

## Low complexity MIMO detection algorithm

Igor Jelovčan, Tomaž Javornik

Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia  
E-mail: igor.jelovcan@ijs.si, tomaz.javornik@ijs.si

**Abstract.** The multi antenna systems offer a much larger channel capacity than the traditional single antenna systems. The price which has to be extra paid is detection complexity. In this paper we examine performance of the low complexity algorithm called zero forcing–maximum likelihood (ZFML) [5]. We find this algorithm appropriate for MIMO (Multiple Input Multiple Output) systems using more receive than transmit antennas. For the 2x4 VBLAST (Vertical Bell-Labs Layered Space-Time) MIMO system almost maximum likelihood (ML) performance is achieved. To overcome large performance gaps in NxN MIMO systems, some modifications of the algorithm are needed. In future, we intend to upgrade this algorithm to allow for iterative decoding in turbo MIMO systems.

**Keywords:** MIMO, ML, ZF, ZFML, sphere detection

## Algoritem za detekcijo MIMO z nizko kompleksnostjo

**Povzetek.** Večantenski sistemi ponujajo precej večjo zmogljivost kanala kot tradicionalni enoantenski sistemi. Cena, ki jo je treba plačati, je kompleksnost detekcije. V tem članku smo proučili zmogljivost ZFML algoritma z nizko kompleksnostjo [5]. Ugotovili smo, da je algoritem primeren za sisteme MIMO, ki uporabljajo več sprejemnih kot oddajnih anten. Za sistem 2x4 VBLAST MIMO je bila dosežena skoraj ML zmogljivost. Rezultati za sisteme NxN MIMO so bistveno slabši, zato je potrebna izboljšava algoritma. V prihodnosti nameravamo algoritem nadgraditi tako, da bo primeren za iterativno dekodiranje v sistemih turbo MIMO.

**Ključne besede:** MIMO, ML, ZF, ZFML, sferična detekcija

---

### 1 Introduction

The increasing demands for higher data rates and higher quality of service in wireless communication systems have forced the designers of the emerging communication systems to invent new approaches toward improvements of the spectral efficiency and reliability. Among them, the use of multiple transmit and receive antennas, known as the MIMO approach, is one of the most promising. In the rich scattering environment the MIMO approach can provide either higher spectral efficiency or higher diversity gain. One

of the major obstacles to apply the MIMO approach is the complexity of the optimal ML detection. The available diversity can be fully exploited applying the ML detection that uses exhaustive search over all the possible input vectors to find the transmitted signal, yet, at the price of exponential complexity. For that reason, the ML detection is not suitable when the number of antennas is large and for communication systems with limited power resources at the base or mobile station, like terrestrial-mobile or land-mobile satellite systems and wireless communications via high-altitude platforms.

Apart from the optimal ML detection, there are several different algorithms for the MIMO detection, all of them requiring the channel state information (CSI) to be known at the receiver. The basic linear methods are zero forcing (ZF) and minimum mean-square error (MMSE), whose performance is significantly inferior to the ML detection. These two algorithms are the basis for some other non-linear algorithms, e.g. ordered successive interference cancellation (OSIC) [1]. The large gap in performance and complexity between the ML and those conventional detectors has motivated many researchers to envisage development of suboptimum ML detection schemes. Lots of them are based on the sphere detection (SD) [2], which gives a near-ML solution. Its complexity varies depending on the estimated radius.

The paper examines the performance of the suboptimum algorithm based on sphere detection that efficiently combines linear processing with local ML search [5]. The algorithm is a modified complex-sphere detector, appropriate for the use in the VBLAST MIMO systems with a higher-level quadrature amplitude

modulation (QAM) on each transmit antenna. We first give a system model description and a short survey of the basic MIMO detection algorithms and then we briefly describe the ZFML algorithm and the simulation setup. The next to follow are simulation results. In the conclusion of the paper we propose the work to be done in future.

## 2 System description

Our focus is on the MIMO systems with  $M$  transmit and  $N$  receive antennas ( $M < N$ ), where the binary data stream is demultiplexed into  $M$  data substreams. Each substream is mapped onto m-QAM symbol  $x_i$  and transmitted over  $M$  antennas simultaneously. To place more emphasis on the detection algorithms, we consider a simple but most commonly used independent block flat Rayleigh fading channel. In such a baseband system the received  $N \times 1$  complex signal vector  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$  is given by

$$\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{x}$  denotes the complex value  $M \times 1$  transmit signal vector,  $\mathbf{n}$  is a  $1 \times N$  vector of independent zero-mean complex Gaussian noise entries with variance  $\sigma_n^2$  and  $\mathbf{H}$  is a  $N \times M$  complex channel matrix, perfectly known at the receiver. The entries of  $\mathbf{H}$  are independent, zero-mean and complex Gaussian variables of unit variance ( $\sigma_c^2 = 1$ ). The vector  $\mathbf{x}$  is assumed to have zero-mean independent complex values with variance  $\sigma_x^2$ . The total transmitted power is  $P$ , which is equally allocated along all  $M$  transmit antennas, thus the expected signal to noise ratio (SNR) per receive antenna  $\rho$  equals  $P/\sigma_n^2$ .

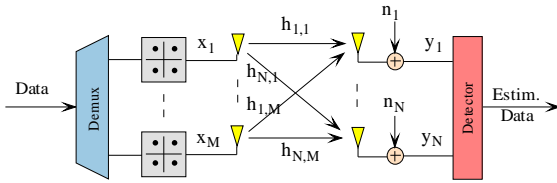


Figure 1: MIMO system

The standard way of detecting the MIMO signal is ZF, where the interference is completely suppressed by multiplying the received signal vector  $\mathbf{y}$  with the pseudo-inverse of the channel matrix  $\mathbf{H}^+ = (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^*$ , where  $*$  denotes the hermitian operator. Each calculated value has to be mapped onto the nearest symbol from the modulation symbol alphabet.

$$\mathbf{x}_{ZF} = \mathbf{H}^+ \mathbf{y} = \mathbf{x} + (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* \mathbf{n} \quad (2)$$

Another basic approach for detecting the MIMO signal is called MMSE, where the noise term is considered. Instead of pseudo-inverse matrix  $\mathbf{H}^+$ , matrix  $\mathbf{G}$  is used:

$$\mathbf{G} = (\mathbf{H}^* \mathbf{H} + \frac{M}{\rho} \mathbf{I}_M)^{-1} \mathbf{H}^*, \quad (3)$$

which minimizes the mean-square error:

$$\begin{aligned} \varepsilon^2 &= E[(\mathbf{x} - \mathbf{x}_{MMSE})^* (\mathbf{x} - \mathbf{x}_{MMSE})] = \\ &= [(\mathbf{x} - \mathbf{G} \mathbf{y})^* (\mathbf{x} - \mathbf{G} \mathbf{y})] \end{aligned} \quad (4)$$

Both ZF and MMSE are linear methods where all layers are detected jointly and can achieve the diversity order of only  $N-M+1$ . Superior performance is obtained applying non-linear detection methods. The first one to be explained is successive interference cancellation (SIC), also called the nulling and cancelling algorithm. Each layer is detected separately with ZF or MMSE detection. When the first layer is estimated, its corresponding contribution to the received vector  $\mathbf{y}$  is cancelled. Then the corresponding column from the matrix  $\mathbf{H}$  is removed and a new pseudo-inverse is calculated. This procedure is iteratively repeated until the system is reduced to one dimension. At each step we assume that the previously made decisions are correct. However, the SIC can suffer from error propagation, which can be minimized by detecting layers starting from the strongest and proceeding to the weakest (OSIC).

When the likelihood of all the data vectors is equal, the optimal solution is the ML detection. It maximizes the likelihood that  $\mathbf{y}$  is received given that  $\mathbf{x}$  is sent. The ML detector is given by:

$$\mathbf{x}_{ML} = \arg \min_{\mathbf{x}_j \in \{\mathbf{x}_1, \dots, \mathbf{x}_K\}} \|\mathbf{y} - \mathbf{H} \mathbf{x}_j\|^2. \quad (5)$$

This requires an exhaustive search through all possible vectors  $K = m^M$ . Thus the complexity grows exponentially with the number of transmit antennas. The ML search is limited in the sphere detection algorithm, which searches only over those points that lie inside an  $M$ -dimensional hypersphere with given radius  $r$  and centred at the received point  $\mathbf{y}$ . Two key questions have to be answered. How to choose  $r$  and how to determine which lattice points are inside the hypersphere? The lattice point  $\mathbf{H} \mathbf{x}$  lies in a hypersphere of radius  $r$  if

$$r \geq \|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2. \quad (6)$$

Selecting lattice points within the  $M$  dimensional sphere involves calculation of Eq. (6) for all hypothetical lattice points, which is a complex task when  $M$  is high. The SD simplifies the lattice points selection, by solving the problem in the hyperspheres at first in one dimension assuming the radius  $r$ , and then it recursively proceeds by increasing the dimension by one. Thus, it constructs a searching tree, where the nodes in the  $k$ -th level of the tree correspond to the lattice points inside the  $k$ -th dimension hypersphere of radius  $r$ . The path, also called the vector point, which survives to the level  $M$  and has a smaller metric than  $\|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2$ , is the solution to the closest point search.

The radius  $r$  should be chosen very carefully considering the current noise power. If radius is too small, no lattice point is found inside a sphere, if it is too big, decoding is still a demanding task. Using the Finke-Phost sphere decoding algorithm, ML detection can be achieved at an average complexity  $O(M^3)$  [3].

SD was originally derived for a real lattice, thus the Eq. (1) has to be transformed into a real value equivalent. The sizes of matrices  $\mathbf{y}$ ,  $\mathbf{H}$ ,  $\mathbf{x}$  and  $\mathbf{n}$  are doubled and comprise the real and imaginary parts of previous matrices. This decoupling is not possible for an arbitrary system. It is possible, however, if QAM modulation is used. For systems where decoupling is unfeasible, complex SD has to be used [4].

There are many other algorithms available that improve performance and/or reduce complexity. In the following section we will concentrate on an algorithm which allows, by applying additional computations, the calculation of soft information required for iterative decoding in turbo MIMO systems. This algorithm avoids exhaustive search for maximum a posteriori (MAP) detection. It was originally presented in [5] and is known as the ZFML algorithm. Soft decoding is possible also with some other algorithms like the list-sphere detector (LSD) [4], space-time Chase decoder [6]. The latter makes no assumption of the transmit constellation, therefore it does not exploit all the available information.

### 3 ZFML algorithm

The ZFML algorithm reduces exhaustive ML searching and is appropriate for higher-order modulations with 16 or more levels. It is composed of several steps. The first step is a conventional detection of the received signal for evaluating initial estimating symbol vector  $\mathbf{x}_{est}$ . Then a subset of neighbouring symbols is generated for each antenna. At the end an exhaustive local ML detection is performed over this relatively small searching region.

For conventional detection any detector (ZF, MMSE, OSIC, etc.) can be chosen. The better it is the better performance is obtained, at the price of higher complexity. On the other hand, the result of better estimating vector  $\mathbf{x}_{est}$  is much closer to the  $\mathbf{x}_{ML}$  solution, thus the searching region can be decreased, which further reduces detection complexity.

The searching region is the result of all possible combinations over the neighbouring symbols of all antennas. If the number of neighbouring symbols is fixed and the same for all antennas, then the easiest way of generating the subset is via a lookup table. An example for 16QAM modulation is shown in Fig. 2, where nine different regions are possible, each containing four constellation points. The neighbouring symbols on the  $i$ -th antenna are constellation points of that region, within which estimated solution  $x_{est}^{(i)}$  is

located. With a higher order modulation the number of regions grows quickly, thus the lookup table becomes impractical. For example, if 64QAM modulation is used, then there are 49 regions with four constellation points. The QAM constellation diagram being symmetric, the neighbouring symbols can be determined dynamically with algebraic equations. In that case the number of the neighbouring symbols can be variable and different for each antenna, depending on current channel conditions.

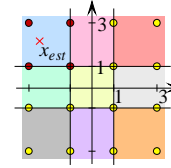


Figure 2: Example of the neighbouring symbols (black) for 16QAM

Once the searching region is determined, the ML detection is performed over it. In the case of a fixed-size neighbouring list containing  $s$  constellation points, there are  $s^M$  possible symbol vectors which have to be examined. Compared to ordinary ML searching over the entire constellation, the reduction rate in the number of searches is  $(m/s)^M$ .

The same algorithm is extended in [7], where the searching region is defined so that the ML solution is always included. The ML solution is located within a sphere of radius  $d$  and centred at the received signal  $\mathbf{y}$ :

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}_{ML}\|^2 \leq \|\mathbf{y} - \mathbf{H}\mathbf{x}_{est}\|^2 = d^2. \quad (7)$$

When the QAM modulation is used, point searching within a certain region is much easier, if the structure of the searching region is similar to the one of the QAM constellation diagram. Thus the polygon searching region is proposed in [7], which always contains a sphere with radius  $d$ . The neighbouring symbols for the  $i$ -th antenna are simply determined by solving the below expressions:

$$\begin{aligned} -d &\leq \operatorname{Re} \left\{ \frac{x_{est}^{(i)} - x_{neighboring}^{(i)}}{\|g^{(i)}\|} \right\} \leq d, \\ -d &\leq \operatorname{Im} \left\{ \frac{x_{est}^{(i)} - x_{neighboring}^{(i)}}{\|g^{(i)}\|} \right\} \leq d, \end{aligned} \quad (8)$$

where  $\|g^{(i)}\|$  is the norm of the  $i$ -th column of the inverse matrix  $\mathbf{G}$ . At low SNR or at a badly conditioned channel ( $g^{(i)}$  is high), the number of neighbouring symbols increases and the overall detection becomes complex. In the case of the time varying wireless channel they may lead to variable processing time at the receiver causing an additional and also variable transmission delay, which is unacceptable for some communication services. In order to solve this problem we propose a

search region which is independent of the channel matrix and is clearly described in the previous section and depicted in Fig. 2.

### 4 Simulation setup

This paper is the result of our ongoing research in turbo MIMO aiming at finding a simple but still efficient way for MIMO detection and also enabling a later extension in the soft MIMO detection algorithm. From a group of algorithms we opted for a multi-step algorithm with ZF pre-estimation (ZFML). This algorithm has so far been very poorly examined, although results in [5] are quite promising.

We analyzed the properties of this algorithm by using Monte Carlo simulation. For each simulated point in the BER curve at least 400.000 bits were processed or 200 errors detected. Simulations were run for various antenna combinations with 16QAM modulation.

The size of the neighbouring region varied depending on the simulation type but was the same for all antennas. Fig. 3 clearly indicates how the neighbouring constellation points are determined. Once the sub-region into which  $x_{est}$  falls is determined, the chosen constellation points are ordered from the closest ones to the most distant ones. According to the desired extent of searching for the transmitted symbol, two, three or four constellation points for each antenna are chosen. Whenever only one point is selected, we talk about the ZF solution. Unless otherwise defined, the neighbouring region in our simulations comprises of four neighbouring constellation points.

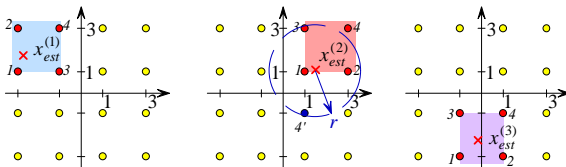


Figure 3: Neighbouring region for MIMO with three transmit antennas and 16QAM modulation

One of the main disadvantages of the squared searching region is shown in Fig. 3. The circle on the second antenna represents the region within which all constellation points are distant from  $x_{est}^{(2)}$  for  $r$  or less. It is clear that the more distant point 4 is selected instead of the closer point 4'. This happens when radius  $r$  to the most distant point 4 is approximately 2.3 or higher.

### 5 Simulation results

Fig. 4 shows the bit error rate (BER) curves of the ZF and ML detectors for various MIMO systems, where  $E_b$  denotes the average energy per information bit arriving at the receiver ( $E_b/N_0 = \rho \cdot N/M \log_2 m$ ). When only

one transmit antenna is used, the ZF detection gives the same results as ML detection. By increasing the number of transmit antennas, a large gap arises between the ML and ZF detection. The performance gap between those two detections will decrease if the number of receive antennas is higher than the number of transmit antennas. The diversity order exploited by ZF is only  $N-M+1$ , while at the ML detection it is equal to  $N$ .

Performance results of the ZFML algorithm compared to the ZF and ML detection are shown in Fig. 5. We simulated various antenna configurations with 16QAM modulation. As the simulations results show similar system performance only two cases (2x4, 4x4) are presented in our paper. The performance gains compared to ZF detection are at BER =  $10^{-4}$  for the 2x4 system 2.6 dB and 3.9 dB for the 4x4 system. Much more remarkable is the fact that the ML performance can almost be achieved by greatly reducing detection complexity of the MIMO systems, where the number of receive antennas is higher than the number of transmit antennas. For the 2x4 MIMO system the ordinary ML detection has to examine all the 256 possible symbols while ZFML searches for the best symbol within a subset of only 16 symbols. Some simulations were made also for the 64QAM modulation, where the results were very similar to those made with 16QAM and only the complexity reduction was much bigger (from 4096 to 16 symbols).

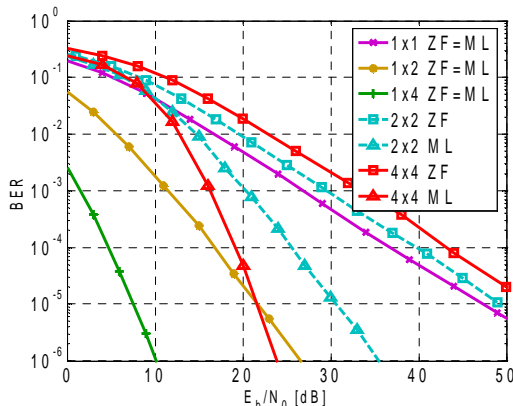


Figure 4: Performance comparison of the ZF and ML detection for various V-BLAST MIMO systems using 16QAM modulation

All currently made simulations were made with four neighbouring constellation points per antenna. Fig. 6 shows the achieved gain when using only two or three neighbouring points. Ratio between the added complexity and detection gain is the largest at the smallest neighbouring subset. With the ML detection over four MIMO symbols a 2.3 dB gain is obtained at BER =  $10^{-5}$ . An additional 1.4 dB gain is obtained when detection is made within nine MIMO symbols. When

$S = 16$ , the performance curve is 4.4 dB better than the ZF curve but still about 2 dB from the ML solution

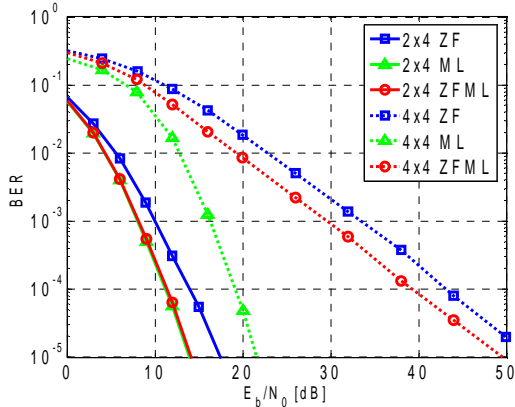


Figure 5: Performance comparison of different detection algorithms for the 2x4 and 4x4 V-BLAST MIMO systems using 16QAM modulation

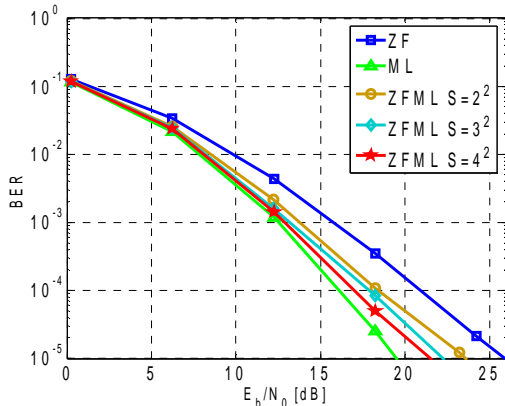


Figure 6: Performance of the ZFML algorithm with a different neighbouring list size for the 2x3 V-BLAST MIMO systems using 16QAM modulation

## 6 Conclusion

We present and analyze the ZFML MIMO detection algorithm that extremely reduces computational complexity when compared to the ML detection. The higher is the number of the transmit antennas and modulation level the higher is the complexity reduction. This complexity reduction indirectly results in power consumption reduction.

Simulation results also show that the algorithm can achieve almost the ML performance if there are more receive than transmit antennas. Otherwise, a large performance gap is obtained that can be suppressed by adaptive determination of neighbouring constellation points. The searching region on each antenna should be of a circular shape of a variable radius, depending on

the current channel conditions. Some preliminary simulations show that radius suggested in [7] is defined too loosely.

In future we intend to modify the algorithm, in order to make it more adaptive and to be used in the iterative decoding of the encoded MIMO signals.

## 7 References

- [1] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky; Detection algorithm and initial laboratory results using VBLAST space-time communication architecture, *Electronic Letters*, vol. 35, no. 1, pp. 14–15, Jan., 1999.
- [2] U. Fincke, M. Pohst; Improved methods for calculating vectors of short length in a lattice, including a complexity analysis, *Math. Comput.*, vol. 44, no. 4, pp. 463–471, Apr. 1985.
- [3] B. Hassibi, H. Vikalo; On the Sphere Decoding Algorithm. I. Expected Complexity, *IEEE Trans. on Sig. Proc.*, vol. 53, no. 8, pp. 2806-18, Aug 2005.
- [4] B. M. Hochwald and S. ten Brink; Achieving near-capacity on a multiple-antenna channel, *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389-399, Mar. 2003.
- [5] X. Li, H. C. Huang, A. Lozano, G. J. Foschini; Reduced-complexity detection algorithms for systems using multi-element arrays, *Proc. IEEE GLOBECOM 2000*, pp. 1072-1076, Nov. 2000.
- [6] D. Love, S. Hosur, A. Batra, R. W. Heath.; Space-Time Chase Decoding, *IEEE Trans. on Wireless Comm.*, vol. 4, pp. 2035-2039, Sept 2005.
- [7] L. He, H. Ge; Reduced complexity maximum-likelihood detection for V-BLAST systems, *Proc. IEEE MILCOM 2003*, pp. 1386-1391, Oct. 2003.

**Igor Jelovčan** graduated in 2004 from the University of Ljubljana, Faculty of Electrical Engineering by defending his thesis in the field of turbo codes. In the same year he joined the Jožef Stefan Institute, the Department of Communication Systems to work as a Young Researcher of the National Research scheme. After finishing his M.Sc. in 2007 he is currently employed at Telsima d.o.o. His research work includes various iterative algorithms in a special emphasis on MIMO systems and research and development of WiMAX systems.

**Tomaz Javornik** received his B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from the University of Ljubljana, Slovenia, in 1987, 1990 and 1993, respectively. In 1987 he joined the Institute Jozef Stefan, where he currently works as a researcher in the Department of Communication Systems. He is involved in the study of digital radio-rely systems, modulation techniques, coding, adaptive signal processing and digital mobile communication systems.