A Successive Interference Cancellation Algorithm in MIMO Systems via Breadth-First Search

Yongtao Su
Department of Automation
Tsinghua University
Beijing, China 100084
Email: ysu@ee.columbia.edu

Xian-Da Zhang
Department of Automation
Tsinghua University
Beijing, China 100084
Email: zxd-dau@tsinghua.edu.cn

Xiaodong Wang
Department of Electrical Engineering
Columbia University
New York, NY 10027
Email: wangx@ee.columbia.edu

***Abstract*—A successive interference cancellation (SIC) algorithm based on breadth-first search (BFS) is developed to achieve a soft-input soft-output detector via the tree structure of the MIMO system model in this paper. Instead of visiting all nodes of the tree, the proposed BFS-SIC algorithm only browses and extends those paths with large metrics. If paths are enough, the performance of BFS-SIC algorithm can approach that of sphere decoding but is much more flexible due to its providing a good tradeoff between complexity and performance. Moreover, the BFS-SIC algorithm possesses path metrics including only scalar operations rather than matrix operations. Simulation results demonstrate the effectiveness of the proposed algorithm.**

***Index Terms*—Multiple input multiple output (MIMO), successive interference cancellation (SIC), breadth-first search (BFS), tree structure**

I. Introduction

Spatial multiplexing (SM) schemes of multiple-input multiple-output systems (MIMO) provides a linear increased capacity with no additional power expenditure [1]. Detection algorithms for SM schemes have been extensively studied in the literatures, such as sphere decoding (SD) [2]–[4], ordered successive cancellation (SIC) [5], M-algorithm [6] and the List-Sequential (LISS) algorithm [7].

Most detection algorithms can be classified as tree search decoding [8]. The well known SD algorithm computes the maximum-likelihood symbol estimate antenna by antenna as an integer least-squares problem [3], [4]. From the view of intelligent search theory, SD is in fact a type of depth-first search (DFS) with heuristic knowledge regarding the transmitted antennas as the levels of a tree. The most probable symbol of the next level is firstly extended as the offspring node of the current level if the symbol is in a hypersphere, or else remounts to its parent node. The drawback of SD is the exponential complexity particularly in the low signal-to-noise ratio (SNR) region due to the incessant retrospect among the levels.

In order to avoid the variational complexity of SD, a soft-input soft-output (SISO) algorithm, called SGA algorithm, is developed with a fixed complexity [9]. By using the output of decorrelator, it sequentially calculates the posterior distribution by some symbol combinations with the largest probability metrics. If the number of symbol combinations is large enough, SGA is near-optimum and can achieve a comparable performance with SD. However, the computational load of SGA is too huge to be implemented in real applications.

Technical speaking, SGA is actually a kind of M-algorithm which is an example of heuristic breadth-first search (BFS). BFS algorithms are naturally suitable for SISO framework. Moreover, its complexity usually remains constant independent of the channel conditions. In this paper, we propose a novel BFS-based algorithm via QR decomposition and SIC. The BFS-SIC algorithm possesses a near optimal performance but has significant complexity saving comparing to SD and SGA algorithm.

This paper is organized as follows. Section II describes the signal model of MIMO system and its tree structure. Section III provides the basic algorithm of BFS-SIC and some comparison with SGA. In section IV, simulation results are described to demonstrate the effectiveness of the proposed BFS-SIC algorithm. Finally, a brief conclusion is given in section V.

II. System Description

A. Signal Model

Consider an MIMO system with n_T transmit antennas and n_R receive antennas ($n_T \leq n_R$) in a rich scattering flat fading environment. The input-output relation of SM scheme can be put in the form of the linear signal model

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{s} \triangleq [s_1, \cdots, s_{n_T}]^T$ is a transmitted vector whose entries are chosen from a finite alphabet set $\mathcal{A} = \{a_1, \cdots, a_N\}$, $\boldsymbol{r} \triangleq [r_1, \cdots, r_{n_R}]^T$ is a received vector, $\boldsymbol{H}$ is an $n_R \times n_T$ complex fading channel matrix known perfectly to the receiver, and $\boldsymbol{v} \sim \mathcal{N}_c(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is a zeros-mean complex circular symmetric Gaussian noise vector with $\boldsymbol{I}$ denoting the identity matrix of appropriate dimension. To simplify the narration, we assume that QAM is employed in this paper.

This work was supported by the Major Program of the National Natural Science Foundation of China under Grant 60496311, and funded by Basic Research Foundation of Tsinghua National Laboratory for Information Science and Technology (TNList).

For a squared QAM signal set, an equivalent form of the model (1) can be expressed in a real matrix form as

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{w} \tag{2}$$

where $\boldsymbol{y} = [\mathrm{Re}(\boldsymbol{r}^T), \mathrm{Im}(\boldsymbol{r}^T)]^T$, $\boldsymbol{x} = [\mathrm{Re}(\boldsymbol{s}^T), \mathrm{Im}(\boldsymbol{s}^T)]^T$ and $\boldsymbol{w} = [\mathrm{Re}(\boldsymbol{v}^T), \mathrm{Im}(\boldsymbol{v}^T)]^T$ are $\tilde{n}_R \times 1$, $\tilde{n}_T \times 1$ and $\tilde{n}_R \times 1$ real vectors, and

$$\boldsymbol{B} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{H}) & -\mathrm{Im}(\boldsymbol{H}) \\ \mathrm{Im}(\boldsymbol{H}) & \mathrm{Re}(\boldsymbol{H}) \end{bmatrix}$$

is a $\tilde{n}_R \times \tilde{n}_T$ real matrix with $\tilde{n}_T = 2n_T$ and $\tilde{n}_R = 2n_R$. Note that the entries of vector $\boldsymbol{x}$ are taken from a pulse amplitude modulation (PAM) signal set $\mathcal{X} = \{u_i = 2i - 1 - M,\ i = 1, 2, \cdots, M\}$ with $M = \sqrt{N}$ for the squared QAM [10].

Denoting the QR decomposition $\boldsymbol{B} \triangleq \boldsymbol{Q}\boldsymbol{R}$ with the orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{\tilde{n}_R \times \tilde{n}_R}$ and the upper trapezoidal matrix $\boldsymbol{R} \in \mathbb{R}^{\tilde{n}_R \times \tilde{n}_T}$, there exists a unique factorization $\boldsymbol{B} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$ where $\widetilde{\boldsymbol{Q}}$ and $\widetilde{\boldsymbol{R}}$ are submatrices of $\boldsymbol{Q}$ and $\boldsymbol{R}$ given respectively by $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}(1:\tilde{n}_R, 1:\tilde{n}_T)$ and $\widetilde{\boldsymbol{R}} = \boldsymbol{R}(1:\tilde{n}_T, 1:\tilde{n}_T)$. Based on the reduced QR decomposition, the signal model (2) can be rewritten as

$$\underbrace{\begin{bmatrix} \widetilde{y}_1 \\ \vdots \\ \widetilde{y}_{\tilde{n}_T} \end{bmatrix}}_{\widetilde{\boldsymbol{y}}} = \underbrace{\begin{bmatrix} r_{11} & \cdots & r_{1\tilde{n}_T} \\ & \ddots & \vdots \\ & & r_{\tilde{n}_T\tilde{n}_T} \end{bmatrix}}_{\widetilde{\boldsymbol{R}}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_{\tilde{n}_T} \end{bmatrix}}_{\boldsymbol{x}} + \underbrace{\begin{bmatrix} \widetilde{w}_1 \\ \vdots \\ \widetilde{w}_{\tilde{n}_T} \end{bmatrix}}_{\widetilde{\boldsymbol{w}}} \tag{3}$$

where $\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{Q}}^H \boldsymbol{y}$ and $\widetilde{\boldsymbol{w}} = \widetilde{\boldsymbol{Q}}^H \boldsymbol{w} \sim \mathcal{N}_c(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Note that QR decomposition is inefficiency when $\boldsymbol{B}$ is ill-conditioned. Some preprocessing approaches in [8] can tackle this problem and lead to much more efficient decoding algorithms but the detail discussion falls outside the scope of this paper.

According to the model (3), the objective of SM receiver is to make Bayesian inference from the received vector $\widetilde{\boldsymbol{y}}$ and estimate the symbols of each transmit antenna. In other words, it evaluates the marginal posterior distributions $p(x_k|\widetilde{\boldsymbol{y}})$ by summarizing $M^{\tilde{n}_T - 1}$ possible values of $\boldsymbol{x}$ given by $\widetilde{\boldsymbol{y}}$,

$$p(x_k|\widetilde{\boldsymbol{y}}) = \sum_{\boldsymbol{x}_{/k}} p(\boldsymbol{x}|\widetilde{\boldsymbol{y}}), \tag{4}$$

where $\boldsymbol{x}_{/k}$ is the transmitted vector except the k-th antenna. The sum in (4) is prohibitive when a large number of transmit antennas or mutlilevel/phase modulation is employed.

B. Tree Structure

Fortunately, thanks to the upper triangular structure of $\widetilde{\boldsymbol{R}}$ in the signal model (3), the transmitted vector $\boldsymbol{x}$ can be expressed as paths in a tree of depth $\tilde{n}_T$, where the estimation is performed in accordance with the inverse order of $\boldsymbol{x}$. Tree search strategies in the literature can be classified into BFS (e.g. M-algorithm, SGA), DFS (e.g. SD) and Best-First search (e.g. stack algorithm) [8]. In general, BFS naturally adapts to SISO structure which is suitable for acting as a detector in turbo receiver. Moreover, the complexity of BFS usually remains constant independent of variations in SNR and channel conditions. In this paper, we develop a novel low-complexity BFS algorithm with adaptive updating the metrics.

III. BFS-SIC Algorithm

A. Basic Algorithm

BFS visits all possible symbols of a level and computes their probabilities given their ancestor nodes before moving to the next level. This process carries out from the root to the last level and finally chooses the most probable path as the optimal estimate of $\boldsymbol{x}$. However, the exhaustive search makes the original BFS infeasible in MIMO systems. Being aware that only a few paths dominate the posterior distribution, we select J nodes in each level, whose corresponding J paths have the largest probabilities, to limit the computational load. That is to say, according to the model (3), the probability metric of the j-th path $\alpha_k^{(j)}$ $(j = 1, \cdots, J)$ at the k-th transmitted antenna is given by

$$\begin{aligned} \alpha_k^{(j)} &= p(x_k, x_{k+1:\tilde{n}_T}^{(j)} | \widetilde{y}_{k:\tilde{n}_T}) \\ &\propto p(\widetilde{y}_{k:\tilde{n}_T} | x_k, x_{k+1:\tilde{n}_T}^{(j)}) p(x_{k+1:\tilde{n}_T}^{(j)}, x_k) \\ &= \alpha_{k+1}^{(j)} \cdot p(\widetilde{y}_k | x_{k+1:\tilde{n}_T}^{(j)}, x_k) p(x_k) \\ &\propto \alpha_{k+1}^{(j)} \cdot \exp\left\{ -\frac{1}{\sigma^2} \left| \widetilde{z}_k^{(j)} - r_{kk} x_k \right|^2 \right\} p(x_k) \end{aligned} \tag{5}$$

where $x_{k+1:\tilde{n}_T}^{(j)} \triangleq \{x_{k+1}^{(j)}, \cdots, x_{\tilde{n}_T}^{(j)}\}$, $\widetilde{y}_{k:\tilde{n}_T} \triangleq \{\widetilde{y}_k, \cdots, \widetilde{y}_{\tilde{n}_T}\}$ and $\widetilde{z}_k^{(j)} \triangleq \widetilde{y}_k - \sum_{l=k+1}^{\tilde{n}_T} r_{kl} x_l^{(j)}$. From (5), we acquire NJ candidate metrics (corresponding to NJ nodes) for the k-th transmit antenna (corresponding to the $(\tilde{n}_T - k + 1)$-th level of the tree). As mentioned above, J of them are selected and affiliated to their respective paths before the next level is taken into account. Of particular note is the sequential form of the metric calculation from the last antenna to the first one by reason of $\widetilde{\boldsymbol{R}}$ being an upper diagonal matrix. Moreover, the expense of (5) is very low due to the scalar computing.

After the last level is processed, J full probability metrics, $\alpha^{(j)} \triangleq \alpha_1^{(j)} = p(x_{1:\tilde{n}_T}^{(j)} | \widetilde{y}_{1:\tilde{n}_T}), j = 1, \cdots, J$, are obtained for J paths. The marginal posterior probability $p(x_k|\widetilde{\boldsymbol{y}})$ in (4) can be further computed by using the generated paths as follows

$$\begin{aligned} P(x_k = u_i | \widetilde{\boldsymbol{y}}) &= \sum_{j=1}^{J} P\left(x_k = u_i, \boldsymbol{x}_{/k}^{(j)} | \widetilde{\boldsymbol{y}} \right) \\ &\propto \sum_{j=1}^{J} \exp\left(-\frac{1}{\sigma^2} \left\| \widetilde{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{x}_{k,i}^{(j)} \right\|^2 \right) \\ &\quad \times P(x_k = u_i) \prod_{l \neq k} P\left(x_l = x_l^{(j)} \right) \end{aligned} \tag{6}$$

where $\boldsymbol{x}_{k,i}^{(j)} = [x_1^{(j)}, \cdots, x_{k-1}^{(j)}, u_i, x_{k+1}^{(j)}, \cdots, x_{\tilde{n}_T}^{(j)}]^T$. In addition, the hard output can be also obtained by either the MAP estimate from (6) or computing

$$\hat{x}_k = \min_{x_k \in \mathcal{X}} |x_k - \sum_{j=1}^{J} \alpha^{(j)} x_k^{(j)}|^2, \quad k = 1, \cdots, \tilde{n}_T \tag{7}$$

for simplicity.

B. Algorithm Summary

To summarize, the proposed BFS-SIC algorithm for MIMO systems is given below.

1) Initialization:
   - Execute the reduced QR decomposition using $\boldsymbol{B} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$ and calculate $\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{Q}}^H \boldsymbol{y}$.
   - For $k = \tilde{n}_T$ and $x_{\tilde{n}_T} \in \mathcal{X}$, compute

$$\beta_{\tilde{n}_T,i} = \exp\left\{-\frac{1}{\sigma^2}\left|\widetilde{y}_{\tilde{n}_T} - r_{\tilde{n}_T\tilde{n}_T}u_i\right|^2\right\} \times P(x_{\tilde{n}_T} = u_i). \quad (8)$$

   Let $\{\alpha_{\tilde{n}_T}^{(j)}\}_{j=1}^{J_1}$ be the largest $J_1 = \min\{J, N\}$ elements of $\{\beta_{\tilde{n}_T,i}\}_{i=1}^{N}$. The corresponding J_1 symbols form the path set $\{x_{\tilde{n}_T}^{(j)}\}_{j=1}^{J_1}$.

2) For $k = \tilde{n}_T - 1, \cdots, 1$, do the following operations:
   - For $j = 1, \cdots, J_1$ and $x_k \in \mathcal{X}$, compute

$$\beta_{k,i}^{(j)} = \alpha_{k+1}^{(j)} \cdot \exp\left\{-\frac{1}{\sigma^2}\left|\widetilde{z}_k^{(j)} - r_{kk}u_i\right|^2\right\} \times P(x_k = u_i). \quad (9)$$

   - Let $\{\alpha_k^{(j)}\}_{j=1}^{J_2}$ be the largest $J_2 = \min\{J, N^{\tilde{n}_T-k+1}\}$ elements of $\{\beta_{k,i}^{(j)}\}_{i=1,j=1}^{N,J_1}$. The corresponding J_2 symbols form the path set $\{x_{k:\tilde{n}_T}^{(j)}\}_{j=1}^{J_2}$.
   - Set $J_1 = J_2$.

3) Calculate the marginal posterior probabilities.
   - Supposing that the normalized probability metrics $\alpha^{(j)} = \alpha_1^{(j)}, j = 1, \cdots, J$ are in descent order, we construct the cumulative distribution function (CDF) as follows.
     – Initialize the CDF: $c_0 = 0$ and $m = 0$.
     – While $c_m \leqslant \delta$, do
       ∗ $m = m + 1$.
       ∗ $c_m = c_{m-1} + \alpha^{(m)}$.
   - Perform (6) to obtain the marginal posterior probabilities or (7) to obtain the hard output except that the sum is from $j = 1$ to m.

Remark 1: In step 3), since only a few paths or candidate vectors in $\{\boldsymbol{x}^{(j)}\}_{j=1}^{J}$ dominate the posteriori distribution, it is possible to select only m paths with the largest probability metrics to calculate (6), which significantly reduces the whole computational load.

Remark 2: The BFS-SIC algorithm is somewhat similar to the M-algorithm in [6] and the SGA algorithm in [9] because they both employ the breadth-first search strategy by extending the most possible paths first and then selecting those with the largest metrics. However, there are following three differences between them:

- The metrics for each possible symbol combinations in M-algorithm and SGA involve matrix computations whereas, as shown in (5), only scalar computations are required in BFS-SIC due to its adaptive updating of $\alpha_k^{(j)}$.

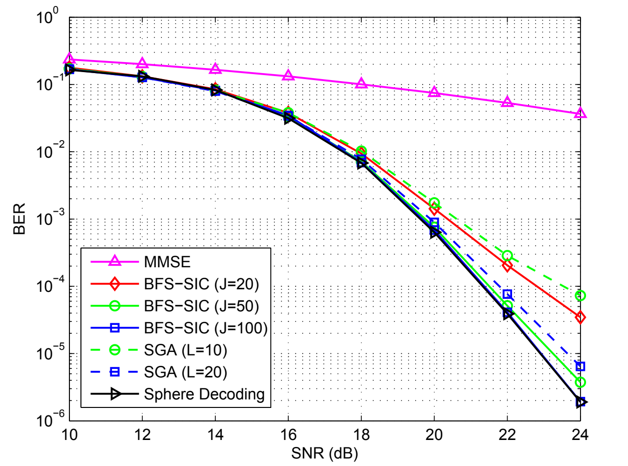

Fig. 1. BER performance of MMSE, BFS-SIC, SGA and sphere decoding algorithm in an uncoded 8×8 MIMO system with 16-QAM.

- SGA operates on the QAM alphabet set and hence its complexity is $\mathcal{O}(n_T JN)$ for recursions. On the other hand, the complexity of BFS-SIC is $\mathcal{O}(2n_T J\sqrt{N})$ due to the operation on the PAM signal set, which greatly speeds up the algorithm. For example, the complexity of BFS-SIC is only half of that of SGA with 16-QAM. In M-algorightm, multilevel bit mapping (MBM) technique is employed to reduce the complexity by partition the QAM constellation. The complexity of M-algorithm is roughly $\mathcal{O}(2n_T J\log_2 N)$ without considering its complexity of metric updates. Actually, the key idea of MBM is to consider a constellation as a tree and constellation points as its leaf nodes, which can be used in not only QAM but also PAM. In other words, by appropriately partition the bit pairs of PAM points, BFS-SIC can also achieve the same number of recursions as M-algorithm.
- SGA intends to use all J paths to estimate the marginal posterior distributions whereas BFS-SIC only browses a very small part of them by computing the CDF especially in high SNR region.

Remark 3: The BFS-SIC algorithm is constructed in a SISO manner and hence can be easily embedded into iterative (turbo) receiver.

IV. Numerical Results

In this section, we demonstrate the performance of BFS-SIC in an 8×8 MIMO system with 16-QAM modulation by comparing to the sphere decoder in [4], [12], SGA in [9] and conventional MMSE detector. The elements of channel matrix $\boldsymbol{H}$ are i.i.d complex Gaussian random variables, i.e. $h_{k,p} \sim \mathcal{N}_c(0, 1/n_T)$. In the simulation, SNR is defined as $E\{\|\boldsymbol{Hs}\|^2\}/E\{\|\boldsymbol{v}\|^2\} = \sigma_s^2/\sigma^2$ where σ_s^2 is the variance of the QAM alphabet set $\mathcal{A}$.

Above all, we demonstrated the performance of different detectors in both uncoded system and coded system. The path numbers are $J = 20, 50, 100$ and the threshold δ is set

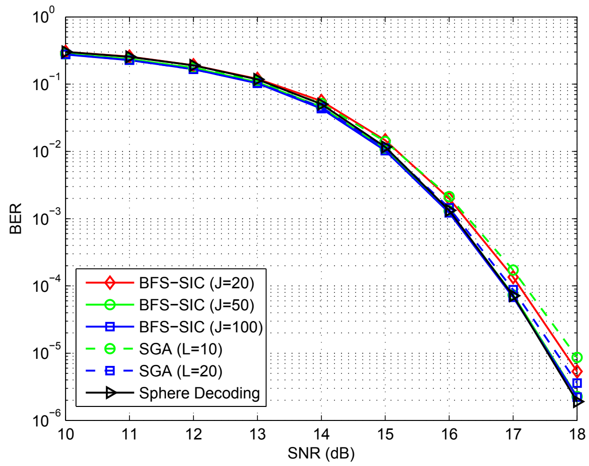

Fig. 2. BER performance of BFS-SIC, SGA and sphere decoding algorithm in a coded 8×8 MIMO systems with 16-QAM. Rate-$1/2$ constraint length-5 convolutional code is used.

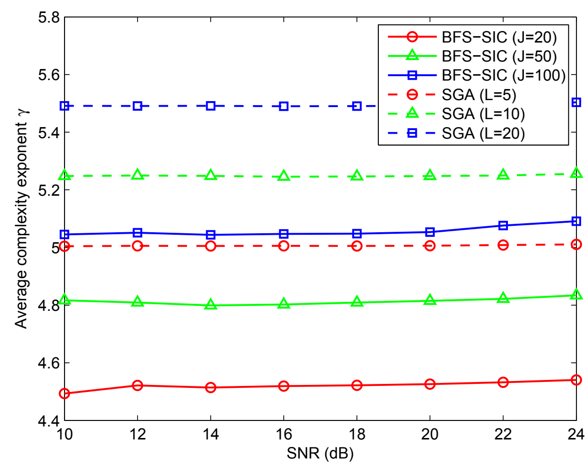

Fig. 3. The average complexity of BFS-SIC and SGA in an uncoded 8×8 MIMO system with 16-QAM.

to be 0.99 in BFS-SIC. For comparison, the path numbers are $L = 5, 10, 20$ in SGA. For a coded MIMO system, a rate $1/2$ convolutional code with constraint length 5 and generator $(23, 35)$ in octal form is employed. The coded bits are random interleaved and then mapped by QAM modulation. Although both BFS-SIC and SGA are designed in SISO manner, we assume no iterations between the detector and the convolutional decoder in the receiver, and only the hard decisions of the detector are fed to the convolutional decoder. Fig. 1 and Fig. 2 show the bit error ratio (BER) performance of detectors in uncoded system and coded system respectively. It is seen that the more the path number is, the more accurate estimates of transmitted bits are. No obvious difference can be aware among BFS-SIC with $J = 100$, SGA with $L = 20$ and sphere decoding in an uncoded MIMO system. However, in a coded system, there is a gap between BFS-SIC and SGA. The performance of BFS-SIC with $J = 50$ is as good as sphere decoding algorithm and is even superior to that of SGA with $L = 20$.

Next, we describe the complexity comparison between BFS-SIC and SGA. As in [11], the average complexity exponent

$$\gamma = \log(\text{average number of flops})/\log \tilde{n}_T$$

is used as a measure for complexity where the flops are averaged over 2000 runs in the simulations. Since the complexity of Max-log-MAP sphere decoder [12] is much larger than that of SGA, as reported in [9], its complexity is not included in the comparison. In Fig. 3, the complexity of BFS-SIC with $J = 100$ is apparently only 30% of that of SGA with $L = 20$ although they can all approach the near optimum BER performance. Note that if only consider the hard decision of detector, DFS such as SD algorithms in [4] have a lower average complexity than BFS [8].

V. Conclusion

Following the search algorithm, this paper developed a BFS-SIC algorithm for MIMO systems which utilizes the tree structure of the signal model after QR decomposition and then employs a breadth-first search to generate multiple paths for the computation of marginal posterior distribution. Thanks to the scalar form of path metrics and some considerations for limit search, BFS-SIC has much lower complexity than SGA and maintains near optimum BER performance.

References

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press 2003.

[2] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463-471, Apr. 1985.

[3] E. Agrell, et al., "Closest point search in lattice," *IEEE Trans. Information Theory*, vol. 48, pp. 2201-2214, Aug. 2002.

[4] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Information Theory*, vol. 49, pp. 2389-2402, Oct. 2003.

[5] G. D. Golden, et al., "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, pp. 14-16, Jan. 1999.

[6] Y. de Jong and T. J. Willink, "Iterative tree search detection for MIMO wireless systems," *IEEE Trans. Communications*, vol. 53, pp. 930-935, June 2005.

[7] J. Hagenauer and C. Kuhn, "The List-Sequential (LISS) algorithm and its application," *IEEE Trans. Communications*, vol. 55, pp. 918-928, May 2007.

[8] A. D. Murugan, et al., " A unified framework for tree search decoding: rediscovering the sequential decoder," *IEEE Trans. Information Theory*, vol. 52, pp. 933-953, March 2006.

[9] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian approximation based mixture reduction for near optimum detection in MIMO systems," *IEEE Commun. Lett.*, vol. 9, pp. 997-999, Nov. 2005.

[10] J. G. Proakis, *Digital Communications (Fourth Edition)*, McGraw-Hill, 2001.

[11] H. Vikalo and B. Hassibi, "On the sphere-decoding algorithm II. Generalizations, second-order statistics, and applications to communications," *IEEE Tans. Signal Processing*, vol. 53, no. 8, Aug. 2005.

[12] M. S. Yee, "Max-log-MAP sphere decoder," in *Proc. IEEE ICASSP*, pp. 1013-1016, Mar. 2005.