Analysis of a fixed-complexity sphere decoding method for Spatial Multiplexing MIMO

M. HARIDIM and H. MATZNER

Department of Communication Engineering
HIT-Holon Institute of Technology
52 Golomb St., Holon 58102
ISRAEL
mharidim@hit.ac.il http://www.hit.ac.il

V. NEDER and D. EZRI

Runcom Technologies LTD,
Rishon Lezion, 75658
ISRAEL
vadimne@runcom.co.il http://www.runcom.com

Abstract- A new detection algorithm for decoding multiple-input multiple-output (MIMO) transmission is proposed and analyzed. The method is based on combining sphere decoding (SD) and zero forcing (ZF) techniques. The proposed method performs a fixed number of operations to detect the signal, independent of the noise level, and hence provides a fixed complexity near optimal performance. The algorithm is especially suited for systems with a large number of transmit antennas and allows efficient implementation in hardware. The high efficiency of this algorithm is obtained by limiting the number of overall SD iterations. Moreover, in the proposed method matrices with high condition number are more likely to undergo SD.

Index Terms- MIMO, detection complexity, spatial multiplexing, sphere decoding, zero forcing, maximum-likelihood.

1 Introduction

Multiple-input multiple-output (MIMO) wireless communication systems are capable to provide very high data rates, without requiring an increased transmission bandwidth. It is believed that future wireless communication systems will incorporate a large number of transmit and receive antennas. IEEE802.16 and WiMax standards are already discussing future user terminals and base stations characterized by a large number of antenna arrays. This kind of setup altogether with high rate QAM modulation schemes make the maximum likelihood (ML) decoding algorithms of MIMO systems a very challenging issue for future hardware implementation. The complexity of the ML detection for MIMO systems grows exponentially with the number of transmit and receiving antennas as well as the constellation points. Actually the ML decoding algorithm is based on searching for the closest point in the lattice to the received vector.

One of the most promising transmission methods for MIMO is spatial multiplexing (SM). In SM, the transmitter endowed with M transmit antennas, transmits M independent information streams, one from each antenna. In this case the receiver, endowed with $N \geq M$ receive antennas, is to decode the transmitted information streams. It is known that the optimal solution to the decoding of SM signals is ML, which involves exhaustive search in case of multiple dimensions used in high-performance applications [1]. Spherical decoding (SD) is an iterative method for the computation of the ML estimator in SM MIMO [2]. Like the ML algorithm, SD finds the lattice points closest to the received vector, but the search is limited to the points located inside a sphere centered at the received vector, leading to a significant reduction of the decoding complexity. SD offers a computationally efficient decoding algorithm with ML performance. However, one of the most severe problems in the implementation of SD lies in the fact that the number of iterations per realization is neither defined nor bounded. Thus, usually, SD methods are not suitable for hardware implementation. Several works have been dedicated to develop SD with fixed complexity ([3]-[6]).

In the K-Best lattice decoder with breadth-first tree search was presented. This method uses the breadth-first tree search technique which introduces fixed throughput. In this method, the best K candidates, which have the smallest overall Euclidian distance, are kept at each search so that that a fixed amount of nodes are visited each time. Main disadvantage of this method is that the K parameter cannot be defined analytically and it is highly dependent of the channel conditions.

In [5] the authors propose the depth-first tree search SD. This is a straightforward way of enforcing a run-time constraint to terminate the search, on a symbol vector by symbol vector basis, after a maximum number of visited nodes. The detector then returns the best solution found so far, i.e., the current ML and counter-hypotheses. The detecting performance of this method can be degraded in case of bad channel conditions.

In[5] an unconstrained list sphere detector with a search method that is bounded independently per search level is proposed. The bound is determined based on the distribution of the candidates found in each search level for the large number of detected sub-carriers. It is shown that the search process cannot be bounded for the first search level without a substantial performance loss. This method exploits the main idea of [4] but with lower upper bound, also it doesn't provide the constant rate, but only bounds it.

In [6] it is shown that diversity achieving schemes may be devised by combination of the low complexity zero forcing (ZF) algorithm and ML detection. This method is based on division of the channel matrices into 2 sets according to the condition number. Matrices with condition number lower than a predefined threshold are ZF decoded, while the others are ML decoded. However, this result does not allow hardware implementation of SD for the ML estimates since, again, the number of iterations is not defined. Moreover, the threshold based technique implies receiver calibration, which should be recalibrated for different channel conditions.

In order to further reduce the complexity of SD, we propose a new decoding algorithm, based on combining SD and ZF. The main idea behind the proposed algorithm is to limit the number of overall SD iterations such that the matrices with high condition number, identified as more problematic ones, will be fully SD decoded. In this manner we assure high efficiency performance. This paper is organized as follows. In Section II the SM system is briefly introduced. In Section III we consider different implementations of the SD algorithms, namely, the zero-forcing (ZF), ML, and SD algorithms. Section IV describes the principles of the proposed algorithm. Section V presents and discusses the simulations results. Conclusions are given in Section VI.

2 SM System Description

The mathematical model for the received vector $\mathbf{y}$ in the case of SM is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \rho\mathbf{v} \qquad (1)$$

where

$$\mathbf{y} = \left[y_1 \;\; y_2 \ldots y_N\right]^T$$
$$\boldsymbol{s} = \left[s_1 \;\; s_2 \ldots s_M\right]^T$$
$$\boldsymbol{v} = \left[v_1 \;\; v_2 \ldots v_N\right]^T,$$

$\mathbf{H}$ is the channel matrix, ρv is a vector of independent complex valued Gaussian random variables each with variance ρ^2. In case of 2×2 MIMO:

$$\begin{matrix} y_1 \\ y_2 \end{matrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \rho \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

A schematic illustration of the SM scenario is given in Fig. *1*.

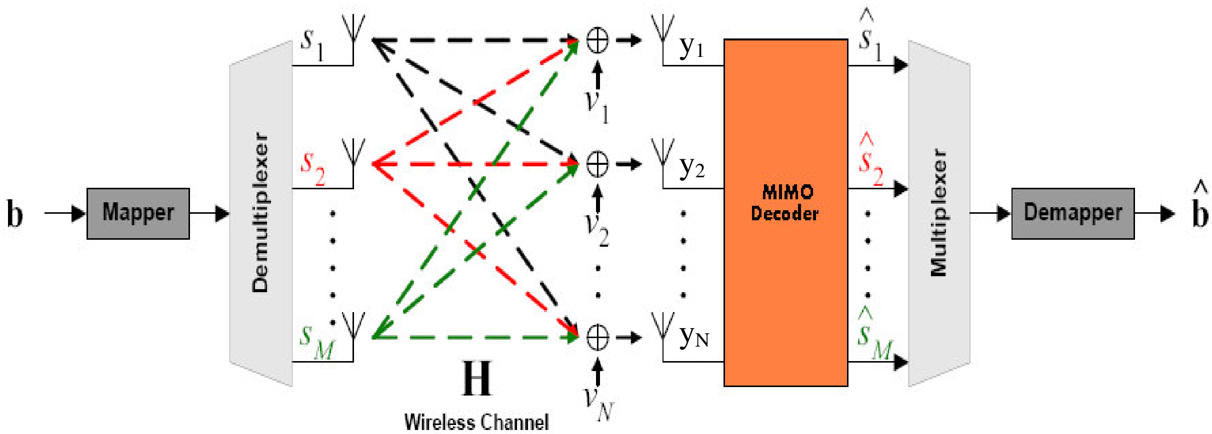

Fig. 1- Schematic representation of an SM system

3 Review of Prominent SM Decoding Algorithms

3.1 Zero Forcing

The linear zero forcing (ZF) algorithm computes the least square estimator [8]

$$\hat{\mathbf{s}}_{\mathbf{ZF}} = \mathbf{H}^{+}\mathbf{y} \qquad (2)$$

where $\mathbf{H}^{+}$ denotes the left pseudo-inverse of **H**. The estimator $\hat{\mathbf{s}}_{\mathbf{ZF}}$ undergoes standard processing as in the single input single output case (SISO). The complexity of finding the ZF estimate is essentially determined by the complexity of finding the pseudo-inverse of matrix **H** in Eq. (1). In the case of 2×2 MIMO, assuming that the transmission power is 1W, it can be shown that the SNR at any receiving antenna is given by:

$$SNR_1=\frac{S_1}{N_1}=\frac{1}{\rho^2 v^2 \cdot \left[\left(H_{11}^{-1}\right)^2+\left(H_{12}^{-1}\right)^2\right]}$$

This equation shows that a relatively high spatial correlation between the antennas results in a small values of H^{-1} ($\det\{H\}$is small in this case), which in turn leads to high noise around any lattice point and hence a low SNR.

For large matrices, the simplest way of calculating the pseudo-inverse is by means of QR factorization, **H** = **QR**. It can also be calculated in a more stable way (which avoids inverting the upper triangular matrix **R**) by means of the singular value decomposition (SVD) of **H**. The ZF algorithm is not optimal in the case of MIMO, but remains attractive due to its low implementation complexity. The problem with the ZF approach is evident when the channel matrix **H** is ill conditioned (has a large condition number), corresponding to strong correlation between the channels. In this case, the entries of $\mathbf{H}^{+}$ in Eq. (**שגיאה! מקור ההפניה לא נמצא.**) are large. This leads to large noise at the output of the ZF estimator. The ZF solution provides diversity order of $N-M+1$ and array gain of $\frac{N-M+1}{M}$.

3.2 Maximum Likelihood

We now address the optimal ML decoder for SM. In this case, the ML detection comes down to finding the log-likelihood ratio (LLR):

$$LLR(b) = \log\frac{\Pr\{b=1\,|\,\mathbf{y}\}}{\Pr\{b=0\,|\,\mathbf{y}\}}$$

Applying Bayes' formula we obtain:

$$LLR(b)=\frac{\Pr\{b=1|\mathbf{y}\}}{\Pr\{b=0|\mathbf{y}\}}=\frac{\frac{p\{\mathbf{y}|b=1\}\cdot\Pr\{b=1\}}{p\{\mathbf{y}\}}}{\frac{p\{\mathbf{y}|b=0\}\cdot\Pr\{b=0\}}{p\{\mathbf{y}\}}}=\frac{\sum_{\mathbf{s}:b=1}p\{\mathbf{y}|\mathbf{s}\}}{\sum_{\mathbf{s}:b=0}p\{\mathbf{y}|\mathbf{s}\}}$$

Using Eq. (1) we obtain:

$$LLR(b)=\frac{\sum_{s:b=1}e^{-\frac{\|\mathbf{y}-\mathbf{Hs}\|^2}{\rho^2}}}{\sum_{s:b=0}e^{-\frac{\|\mathbf{y}-\mathbf{Hs}\|^2}{\rho^2}}}$$

which can be approximated as:

$$LLR(b)=\log\frac{e^{-\min_{s:b=1}\left(\frac{\|\mathbf{y}-\mathbf{Hs}\|^2}{\rho^2}\right)}}{e^{-\min_{s:b=1}\left(\frac{\|\mathbf{y}-\mathbf{Hs}\|^2}{\rho^2}\right)}}=\frac{1}{\rho^2}\left(-\min_{s:b=1}\|\mathbf{y}-\mathbf{Hs}\|^2+\min_{s:b=0}\|\mathbf{y}-\mathbf{Hs}\|^2\right)$$

Finally, we end up with SM decoder definition:

$$\mathbf{s}_{ML}=\arg\min_{\mathbf{s}\in\Gamma}\|\mathbf{y}-\mathbf{Hs}\|^2 \quad (3)$$

This solution implies exhaustive search and therefore it quickly becomes impractical when the number of streams or number of constellation points is large. The ML solution provides diversity order of N and array gain of $\frac{N}{M}$.

3.2 Sphere Decoding (SD)

SD is an iterative method that converges to the ML when the number of iterations is not bounded. In SD, the multidimensional search implied by the ML criterion is transformed to multiple searches in one complex dimension.

The building block of the optimal LLR is the search for the minimizer of the cost functional

$$\min_{\mathbf{s}\in\Gamma}\|\mathbf{y}-\mathbf{Hs}\|^2 \quad (4)$$

over some set of points Γ. Continuing for simplicity on the 2×2 case, denoting the ZF solution as $\hat{\mathbf{s}}$, the cost functional in Eq. (4) may be rewritten as:

$$\|\mathbf{H}(\hat{\mathbf{s}}-\mathbf{s})\|^2=(\hat{\mathbf{s}}-\mathbf{s})^*\mathbf{H}^*\mathbf{H}(\hat{\mathbf{s}}-\mathbf{s})$$

Applying QR decomposition on $\mathbf{H}$ we end up with:

$$\|\mathbf{H}(\hat{\mathbf{s}}-\mathbf{s})\|^2=(\hat{\mathbf{s}}-\mathbf{s})^*\boldsymbol{R}^*\boldsymbol{R}(\hat{\mathbf{s}}-\mathbf{s})=\sum_{i=1}^{m}r_{ii}^2\left(s_i-\hat{s}_i+\sum_{j=i+1}^{m}\frac{r_{ij}}{r_{ii}}\left(s_j-\hat{s}_j\right)\right)^2=$$

$$r_{\mathrm{mm}}^2\left(s_{\mathrm{m}}-\hat{s}_{\mathrm{m}}\right)^2+r_{\mathrm{m-1,m-1}}^2\left(s_{\mathrm{m-1}}-\hat{s}_{\mathrm{m-1}}+\frac{r_{\mathrm{m-1,m}}}{r_{\mathrm{m-1,m-1}}}\left(s_{\mathrm{m}}-\hat{s}_{\mathrm{m}}\right)\right)^2+\ldots \quad (5)$$

We begin with searching for points $\mathbf{s}$ for which the cost functional Eq. (4) is smaller than an arbitrary r^2. Taking only the first term in the sum Eq. (4) we obtain a necessary (but not sufficient) condition for a point $\mathbf{s}$ to lie inside the sphere as $r_{\mathrm{mm}}^2\left(s_{\mathrm{m}}-\hat{s}_{\mathrm{m}}\right)^2<d^2$. This condition is equivalent to s_m being bounded between:

$$\left\lceil \hat{s}_m-\frac{d}{d_{mm}}\right\rceil\le s_m\le\left\lfloor \hat{s}_m+\frac{d}{d_{mm}}\right\rfloor \quad (6)$$

For every s_m satisfying Eq. (6), we shall define $d_{m-1}^2=d^2-r_{\mathrm{mm}}^2\left(s_{\mathrm{m}}-\hat{s}_{\mathrm{m}}\right)^2$. A stronger necessary condition can be found by looking at the first two terms in (5), which leads $s_{\mathrm{m-1}}$ to be bounded between

$$\left\lceil \hat{s}_{\mathrm{m-1,m}}-\frac{d_{m-1}}{d_{m-1,m-1}}\right\rceil\le s_{\mathrm{m-1}}\le\left\lfloor \hat{s}_{\mathrm{m-1,m}}+\frac{d_{m-1}}{d_{m-1,m-1}}\right\rfloor.$$

Continuing in the similar manner for s_{m-2} and so on we can obtain all the points inside the sphere.

Fig.2 illustrates the SD for the case of 16-QAM with two antennas. In this case, for each point of S_1 in the 1st circle, we search for points in the circle inside S_2.

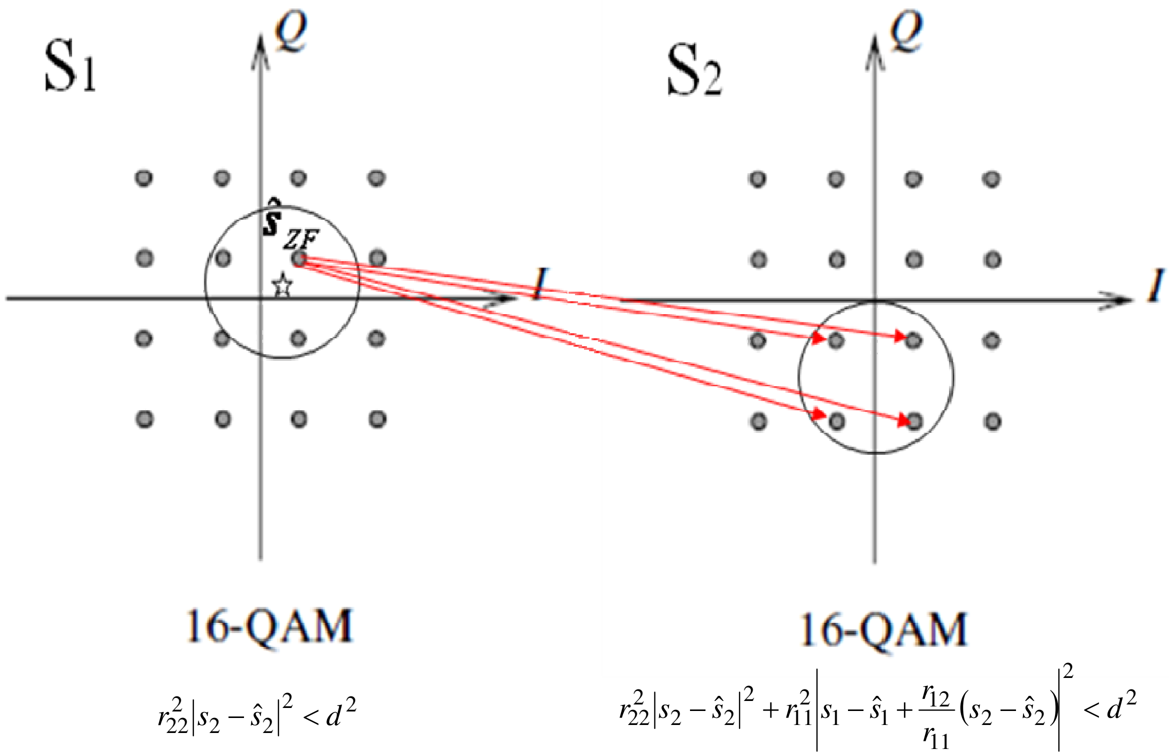

Fig.2- Illustration of the SD for the case of 16 QAM with two antennas

One of the major problems with the SD algorithm is that the number of iterations is not constant and may significantly vary between matrices. This makes hardware implementation of SD very difficult. When the number of iterations of SD algorithm is not limited, the array gain and diversity order are the same as for the ML.

4. The Proposed Algorithm Combining SD and ZF

The SD algorithm provides an exact ML solution, and exhibits the exponential worst-case complexity. On the other hand ZF is a suboptimal algorithm, especially when **H** is ill conditioned (large condition number), but has a polynomial complexity. The idea behind the proposed algorithm is to combine SD with ZF in a manner that exploits the advantages of both algorithms. Observations show that under Rayleigh fading environment, there is a small number of channel matrices, which have relatively large condition number. In such a case most of the matrices could undergo SD without significant noise amplification and therefore without much performance degradation. The channel matrices with large conditional number should undergo SD. Moreover, we can achieve fixed throughput operation of the algorithm by limiting the overall number on channel matrices intended for the SD. Such SD-ZF hybrid scheme can provide a fixed complexity near optimal MIMO decoding solution.

The proposed method assumes a constant hardware clock budget for the decoding of the K matrices, each with dimensions N×M. We further assume that the clock budget is larger than that needed for ZF decoding of all matrices.

The proposed decoding method is based on two main assumptions. Fist, we assume that matrices with high condition number should be likely to undergo SD. The second assumption is that the hardware clock budget must remains constant for the decoding of K matrices. Note that we do not attempt to construct an SD algorithm with finite number of iterations for each matrix, but restrict the number of overall iterations for the decoding of multiple matrices. Utilizing all of the above ideas, a complete algorithm sums up to the following steps:

1. Compute the linear ZF decoder for each of the K SM inputs (or matrices).

2.

$$\hat{\mathbf{s}} = \mathbf{H}^{+}\mathbf{y} \tag{2}$$

An SVD based approach is preferred here since it expedites the calculation of the condition number.

3. Order the K matrices according to the condition number, in descending order (largest first). This way the more problematic matrices in terms of decoding are assigned with a high priority.

4. Apply SD to the matrices according to the above-mentioned order until the hardware clock budget runs out. We note that the SD algorithm requires the ZF solution already obtained in the first step, resulting in no waste of clock budget in the first step.

5.

Thus, in the proposed algorithm, the matrices with high condition number are first to invoke the SD mechanism, which means efficient use of the hardware resources. The performance of the proposed algorithm is low bounded by that of ZF (in case the clock budget is identical to that required for ZF decoding), and it is high bounded by the performance of ML (in case the clock budget is sufficient for SD of all matrices).
The performance of the algorithm in actual scenarios is determined by the clock budget allocated and the distribution of the condition number of the channel matrices.

5 Simulation Results and Discussion

In this section, we evaluate the performance of the proposed algorithm by means of simulation. For simplicity and runtime limitation of simulation, we consider an uncoded 2x2 MIMO spatial multiplexing system with QPSK modulation. The system is tested for a Rayleigh fading channel. Details are shown in Appendix 1. For simplicity we assume perfect channel knowledge at the receiver. Simulation results for the proposed algorithm are shown in Fig.3.

We define the clock relation parameter n as

$$n = \frac{n_{TOT}}{n_{ZF}} \tag{3}$$

Where n_{TOT} is the number of overall hardware clocks, reserved for the decoding, and n_{ZF} is the number of hardware clocks, reserved for ZF decoding.
Fig.'s 3-5 show the BER curves corresponding to the performance of the proposed algorithm with different clock budgets. The BER curves for ZF and ML are added to for comparison purposes. Obviously the BER is smaller as the clock budget is increased. This figure clearly shows that in our method when the matrices are sorted by their conditional number, a small fraction of them undergoes SD, but still we can get significant enhancement in the performance. Moreover, for n=10, where on average 18% of the matrices is undergoing SD, the performance is almost identical to the optimal ML decoder.

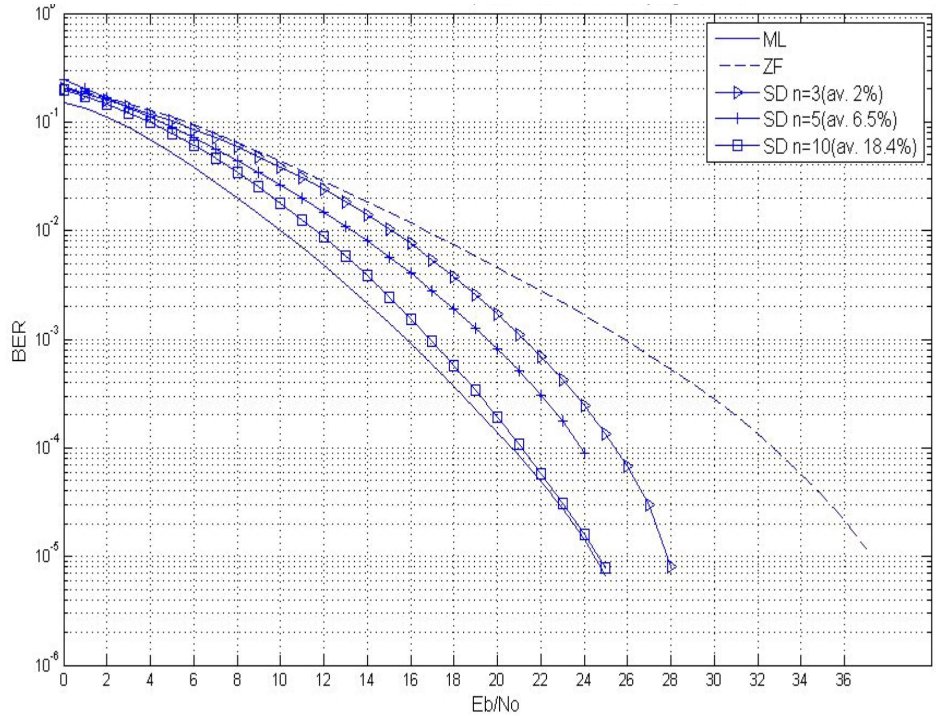

Fig. 3- Simulation results of ML, ZF and proposed SD algorithms in Raleigh channels

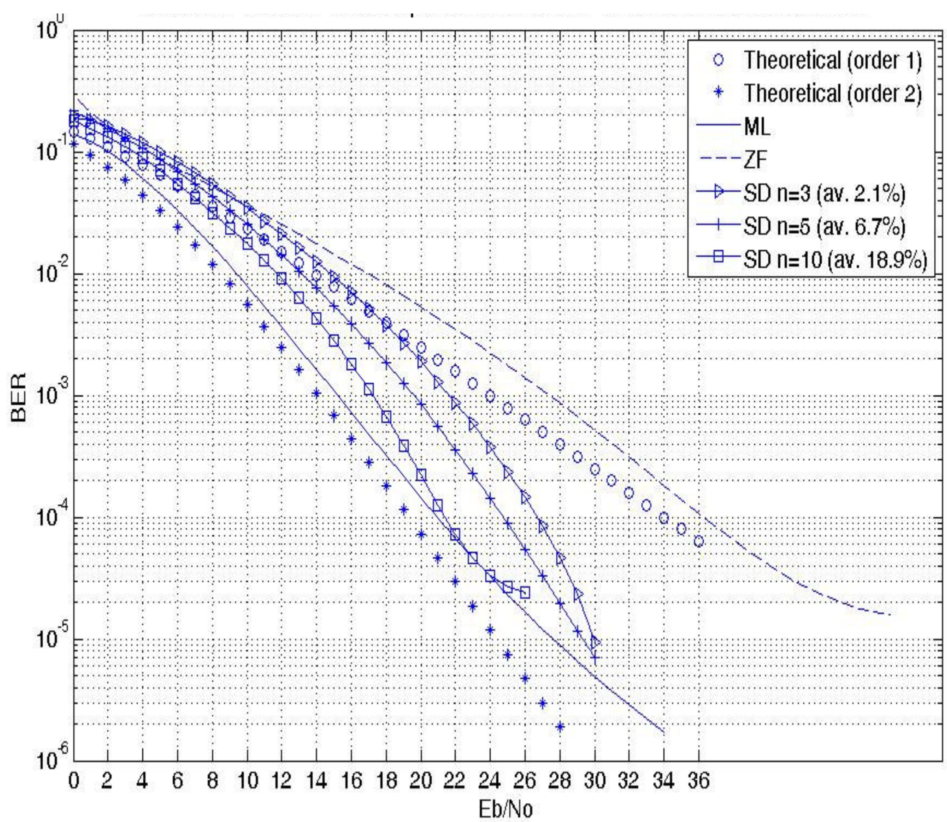

Fig. 4- Simulation results of ML, ZF and proposed SD algorithms in the WiMax Ped-B 3 kM/h channel

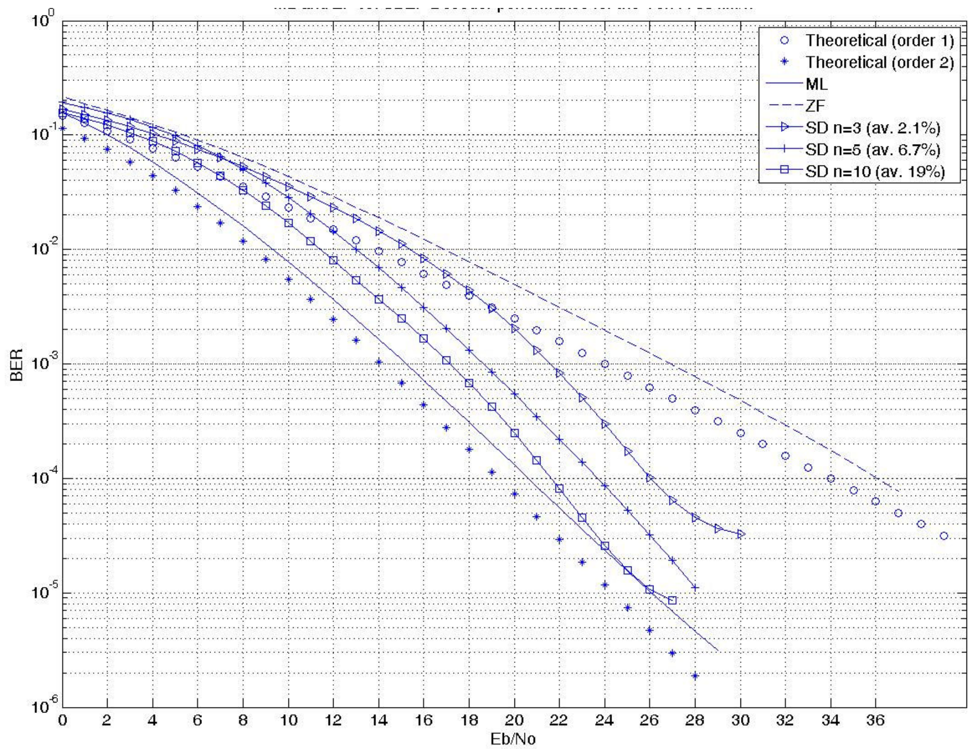

Fig.5- Simulation results of ML, ZF and proposed SD algorithms in the WiMax Veh-A 60 kM/h channel

6 Conclusion

We have presented a novel fixed complexity combined SD-ZF algorithm for decoding MIMO transmission, which is upper bounded by exact ML solution, depending of overall number of iterations, reserved for the decoding. Higher number of overall iterations causes the algorithm to be closer to the optimal ML solution. Our simulation results show that using the proposed algorithm only few percent of the receive matrices need to undergo SD to achieve near optimal performance; therefore the overall number of iterations could be relatively low, which provides low complexity hardware implementation. Moreover, the employment of the SD to a small portion of the matrices, will allow the accommodation of large antenna arrays featuring a large number of spatial streams.

References

[1] B. Hassibi and H. Vikalo, "On the expected complexity of integer least-squares problems" IEEE ICASSP, Orlando, FL, 2002

[2] U. Fincke, M. Pohst, “Improved methods for calculating vectors of short length in a lattice, including a complexity analysis”. Mathematics of Computation, vol. 44, 1985, pp. 463–471,

[3] Kwan-Wai Wong, Chi-Ying Tsui, R. S.-K. Cheng, Wai Ho Mow, “A VLSI architecture of a K-best lattice decoding algorithm for MIMO channels”, ISCAS (3) 2002: 273-276

[4] D. Garrett, L. Davis, S. ten Brink, B. Hochwald, and G. Knagge, “Silicon complexity for maximum likelihood MIMO detection using spherical decoding”, IEEE Journal of Solid-State Circuits, vol. 39, pp. 1544–1552,Sept. 2004.

[5] B. M. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel”, IEEE Trans. Communication., Vol. 51, No. 3, 2003, pp. 389-399.

[6] J. Maurer, G. Matz, and D. Seethaler, “Low complexity and full diversity MIMO detection based on condition number thresholding”, ICASSP 2007

[7] G.J. Foschini, “Layered space - time architecture for wireless communication in a fading environment when using multiple antennas’, Bell Lab. Tech. J., Vol. 1, (2), 1996, pp. 41–59

[8] M. Grotschel, L. Lov´asz, A. Schriver, "Geometric Algorithms and Combinatorial Optimization", Springer Verlag, 2nd ed., 1993

[9] T. Kailath, H. Vikalo, B. Hassibi, "MIMO Receive Algorithms" in Space-Time Wireless Systems: From Array Processing to MIMO Communications, (editors H. Bolcskei, D. Gesbert, C. Papadias, and A. J. van der Veen), Cambridge University Press, 2005

[10] A. Paulraj, R. Nabat, D. Gore, "Introduction to Space-Time Wireless Communications", Cambridge University Press, Cambridge, UK, 2003

Appendix 1

WiMax Ped-B channel definition:

Channel B		Doppler spectrum
Relative delay (ns)	Average power (dB)	
0	0	Classic
200	–0.9	Classic
800	–4.9	Classic
1 200	–8.0	Classic
2 300	–7.8	Classic
3 700	–23.9	Classic

WiMax Veh-A channel definition:

Tap	Channel A	
	Relative delay (ns)	Average power (dB)
1	0	0.0
2	310	–1.0
3	710	–9.0
4	1 090	–10.0
5	1 730	–15.0
6	2 510	–20.0