# Low-complexity Prediction Techniques of K-best Sphere Decoding for MIMO Systems

Hsiu-Chi Chang, Yen-Chin Liao, and Hsie-Chia Chang

Department of Electronics Engineering
National Chiao Tung University,
1001 Ta-Hsueh Road, Hsinchu, Taiwan, R.O.C.
Tel: +886-3-5712121 ext.54246 email: jasper.ee94g@nctu.edu.tw

***Abstract*— In multiple-input multiple output (MIMO) systems, maximum likelihood (ML) detection can provide good performance, however, exhaustively searching for the ML solution becomes infeasible as the number of antenna and constellation points increases. Thus ML detection is often realized by $K$-best sphere decoding algorithm.**

**In this paper, two techniques to reduce the complexity of $K$-best algorithm while remaining an error probability similar to that of the ML detection is proposed. By the proposed $K$-best with predicted candidates approach, the computation complexity can be reduced. Moreover, the proposed adaptive $K$-best algorithm provides a means to determine the value $K$ according the received signals. The simulation result shows that the reduction in the complexity of 64-best algorithm ranges from 48% to 85%, whereas the corresponding SNR degradation is maintained within 0.13dB and 1.1dB for a 64-QAM $4 \times 4$ MIMO system.**

## I. Introduction

Recently, multiple-input multiple-out (MIMO) systems are applied in many wireless applications for better transmission efficiency and signal quality due to the inherent diversity gain provided by the multi-path environment. Maximum-likelihood (ML) sequence detection is one of the detection schemes for detecting the received signals in MIMO systems. By searching for the constellation point nearest to the received signal, ML detection is optimized for minimizing the symbol error probabilities, but exhaustive search becomes infeasible since the computation complexity grows as the number of antenna or the constellation points increases. Sphere decoding (SD) algorithm can reduce the computation complexity by confining the number of constellation points to be searched, Fincke-Pohst [1] and Schnorr-Euchner [2] are two of the most common computationally efficient search strategies for realizing the ML detection. Nevertheless, the difficulties in hardware implementation arise because of the non-constant computation complexity and decoding throughput. Alternatively, K-best SD algorithm [3], [4] simplifies the hardware implementation of SD algorithm by keeping at most K best paths in each layer, leading to fixed-throughput and predictable complexity. Note that the term layer refers to the signal constellations of an transmit antenna. However, K-best SD algorithm can not guarantee ML performance since the ML path might be eliminated due to the *breadth-first* nature of K-best SD search approach. Thus the value of K should be large enough, and the value K dominates the performance and computation complexity.

In this paper, two modified K-best SD algorithms are proposed for reducing the computation complexity while remaining the performance similar to ML detection. The *K-best algorithm with predicted candidates*, one of our proposed methods, reduces the computation by only computing a fraction of the paths before selecting the K best candidates. Moreover, an adaptive $K$-best algorithm is proposed, providing an adaptive selection of K by observing the ratio of the second minimum and minimum of all paths at the first decoding layer. According to our simulation results, the proposed techniques can achieve at most 85% complexity reduction when comparing to conventional 64-best SD algorithm.

The rest of this paper is organized as the following. The system model, SD algorithm, and K-best SD algorithm are briefly described in Section II. In Section III, the two proposed detection schemes are presented. The bit error probabilities of the proposed schemes are simulated in a $4 \times 4$ MIMO system of uncorrelated flat-fading channels, and the simulation results and comparisons are given in Section IV. Finally, Section VI concludes this work.

## II. Sphere Decoding for MIMO System

For a MIMO system with $N_T$ transmit antennas and $N_R$ receive antennas, the transmitted and received signals can be represented by

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

where $\tilde{\mathbf{y}}$ is the $N_R \times 1$ received complex signals, $\tilde{\mathbf{H}}$ is an $N_R \times N_T$ matrix of independent and identical distributed (i.i.d.) circular Gaussian random variables (flat fading is assumed), $\tilde{\mathbf{s}}$ is an $N_T \times 1$ complex vector representing the signals transmitted by each transmit antenna, and $\tilde{\mathbf{n}}$ is the $N_R \times 1$ i.i.d. complex Gaussian noise vector. Moreover, the complex model in (1) is often described by the equivalent real-valued representation, which is

$$\begin{aligned}
\mathbf{y} &= \begin{bmatrix} Re\{\tilde{\mathbf{y}}\} \\ Im\{\tilde{\mathbf{y}}\} \end{bmatrix} \\
&= \begin{bmatrix} Re\{\tilde{\mathbf{H}}\} & -Im\{\tilde{\mathbf{H}}\} \\ Im\{\tilde{\mathbf{H}}\} & Re\{\tilde{\mathbf{H}}\} \end{bmatrix} \begin{bmatrix} Re\{\tilde{\mathbf{s}}\} \\ Im\{\tilde{\mathbf{s}}\} \end{bmatrix} + \begin{bmatrix} Re\{\tilde{\mathbf{n}}\} \\ Im\{\tilde{\mathbf{n}}\} \end{bmatrix} \\
&= \mathbf{Hs} + \mathbf{n}. \qquad (2)
\end{aligned}$$

This is also referred to as the *real value decomposition*. For QAM signals, real value decomposition transforms the complex constellation into two real-valued PAM constellations, which can result to fewer computation.

For detecting the received signals, maximum likelihood (ML) sequence detection is one of the MIMO system detection technique that optimizes the symbol error probability. According to the system model described in, Fig.1 ML detection is equivalent to searching for the vector $\hat{s}$ that minimizes $||\mathbf{y} - \mathbf{H}s||^2$. That is,

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathbf{\Omega}} ||\mathbf{y} - \mathbf{H}s||^2 , \qquad (3)$$

where $\Omega$ is the set consisting of all possible $2N_t$-dimensional signal constellation points. Fig.1 shows the simplified block diagram of a MIMO receiver. The channel estimator provides the required channel state information $\mathbf{H}$. By QR decomposition, the channel matrix $\mathbf{H}$ is decomposed by $\mathbf{H} = \mathbf{QR}$, and (3) can be rewritten as

$$\begin{aligned}
||\mathbf{y} - \mathbf{Hs}||^2 &= (\mathbf{s} - \mathbf{s_{zf}})^H \mathbf{H}^H \mathbf{H} (\mathbf{s} - \mathbf{s_{zf}}) \\
&+ \mathbf{y}^H (I - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{y}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mathbf{s}} &= \arg\min_{\mathbf{s}\in\mathbf{\Omega}} (\mathbf{s} - \mathbf{s_{zf}})^H \mathbf{H}^T \mathbf{H} (\mathbf{s} - \mathbf{s_{zf}}) \\
&= \arg\min_{\mathbf{s}\in\mathbf{\Omega}} \bar{\mathbf{s}}^H \mathbf{R}^H \mathbf{R} \bar{\mathbf{s}}. \qquad (4)
\end{aligned}$$

Note that the matrix $\mathbf{R}$ derived from QR decomposition is an upper triangular matrix with non-negative diagonal elements, and $H^H H = R^H R$. Moreover, $\mathbf{s_{zf}}$ is the zero-forcing (ZF) solution that can be derived by $\mathbf{s_{zf}} = \mathbf{H}^+\mathbf{y}$ for $\mathbf{H}^+$ is the pseudo-inverse of $\mathbf{H}$. It is perceived that $\bar{\mathbf{s}} = \mathbf{s} - \mathbf{s_{zf}}$ is the distance from the candidates of signal to the ZF solution.

Due to the triangular form of $\mathbf{R}$, we can rewrite (4) as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathbf{\Omega}} \sum_{i=1}^{N_R} \left\| y_i - \sum_{j=i}^{N_T} R_{ij} s_j^{(i)} \right\|^2 , \qquad (5)$$

where $R_{ij}$ and $s_j$ denote the $i$-th row, $j$-th column of $\mathbf{R}$ and the $j$-th element of $\mathbf{s}$. Moreover, we can define $e(\mathbf{s}^{(i)})$, the partial square Euclidean distance(PED) of the $i$-th layer, by

$$e(\mathbf{s}^{(i)}) = \left\| y_i - \sum_{j=i}^{N_T} R_{ij} s_j^{(i)} \right\|^2 , \qquad (6)$$

where $\mathbf{s}^{(i)} = [s_i^{(i)} s_{i+1}^{(i)} \cdots s_{N_T}^{(i)}]^T$ and $s_j^{(i)}$ is the $j$-th element of $\mathbf{s}^{(i)}$. Then the accumulated Euclidean distance corresponding to the candidate $\mathbf{s}^{(i)}$ can be derived recursively from the PED and the accumulated Euclidean distance corresponding to $\mathbf{s}^{(i+1)}$, denoted by $T(\mathbf{s}^{i+1)})$, that is

$$T(\mathbf{s}^{(i)}) = T(\mathbf{s}^{(i+1)}) + e(\mathbf{s}^{(i)}). \qquad (7)$$

The detection process starts from i=$N_T$, resulting to a tree-structure, or called depth-first, search strategy. However, exhaustively searching for the ML solution becomes infeasible [5] since the computation complexity grows exponentially with $N_t$ or the number of constellation points. Thus, sphere decoding (SD) algorithm has been proposed and recognized as a powerful means to solve the ML detection problems [6] [4]. SD algorithm reduces the computation by restricting the search range. Instead of searching all candidates in $\mathbf{\Omega}$, SD algorithm constrains a much smaller search range $\mathbf{\Omega_{SD}} = \{\mathbf{s} : \bar{\mathbf{s}}^H \mathbf{R}^H \mathbf{R} \bar{\mathbf{s}} \leq r^2\}$; only the candidates in $\mathbf{\Omega_{SD}}$ will be compared. By the aforementioned procedure, the candidate of the smallest $T(\mathbf{s}^{(1)})$ is always the ML solution as long as $r$ is properly defined. However, not only the value $r$, but the computation varies with SNR, leading to a non-constant decoding throughput. Hardware implementation of SD algorithm becomes complicated.

$K$-best SD algorithm is an alternative method that improves the decoding throughput. It simplified the original SD algorithm and maintains a constant throughput by keeping only the $K$ smallest accumulated PED at each layer. However, $K$-best SD algorithm can not guarantee the performance of ML detection since the ML solution may be eliminated when it is not of the $K$ best accumulated PEDs. Thus, larger $K$ is required and the value $K$ becomes a tradeoff between complexity and error performance.

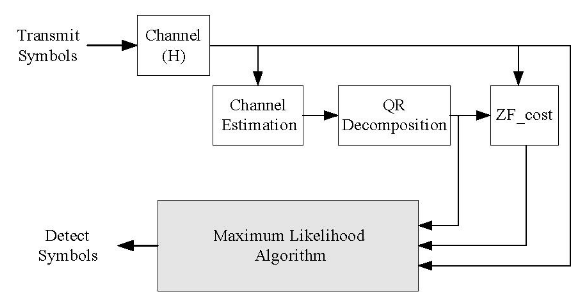



Fig. 1. Block diagram of MIMO detection

Fig. 2 illustrates the bit error rate of a $4 \times 4$ MIMO detector of different values of $K$, and there is performance degradation when $K$ is chosen too small.

## III. Proposed K-best SD Algorithm with Predicted Candidates

Although $K$-best SD algorithm remains constant throughput and computation, its computation complexity is not necessarily lower than the conventional SD algorithm since all the PEDs of each layer still need to be calculated. However, only the $K$ PEDs resulting to the $K$ best accumulated PEDs can affect the PED calculation in the next decoding layer. That is, part of computations of the PEDs are unnecessary. A method to

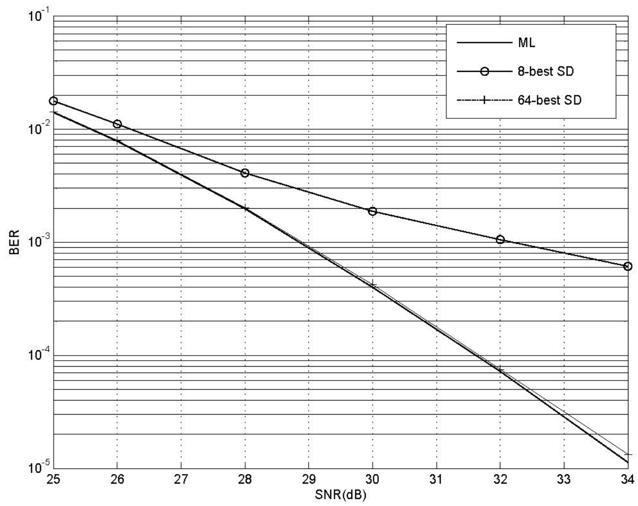



Fig. 2. Comparisons of ML and K-best SD algorithm

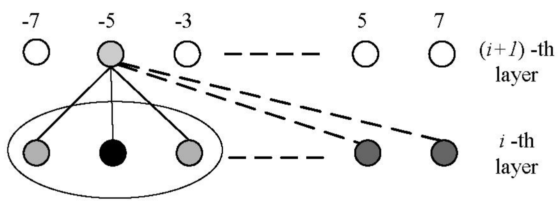



Fig. 3. K-best with predicted candidates

predict the more likely PEDs is presented in the following. Only a fraction of the PEDs are computed, and thus, the computation can be greatly reduced.

At decoding layer $i$, the point $\hat{s}_i$ resulting in the smallest PED for a given $\mathbf{s}^{(i+1)}$ can be derived by

$$\hat{s}_i^{(i+1)} = \frac{y_i - \sum_{j=i+1}^{N_T} R_{ij} s_j^{(i+1)}}{R_{ii}}, \tag{8}$$

and only the $L-1$ points nearest to $\hat{s}_i^{(i+1)}$ will be computed for $e(\mathbf{s}^{(i)})$. That is, the $s_i^{(i)}$ of the vector $\mathbf{s}^{(i)}$ will be $\hat{s}_i^{(i+1)}$ and its $L-1$ nearest constellation points. Only $L$ PEDs from $e(\mathbf{s}^{(i+1)})$ should be calculated instead. Accordingly, we can always have the PED values computed in an ascending order, and the first $L$ smallest PEDs will contribute to more likely candidates. Fig.3 is a 64-QAM example with $L = 3$. The constellation corresponds to the $i$-th layer is denoted as $S_i$, as the figure shows, the points with of cross mark is the $\hat{s}^{(i+1)}$, and only the three constellation points (linked by solid lines) will be computed. Thus, the computation complexity can be reduced, especially when $N_T$ is large.

## IV. Proposed Adaptive $K$-Best Sphere Decoding Algorithm

In the previous section, a scheme was proposed to predict the more likely PEDs. Instead of $K$, only $L$ PEDs are required to be computed for the given $\mathbf{s}^{(i+1)}$. However, the error probability arises when $L$ is chosen too small, and $L$ should be increased to reduce this degradation. Alternatively, determining a proper $K$ value is another way to reduce complexity and error probability. Due to fading, the signals suffer from low SNR when they are in deep fades, and $K$ should be chosen larger. Contrarily, smaller $K$ is sufficient when the signal strength is high. Dynamic $K$ implies an signal quality indicator is required.

A technique for supporting dynamic $K$ which is referred as adaptive $K$-best algorithm , provides a means to observe the required signal quality. For a MIMO system of $N_T$ transmit antennas, this indicator can be acquired by the ratio

$$R = \frac{M_2}{M_1}, \tag{9}$$

where $M_2$ and $M_1$ are the second minimum and minimum of the $N_t$-th decoding layer, respectively. It can be observed that when the value $R$ is below some threshold, the probability of the ML path being eliminated during the $K$-best SD processing increases.

Fig.4 is an illustrative example of a $4 \times 4$ 64-QAM system, which shows the relation between $T$ and the symbol error probability conditioned on the value $T$. The curve stands for the probability $Pr(R < T)$, and the histogram shows the the conditional symbol error probability. It is perceived that symbol error probability is small as $T$ increases. Thus, the value $K$ can be determined by first computing $R$ in (9), then

$$K = \begin{cases} K_1 & \text{if } R \leq T; \\ K_2 & \text{otherwise.} \end{cases} \tag{10}$$

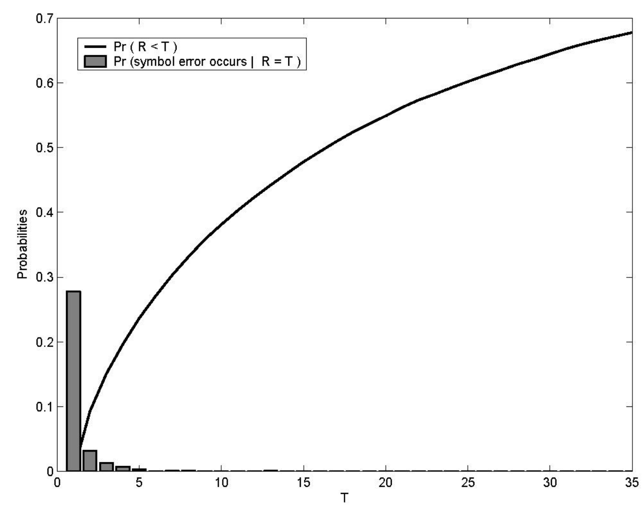



Fig. 4. The probability of $R < T$ and the conditional symbol error probability.

The value $R$ can be regarded as a signal quality indicator of the visited signals. In fact, at each decoding layer, there is always a corresponding $R$, and the layer number in which $R$ is determined becomes a tradeoff between computation complexity and performance. If $R$ is determined at the first few decoding layers, the computation of the rest of the decoding layers can be reduced if $K = K_2$ is chosen. However, if $R$ is determined earlier, there are chances that $R$ cannot provide sufficient information to report the signal quality and the performance will degrade.

## V. Simulation Results

In this section, a $4 \times 4$ MIMO system is simulated for comparing the proposed schemes and the conventional SD and $K$-best algorithms ($K = 64$), whereas the ML detection provides a performance baseline. The signal is modulated by 64-QAM and the MIMO channel is assumed to fade uncorrelatedly and independently. Totally $10^6$ bits are simulated when the SNR is below 30dB, and $10^7$ bits are simulated for SNR $\geq 30dB$.

The proposed adaptive $K$-best algorithm can be applied with the above mentioned candidate prediction technique, whereas the $K_1$ and $K_2$ can have distinct $L_1$ and $L_2$ values, respectively. Fig.5 presents the error probabilities versus SNR for different detection methods. It is perceived that for SNR lower or equal to 30 dB, all the proposed schemes can provide performance very close to that of the ML detection. When SNR is greater than 30dB, a slight degradation is shown, and the value $L$ dominates the degradation. As shown in Fig. 5, for $K_1 = K_2 = 64$, the one with $L_1 = L_2 = 8$ outperforms the one with $L_1 = L_2 = 3$.

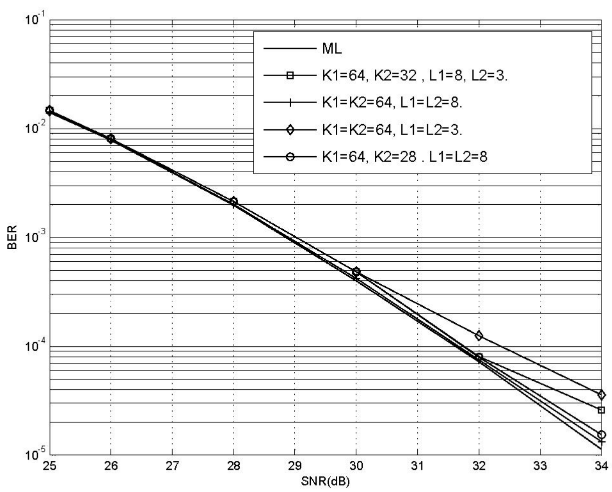
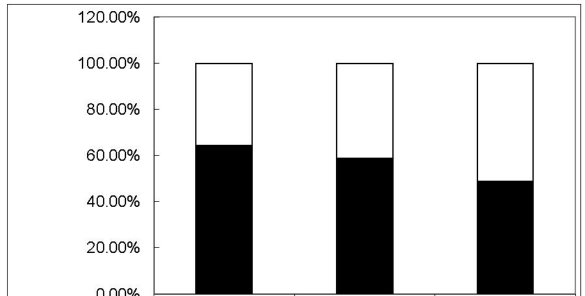


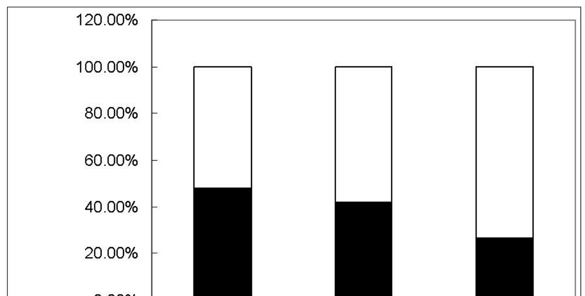

Fig. 5. BER comparisons of different detection schemes

The value $T$ provides a tradeoff between the complexity and error probability. Since smaller $K_2$ may lead to performance degradation in high SNR, a larger $T$ will be required. On the other hand, Fig. 4 shows that symbol error probability drops when $T > 10$. Accordingly, we compare the two cases $K_1 = 64, K_2 = 32, T = 15$ with $L_1 = 8, L_2 = 3$ and $K_1 = 64, K_2 = 28, T = 30$ with $L_1 = L_2 = 8$, whereas the parameters chosen will result to similar computation complexities. As Fig. 5 shows, the latter results to slightly smaller error probabilities. Thus, it can be observed that the value $L$ affect error probability. The maximum value of L is the dimension of the PAM constellation. Smaller L will reduce computation effort, however, the performance will also degrade since some computation is ignored.



| SNR(dB)/Adaptive K-best | 30 | 32 | 34 |
|---|---|---|---|
| □ K2=28,L2=8 | 35.54% | 41.18% | 51.37% |
| ■ K1=64,L1=8 | 64.46% | 58.82% | 48.63% |

Fig. 6. Reduce computation effort in SNR = 30, 32, and 34dB for T = 30.



| SNR(dB)/Adaptive K-best | 30 | 32 | 34 |
|---|---|---|---|
| □ K2=32,L2=3 | 52.17% | 58.02% | 73.52% |
| ■ K1=64,L1=8 | 47.83% | 41.98% | 26.48% |

Fig. 7. Reduce computation effort in SNR = 30, 32, 34dB for T = 15.

Fig.6 and Fig.7 shows the percentage of $K_1$ and $K_2$ are selected for SNR = 30, 32, and 34 dB. As the SNR increases, the percentage of $K_2$ being selected also increases, and more computation complexity can be reduced. For all detection schemes, sorting always contributes the most to the overall computation complexity. Thus, the number of sorting operations are recorded and shown in TABLE I for comparing the complexities. The normalized sorting complexity refers to the number of sorting operation af all methods normalized to that of the conventional 64-best SD algorithm. The table shows that the reduction in the complexity of 64-best algorithm ranges from 48% to 85%, whereas the corresponding SNR degradation is maintained within 0.13dB and 1.1dB for a 64-QAM $4 \times 4$ MIMO system.

## VI. Conclusion

Two techniques reducing the complexity of $K$-best SD algorithm for signal detection in MIMO systems are presented. By the proposed $K$-best algorithm with predicted candidates reduces the number of sorting operation. Moreover, the proposed adaptive $K$-best SD algorithm provides a means to

TABLE I

COMPARISON OF ML AND K-BEST SPHERE DECODING AND RATIO SPHERE DECODING DESIGN

| Method | ML | $K_1 = K_2 = 64$ $L_1 = L_2 = 8$ | $K_1 = 64, K_2 = 28$ $L_1 = L_2 = 8$ | $K_1 = 64, K_2 = 32$ $L_1 = 8, L_2 = 3$ | $K_1 = K_2 = 64$ $L_1 = L_2 = 3$ |
|---|---|---|---|---|---|
| Number of Sorting Operations | $1.19 \times 10^{19}$ | $6.59 \times 10^{10}$ | $3.43 \times 10^{10}$ | $1.9 \times 10^{10}$ | $9.39 \times 10^{9}$ |
| Normalized Sorting Complexity | $1.8 \times 10^{8}$ | 100% | 52.04% | 28.83% | 14.2% |
| SNR (dB) for BER $= 5 \times 10^{-4}$ | 32.64 | 32.72 | 32.85 | 33.24 | 33.82 |

determine the value $K$ by observing the received signals. These two schemes can be applied at the same time when considering the error probability and complexity, providing flexibility and tradeoff between system performance and implementation cost. According to our simulation results, the reduction in the complexity of 64-best algorithm ranges from 48% to 85%, whereas the corresponding SNR degradation is maintained within $0.13$dB and $1.1$dB for a 64-QAM $4 \times 4$ MIMO system.

## REFERENCES


[1] U. Fincle and M. Phost, "Improved methods for calculating vectors for short length in a lattice,includeing complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, April. 1985.

[2] C. Schnorr and M. Euchnerr, "Lattice basis reduction: improved practical algorihtms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181–191, 1994.

[3] K. W. Wong, C. Y. Tsui, R. S. K. Cheng, and W. H. Mow, "A vlsi architecture of a k-best lattice decoding algorithm for mimo channels," *PIMRC*, vol. 02, 2002.

[4] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 1639–1642, July. 1999.

[5] D. Micciancio, "The hardness of the closest vector problem wtih preprocessing," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 1212–1215, March. 2001.

[6] E. Angrell, T.Eriksson, A.Vardy, and K.Zeger, "Close point search in lattices," *IEEE Trans. on Inform. Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.