) hereafter stands for that i^{th} symbol in the same layer. Symbol $\tilde{l}_{(i)}$ should be evaluated before $\tilde{l}_{(j)}$ if $\tilde{s}$ and $\tilde{l}_{(i)}$ lie on the same side of this perpendicular bisector $b_p(\tilde{l}_{(i)}, \tilde{l}_{(j)})$. With *ABCD* being the decision region that decodes $\tilde{s}$ as $(0,0)$, the ambiguous segments are identified by

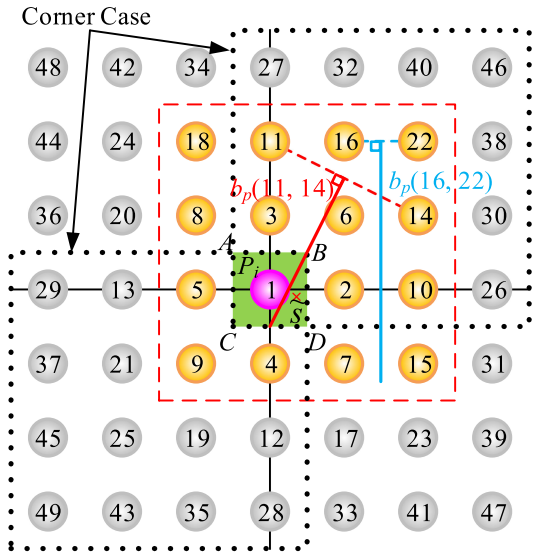

Fig. 1. The extended constellation that covers the corner cases is applied for 16-QAM. The ambiguity among the symbols is identified by using the perpendicular bisectors.

Algorithm 1: Ambiguity Identification.

Input: The extended symbols $\{\tilde{l}_{(i)}\}_{i=1}^{(2\sqrt{M}-1)^2}$.
Output: The ambiguity set $A_{(i)}$, $i = 2, 3, \cdots (2\sqrt{M}-1)^2$.

```
R is the decision region for (0, 0).
for i = 1 : (2√M − 1)^2 − 1
    A_(i) = ∅;
    for j = i + 1 : (2√M − 1)^2
        if b_p(l̃_(i), l̃_(j)) intersects R      // Ambiguity occurs.
            A_(i) = A_(i) ∪ j;
        end
    end
end
```

scanning the candidates with the perpendicular bisectors pair by pair as shown in Alg. 1. The ambiguity set $A_{(i)}$ contains the indices of symbols that have ambiguity with the symbol $\tilde{l}_{(i)}$. We define the number of symbols in $A_{(i)}$ as the **ambiguity length** for symbol $\tilde{l}_{(i)}$. With Alg.1, one can identify the ambiguity sets for the kept 16 symbols as shown in Fig. 2 where the links denote the ambiguity among the symbols, and only 11 pairs of ambiguity need to be resolved on-line. The proposed sorting of all the M symbols is summarized in Algorithm 2. Additionally, the sorting of the segments can be conducted concurrently.

The proposed bound dynamically sets the end of the visit among the siblings. During the decoding, λ^{MAP} in (14) is updated as the smaller sphere constraint is available and so is $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$ in (17) as the smaller counter hypothesis is discovered. The update of $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$ peruses $\mathbf{s}^{(i)}$ bit-by-bit, and pruning the siblings as early as possible can reduce enormous complexity.

The update of (17) is skipped if the newly computed $f(d(\mathbf{x}^\dagger), L_{i,b}^A, x_{i,b}^{\mathrm{MAP}})$ is equal or larger

$$\Lambda_{i,b}^{\overline{\mathrm{MAP}}} \leq \begin{cases} d(\mathbf{x}^\dagger) - L_{i,b}^A, & x_{i,b}^{\mathrm{MAP}\dagger} = +1 \\ d(\mathbf{x}^\dagger) + L_{i,b}^A, & x_{i,b}^{\mathrm{MAP}\dagger} = -1. \end{cases} \tag{25}$$

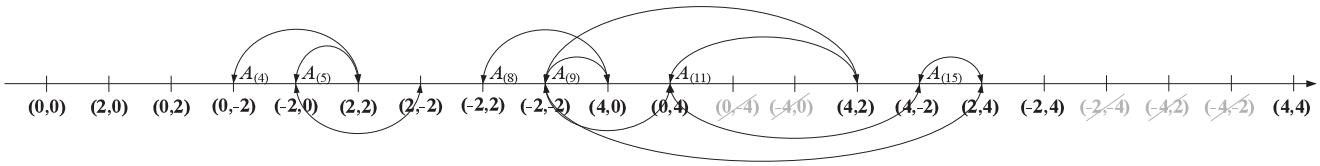

Fig. 2. The ambiguity after trimming. The arrows denote the ambiguity among symbols.

Algorithm 2: Proposed Fast Algorithm for Sorting.

Input: $\mathbf{A}$, M, predefined ordered symbols $\{\tilde{l}_{(i)}\}_{i=1}^{M}$.
Output: Ordered symbols
1: $L = \varnothing$; // Stack for ambiguity solving;
2: **for** $i = 1 : M$
3: $A_{(i:1)} = A_{(i:1)} - \tilde{l}_{(i)}$ // Remove $\tilde{l}_{(i)}$ from $A_{(i)}, \ldots, A_{(1)}$
4: $L = L \cup \tilde{l}_{(i)}$ // Add $\tilde{l}_{(i)}$ to L
5: Solving ambiguity in L by bubble sorting $\tilde{l}_{(i)}$.
6: **while** (first q points in L whose $A_{(i)} = \varnothing$)
7: The order of these q points is finalized.
8: Removing these q points from L.
9: **end**
10: **end**

For the symbols $\mathbf{x}^{\dagger}$ that fit the following criterion

$$d(\mathbf{x}^{\dagger}) \geq \Lambda_{i,b}^{\overline{\mathrm{MAP}}} + |L_{i,b}^{A}|, \tag{26}$$

they cannot be applied for updating $\Lambda_{i,b}^{\overline{\mathrm{MAP}}}$. Note that such a bound can be increased or shrunk as x and $L_{i,b}^{A}$ vary at different indexes i and b. Therefore, the threshold at layer i in the traverse is defined as

$$\theta_i = \max\left(\Lambda^{\overline{\mathrm{MAP}}(i)}\right) + \max\left(|L^{A(i)}|\right), \forall b, \tag{27}$$

where the superscript (i) indicates from the layer M_T to the layer i for all bits. Applying the threshold in (27) to the PD in (10) and (11), the counter hypothesis in the layer i is NOT updated when

$$d(\mathbf{x}^{\dagger(i)}) = d_{i+1} + \frac{1}{N_0}\left|\tilde{y}_i - \sum_{j=i}^{M_T} R_{i,j}s_j\right|^2 - \log \mathrm{P_c}\left[\mathbf{s}^{(i)}\right] \geq \theta_i. \tag{28}$$

From (12), since $x_{i,b} \in \pm 1$, $|L_{i,b}^{A}| - x_{i,b}L_{i,b}^{A}$ can only be 0 or $2|L_{i,b}^{A}|$, $-\log \mathrm{P_c}[\mathbf{s}^{(i)}]$ can only be equal to or larger than 0. Additionally, the d_{i+1} for the comparisons among the candidate sets in the same layer are identical. If $\mathbf{x}^{\dagger(i)}$ coincides with

$$d_{i+1} + \frac{1}{N_0}\left|\tilde{y}_i - \sum_{j=i}^{M_T} R_{i,j}s_j\right|^2 \geq \theta_i, \tag{29}$$

$\mathbf{x}^{\dagger(i)}$ is no longer the valid counter hypothesis and is excluded in the following traverse as elaborated in Fig. 3. The conventional tree-pruning needs to check the symbols in the same layer one after another due to lack of the ambiguity maps as shown in Fig. 2. With such a crucial hint, the proposed sibling-pruning vastly reduces unnecessary visits, especially in the large constellation level.

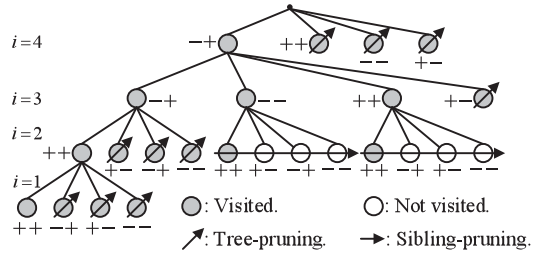

Fig. 3. The tree-pruning eliminates one node per validation. As to the proposed sibling-pruning, if one node fits the pruning threshold, the node and those afterwards are all eliminated.

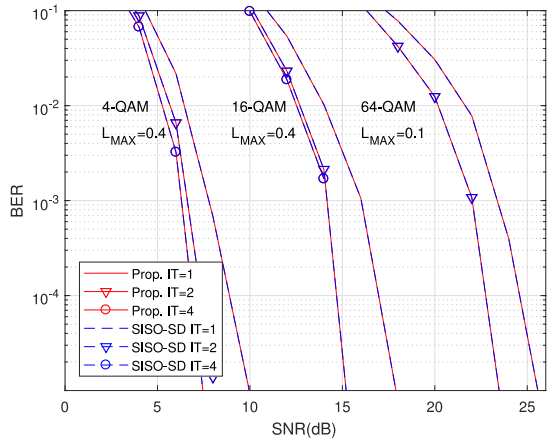

Fig. 4. The BER performance as a function of SNR for soft-decision 4×4 MIMO detector.

IV. Simulation Results

The proposed decoding scheme has been evaluated in a 4×4 MIMO channel with Gaussian noise. Note that the MIMO dimension is scalable since the proposed pruning is independent among different layers. The benchmarking ML algorithm and the K-best SD were compared with the proposed hard-decision scheme, while the SISO-SD [9] was compared with the proposed soft-decision scheme.

Experiment 1: **Performance Comparisons of Various Decoding Schemes.** The BER performance of the proposed SD is compared with typical ML for hard-decision and the SISO-SD [9] for soft-decision. To mitigate the error propagation that degrades the performance of the applied models, sorting the channel matrix $\mathbf{R}$ has been applied. The BER of the proposed scheme in hard-decision is exactly the same as that in ML

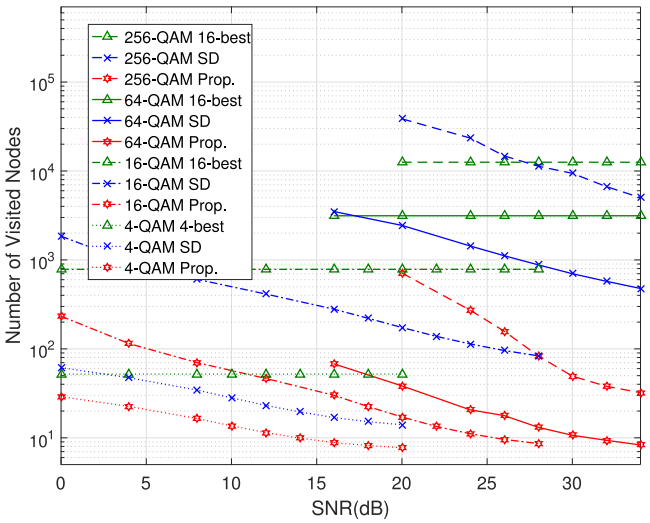

Fig. 5. The number of visiting nodes of hard-decision 4 × 4 MIMO detector.

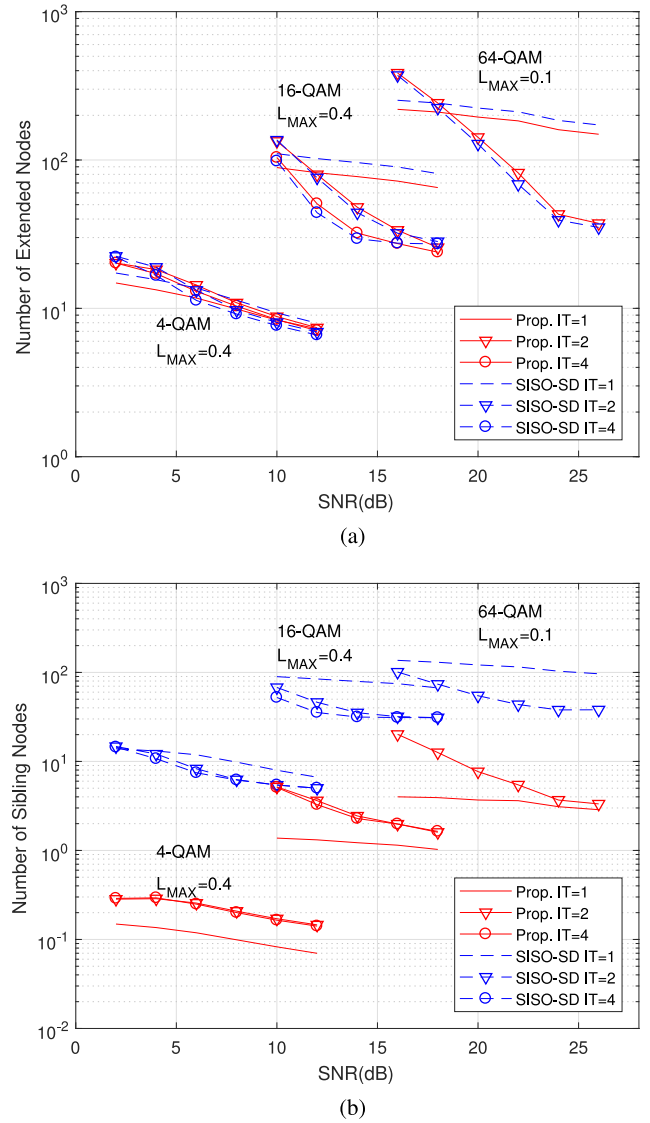

Fig. 6. The numbers of (a) extended and (b) sibling nodes of soft-decision 4 × 4 MIMO detector.

decoder, while the BER in SISO MIMO detection is identical to that of SISO-SD in [9] as shown in Fig. 4

Experiment 2: **Complexity Comparisons of Various Decoding Schemes.** Figure 5 shows the number of visiting nodes as a function of SNR in hard-decision MIMO detector. The complexity of the proposed SISO MIMO detection is as shown in Fig. 6. The extended nodes are the visits that vertically travel down to the leaf and update the sphere constraint in (14), while the sibling nodes are those horizontal peer-compares in (15) that renew the minimal soft-information. The efforts for evaluating a sibling node is heavier than that for evaluating an extended node, since evaluating a sibling node requires evaluating the soft information of all bits using (15), while evaluating an extended node only cares if the sphere constraint is met in (14). Consequently, the complexity is dominated by the sibling nodes.

V. Conclusion

A pruning method for MIMO detector that achieves the same performance as that in ML has been proposed in this work. The reduction in the number of visiting nodes becomes more pronounced in higher modulation. Moreover, we have proposed fast algorithms to reduce the real-time computational complexity in obtaining the correct enumeration. We have found that most of computations can be conducted in a off-line manner and thus the residual effort for on-line computations can be greatly reduced. The proposed MIMO decoder is a promising solution for advanced modulation schemes that are applied in IEEE 802.11ax and LTE, and is applicable in SISO MIMO detectors.

References

[1] 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.

[2] C. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, pp. 181–199, Sep. 1994.

[3] K.-J. Yang, S.-H. Tsai, R.-C. Chang, Y.-C. Chen, and G. C.-H. Chuang, "VLSI implementation of a low complexity 4 × 4 MIMO sphere decoder with table enumeration," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 2167–2170.

[4] L. Xiao *et al.*, "Efficient compressive sensing detectors for generalized spatial modulation systems," *IEEE Trans. Veh. Tech.*, vol. 66, no. 2, pp. 1284–1298, Feb. 2017.

[5] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[6] L. Dai *et al.*, "Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems," *IEEE Trans. Veh. Tech.*, vol. 64, no. 10, pp. 4839–4845, Oct. 2015.

[7] T. Cui, S. Han, and C. Tellambura, "Probability-distribution-based node pruning for sphere decoding," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1586–1596, May 2013.

[8] M.-Y. Huang and P.-Y. Tsai, "Toward multi-gigabit wireless: design of high-throughput MIMO detectors with hardware-efficient architecture," *IEEE Trans. Circuits Syst. I*, vol. 61, no. 2, pp. 613–624, Feb. 2014.

[9] C. Studer and H. Bolcskei, "Soft-input soft-output single tree-search sphere decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4827–4842, Oct. 2010.

[10] E. P. Adeva and G. P. Fettweis, "Efficient architecture for soft-input soft-output sphere detection with perfect node enumeration," *IEEE Trans. VLSI Syst.*, vol. 24, no. 9, pp. 2932–2945, Sep. 2016.

[11] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.