Robust ML and Sphere Decoder

Developed by

Omar Hayat

Reg No 184-FET/MSEE/F08

Supervisor

Dr. Ihsan-ul-Haq

Department of Electronic Engineering

Faculty of Engineering and Technology, (FET)

International Islamic University, (IIU), Islamabad

2011

Robust ML and Sphere Decoder

Omar Hayat

(184-FET/MSEE/F08)

This dissertation is submitted to Faculty of Engineering and Technology, International Islamic University Islamabad Pakistan for partial fulfillment of the degree of **MS Electronic Engineering** With specialization in Communication at the Department of Electronic Engineering

Faculty of Engineering and Technology, (FET)

International Islamic University, (IIU), Islamabad.

Supervisor:

Dr. Ihsan ul Haq

23 Feburary 2011

1. Multiuser detection

Certificate of Approval

Title of Thesis: Robust ML and Sphere Decoder

Name of Student: Omar Hayat

Registration No: 184-FET/MSEE/F08

Accepted by the Faculty of Engineering and Technology, International Islamic University, Islamabad, in partial fulfillment of the requirements for the Master of Science Degree in Electronic Engineering.

Viva voice committee

Chairman/Director/Head

External Examiner

Internal Examiner

Supervisor
Dr. Ihsan Ul Haq
Associate Professor,
FET, IIU, Islamabad

07/04/2011

23 Feburary 2011

Declaration

I certify that except where due acknowledgments has been made, the work has not been submitted previously, in whole, to qualify for any other academic award, the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program, and any editorial work paid or unpaid, carried out by a third party is acknowledged.

Omar Hayat

184–FET/MSEE/F08

Paper acceptance

Research paper has been published in IEEE ICET 2010 held in FAST Islamabad. The topic was "Sphere decoder in the presence of channel uncertainty".

Abstract

The performance of maximum likelihood (ML) and Sphere decoding (SD) are investigated when the channel is uncertain. In this work, a modified low complexity approach is proposed to deal with the effects of the channel uncertainty and obtain robust ML and SD. A higher complexity approach is also suggested but is not investigated here as it appears to be unsuitable for real time applications. The approach adopted involves linearly preprocessing the received signal to suppress the interference caused by channel uncertainty before applying a detector. Parallels of this approach may be found in the channel shortening literature where it is suggested to linearly preprocess the received signal with a channel shortened to reduce the effective delay spread of the inter symbol interference(ISI) channel before applying a reduced state Viterbi decoder to bring down the complexity of ML decoding. In our case, however, the pre processing is applied to reduce the effect of the interference caused by channel uncertainty.

Acknowledgements

There are many dedicated people who just manage to stay behind the scenes to complete my thesis. It is my pleasure to acknowledge their contribution.

First of all, I am thankful to Almighty ALLAH for showering his blessing upon me and make be able to complete this thesis.

It is my great pleasure that I completed my thesis under the supervision of Dr Ihsanulhaq. He really inspired me to step in the broad field of wireless communication. I want to express my deep regards to my supervisor for his guidance and helpfulness, without which I would not have reached so far. His fruitful discussions and the way to approach the particular problem would prove great asset in my future life. I would never forget his informal and cooperative behavior during my thesis. I would take this opportunity to thank all my colleagues at the faculty of engineering, NUML, especially Dr Tariq Nawaz Qaisrani for providing me such healthy and competitive environment throughout the course. I also thankful to Mr. Muhammad Irfan Rafique and Mr. Imran Ali who, gave me the moral supports for each steps of my study and social life.

The one person whom I never forget in my life is my elder brother Hafiz Muhammad Ishaq. His moral, financial support helps me to complete my master degree. I don't have any word to write for his favors. I have great feelings of gratitude for my mother, father and Wife whose endless love provided me peace of mind to concentrate on my study. Their prayers work as a torch bearer to achieve every mild stone in my life.

Omar Hayat

Table of Contents

List of Figures

List of Acronyms

BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BLAST	Bell Labs Layered Space Time
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
D-BLAST	Differential Bell Labs Layered Space Time
FSD	Fixed Complexity Sphere Decoder
ISI	Inter-symbol Interference
ID&CE	Iterative Detection and Channel Estimation
MIMO	Multiple Input Multiple output
MMSE	Minimum mean Square Error
MRRC	Maximal Ratio Receiver Combiner
MISO	Multiple Input Single Output
MLE	Maximum Likelihood Estimate
MU-MIMO	Multiuser MIMO
MMSE-SIC	Minimum Mean Square Error Successive Interference cancellation
MAP	Maximum A-posteriori Probability
ML	Maximum Likelihood
NLOS	Non Line Of Sight
OSIC	Ordered Successive Interference Cancellation
RF	Radio Frequency
SISO	Single Input Single Output
SNR	Signal to Noise Ratio

SIMO	Single Input Multi Output
SIC	Successive Interference Cancellation
SD	Sphere Decoder
V-BLAST	Vertical Bell Labs Layered Space Time
ZF	Zero Forcing
ZF-SIC	Zero Forcing Successive Interference Cancellation

Chapter 1

Introduction to MIMO

1.1 Introduction

The ever increasing demands for higher data rates caused by the multimedia and interactive applications have created a problem for communications engineers. The data rates that can actually be achieved depend on both physical world factors like fading, NLOS (Non Line Of Sight), and path loss and also on regulatory bodies imposing constraints on the available bandwidth, allowed transmit power, etc. From [1], we know that the data rate increase requires more bandwidth and a better signal to noise ratio – hence more transmit power. We are currently operating at the peak with regards to these parameters. Contemporary communications systems cannot increase the transmit power as they are constrained by the regulatory authorities because of safety considerations and to limit interference to adjacent bands. Neither can they afford to introduce more bandwidth as RF spectrum is a precious resource which is scarce and hence expensive. The key to meeting the demands for higher data rates lies in using MIMO (Multiple Input and Multiple Output) technology [2-4].

The gains to be realized through MIMO are indeed impressive with the possibility of spatial multiplexing gain, beam forming, and diversity and array gains [4]. With the introduction of multiple antennas at the transmitter and at the receiver, many signal processing possibilities open up posing several interesting problems. These gains are also verified by information theoretic analysis. With the introduction of MIMO, signal processing becomes far more complex. Furthermore, the requirements are now far greater not just for the signal processing aspects but also from the RF point of view as we would now require multiple RF chains which are expensive. From the signal processing perspective which is our interest, we now require substantially complicated signal processing. Channels are usually time varying and where we once had to train the receiver for only one transmitter in single input single output, we now need to train the receivers for multiple transmitters which means a higher signal processing overhead from a training perspective [4]. Furthermore, the signal processing at the receiver generally assumes knowledge of the channel. This is of course unknown and must be estimated through training. Even with training, we will get an estimate of the channel which may be somewhat different from the actual channel.

Signal processing for MIMO generally involves designing a suitable receiver which may be based on the (ZF) zero forcing, MMSE (Minimum Mean Square Error), (SIC) successive interference cancelation or the (ML) maximum likelihood criterion [5, 6]. Several complexity reducing algorithms have also been proposed, in particular the sphere decoder [7-11]. All these techniques require channel knowledge at the receiver which may not be perfectly available. In this dissertation, we investigate robust techniques for MIMO which by design are not sensitive to variations and or uncertainties in the channel. We will seek robust variations on maximum likelihood and sphere decoding.

1.2 MIMO System Modal

A MIMO wireless communication system has a multiple transmit and receive antennas. We will see that this allows the transmitter and receiver to open up multiple spatial pipes (or streams) for information transfer between each other thereby offering excellent performance gains and increasing the MIMO channel capacity.

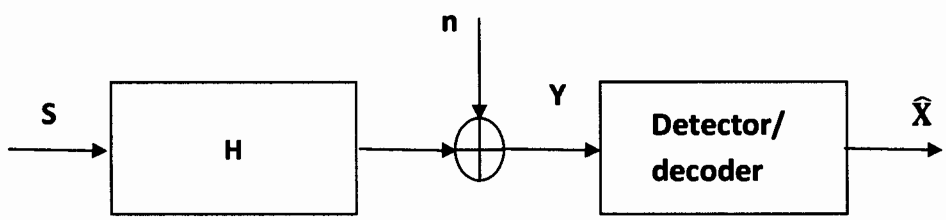

Figure 1: MIMO system model

Assume a MIMO channel with M_T transmit antennas and M_R receive antennas. This flat fading MIMO channel is modeled by the $M_R \times M_T$ matrix **H**. The i, j^{th} entry of **H** i.e. $[h_{ij}]$ is the channel gain from the j^{th} transmit antenna to the i^{th} receive antenna as shown in Figure 2.

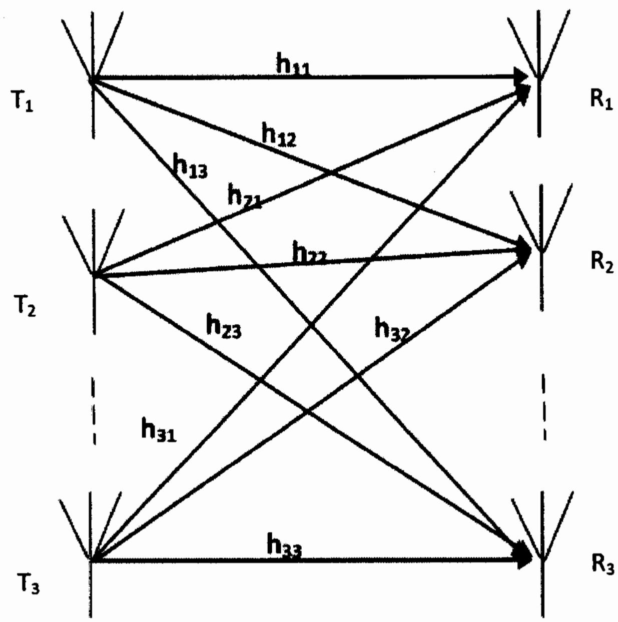

Figure 2: MIMO channel model

The channel matrix for a MIMO system is

$$H=\begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_R,1} & h_{M_R,2} & \cdots & h_{M_R,M_T} \end{bmatrix} \quad 1.1$$

The channel matrix includes the effects of transmit, receive filtering and the effects of the RF chains. We can write the output of system in the form of equation

$$y_i = \sum_{j=1}^{M_T} h_{ij} * s_j + n_i, \text{where } i = 1,2 \ldots M_R \quad 1.2$$

s_j is the transmitted symbols vector $s_j = [s_1, s_2, \ldots\ldots s_j]$

We can also write this equation in term of the channel matrix as follows

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \qquad 1.3$$

Where **y** is the output vector, **H** is the channel matrix and **s** is input vector and **n** is the noise that added at the receiver front end.

1.3 Performance Improvement of MIMO System

MIMO systems can potentially provide the following gains over a single input single output (SISO) system

i) Diversity Gain

ii) Beam Forming

iii) Spatial multiplexing Gain

iv) Array Gain

1.3.1 Diversity Gain

In a wireless channel signal power fluctuates randomly, so the multiple antennas at the receiver and transmitter are used to get the diversity gain. Using the diversity, fading effects can be mitigated in the wireless channel. Basically diversity is to send the same data on independent fluctuating signal paths. There are many methods to achieve the diversity. One is by using antenna arrays called spatial diversity [12]. Another method is to send the same data on the different carriers or transmit the same signal at different time with the condition that the spacing between the successive transmissions must exceed the coherence time of channel ($T > \tau_c$). These techniques are called frequency and temporal diversity. At the receiver, signals receive from different independent paths are combined. There are several ways to combine the received signal [12] like linear combiner, selection combining, threshold combing and Maximal Ratio Combiner (MRC). A famous scheme to achieve transmit diversity is the 'Alamouti Scheme' proposed by Siavash Alamouti [12].

1.3.2 Beam forming

Beam forming allows the antenna array to transmit or receive in a particular direction and to place nulls in desired directions. This is key to using multiple antennas for interference rejection and interference management in multiuser communications. Beam forming requires channel knowledge. As an example, for coherent combining of the received signals, channel knowledge has to be assumed at the receiver. Beam forming requires explicit knowledge of the channel at the transmitter while transmit diversity does not. Beam forming increases the rang, reduces the interference and increase the overall network capacity [12].

1.3.3 Array Gain

Array gain is power gain that improves the performance of a system. Array gain focuses the transmitted energy towards the receiver, so that more power is received. Transmitter or receiver array gain needs channel state information (CSI) at the transmitter and receiver. It also depends on the number of receive and transmit antennas. Channel state information (CSI) at the transmitter is difficult to maintain [3], but channel state information (CSI) is typically available at receiver.

1.3.4 Spatial multiplexing gain

Spatial multiplexing gain can be achieved by transmitting independent data from individual antennas [4]. Channel capacity increases linearly with respect to size of transmit and receive arrays. There are several different techniques that are used to increase the performance and capacity of the system. These methods have been described as Bell Labs Layered Space Time Architecture (BLAST). Different architectures are given below.

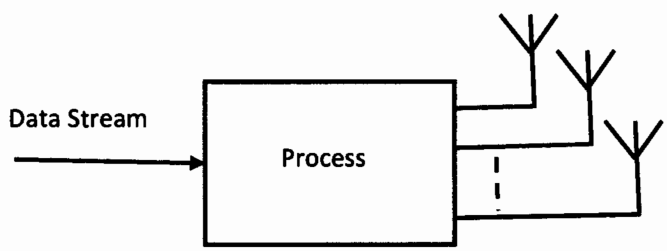

Figure 3: BLAST Architecture

Serial coding: The logic of spatial multiplexing to send the M_T independent symbols in each transmission by using the dimension of space in addition to time. To achieve full order diversity, the encoded bits must be sent over all transmit antennas. This can be performed by using the serially encoding as shown in Figure 4.

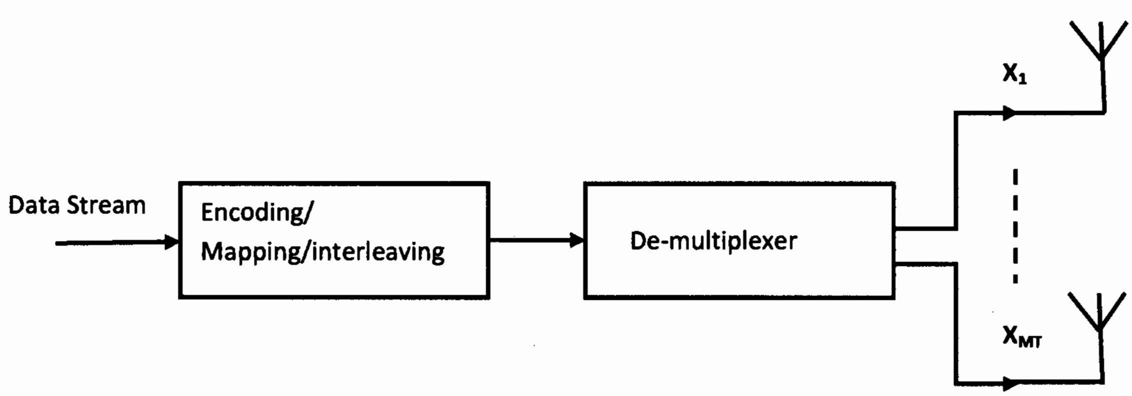

Figure 4: Serial Coding

The bit stream is processed by encoding, interleaving and mapping to a different constellation point. The mapped signal is then de-multiplexed and the multiple streams created are sent over a different transmit antennas. If each code word is adequately long, it can be sent over e all M_T transmitted antennas and received by all M_R receiver antennas. The result is full diversity gain. This system is called vertical encoding (V-BLAST) [4]. This however increases complexity exponentially [12] with the codeword length. The increasing complexity makes it impractical.

Parallel coding: This simple method is described also Bell Laboratories called BLAST (Bell Labs Layered Space Time Architecture) [4]. The architecture of parallel coding is shown in Figure 5. In serial coding, the encoding was performed before de-multiplexing. In the parallel encoding, the data stream is de-multiplexed into the M_T streams and these streams are encoded independently. After these processes send the data stream over the each corresponding transmit antenna as shown in Figure 5. But the problem with the parallel encoding is its complexity increases linearly with number of antennas. These both methods can be used as one transmission technique is called differential BLAST (DBLAST).

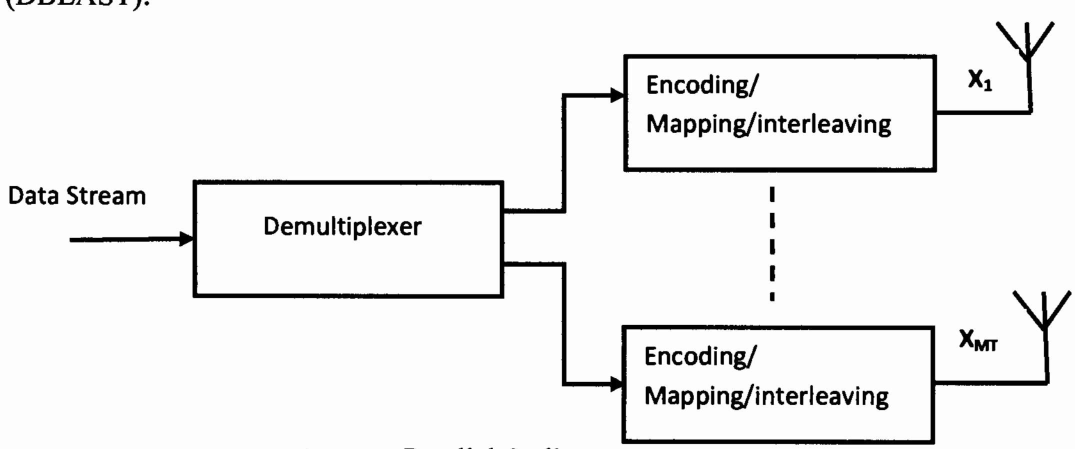

Figure 5: Parallel Coding

VBLAST Receiver and joint detection: Optimal decoding can be achieved by joint detection. The joint detection is the process that transmitted code words from each transmit antenna is received by the all receive antennas. Its detail is available in [13]. The complexity of the receiver can be reduced by using the ordered successive interference cancellation as in [13] as well as in Figure 6. All the streams are ordered in terms of received SNR. The received symbols with highest SNR are detected first, treating the other streams as interference. The estimated symbols of the stream with highest SNR are extracted out and the process is repeated for the stream with the next highest SNR while other treating uncancelled streams as interference. This process is repeated again and again until all M_T transmitted streams have been estimated. In the receiver each transmitted codeword individually decode, so the resulting receiver complexity is to be linear with the number of transmitted codeword. The solution of that complexity is DBLAST where the transmission of data streams rather than the independent codeword on individual antenna, the codeword are rotated on antennas, so that the codeword is extended over all M_T transmit antennas.

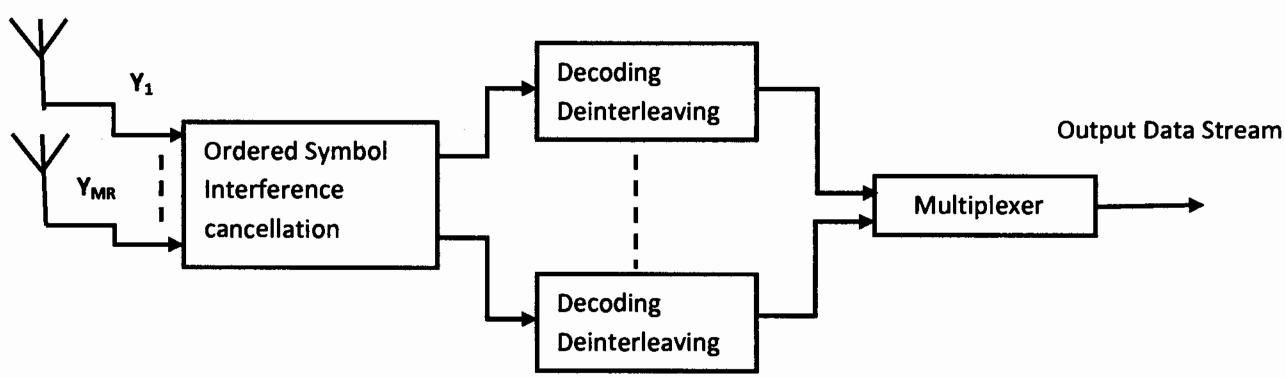

Figure 6: BLAST Receiver and joint detection

1.4 Capacity of MIMO channel

The capacity of wireless channel is very important because it dictates the maximum data rate that can be sent on a channel with arbitrarily small bit error probability. Maximum achievable capacity was discovered by Claude Shannon [1]. The foundation was totally mathematical and depends on mutual information between output and input of the channel. Shannon defined capacity as the maximization of mutual information. We can find the channel capacity by Shannon's famous formula $C = B \times \log_2(1 + SNR)$, where *C* is channel capacity, *B* is bandwidth of the signal and *SNR* is the signal to noise ratio and that was quite a revolutionary idea in communication sectors. In the following sections we will see the capacity advantages of MIMO wireless system and apply Shannon's capacity on a MIMO wireless channel.

1.4.1 Capacity of MIMO channel

The capacity of MIMO wireless channel is extended from the mutual information formula of a SISO system

$$C = \max_{p(s)} I(s; y)$$

above equation shows that the channel capacity can be achieved by maximizing the mutual information between s and y. For MIMO system it can be written as

$$C = \max_{P(s):P \leq P_T} I(s: y)$$

In this P is the average power transmitted by the single transmitter and P(s) is the distribution of the input signal s.

For a flat fading MIMO system, the relation between input and output is

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n} \qquad 1.4$$

Now we assume that channel **H** is known at the receiver. Channel state information (CSI) at the receiver can be keep up through training, but **H** is random variable. The mutual information between input and output is [12]

$$C = \max_{P(s)} I(\mathbf{s};\mathbf{y}) = \max_{P(s)} h(\mathbf{y}) - h(\mathbf{y}|\mathbf{s}) \qquad 1.5$$

Using the relationship between entropy and mutual information

$$I(\mathbf{s};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{s}) \qquad 1.6$$

$$= h(\mathbf{y}) - h(\mathbf{Hs}+\mathbf{n}|\mathbf{s})$$

$$= h(\mathbf{y}) - h(\mathbf{n}|\mathbf{s})$$

$$= h(\mathbf{y}) - h(\mathbf{n}) \qquad 1.7$$

h (.) Shows the differential entropy but the covariance matrix of the received y is given by

$$E[\mathbf{yy^H}] = E[(\mathbf{Hs}+\mathbf{n})(\mathbf{Hs}+\mathbf{n})^{\mathbf{H}}]$$

$$\mathrm{R_y} = \mathrm{E}[(\mathbf{Hs}+\mathbf{n})(\mathbf{s^H H^H}) + \mathbf{n^H}] \qquad 1.8$$

$$= \mathrm{E}[\mathbf{Hss^H H^H} + \mathbf{Hsn^H} + \mathbf{ns^H H^H} + \mathbf{nn^H}] \qquad 1.9$$

$$= \mathbf{HR}_s\mathbf{H}^H + 0 + 0 + \mathrm{N}_o\mathbf{I}_{M_R} \qquad 1.10$$

$$= \mathbf{H}R_s\mathbf{H}^H + \mathrm{N_0}\mathbf{I}_{MR}$$

R_y is output correlation matrix and R_s is input corellation matrix. As the normal distribution maximizes entropy

$$\mathbf{I(s:y)} = \mathrm{B}\log_2 [\mathbf{HR}_s\mathbf{H}^H + \mathbf{I}_{M_R}] \qquad 1.11$$

$$C = \max_{P(s)} B\log_2 \det(\mathbf{HR}_s\mathbf{H}^H + \mathbf{I}_{M_R}) \qquad 1.12$$

In general channel knowledge at the transmitter is difficult to obtain practically. The channel information can be fed back from the receiver to the transmitter or the channel could be estimated from the reverse link. In any case, it is difficult to obtain accurate channel information at the transmitter as the information obtained suffers from limited feedback and channel variations. If channel is unknown at the transmitter then the capacity is achieved by transmitting isotropically, i.e. no particular direction is favored and the capacity can be expressed in terms of the single value decomposition (SVD) and by using the unitary matrix properties [20] as

$$C = \sum_{i=1}^{m} \log_2(1 + \lambda_i) \qquad 1.13$$

where m is rank of **H** and λ_i are the eigenvalues of $\mathbf{HH}^{\mathrm{H}}$ which are positive by virtue of its positive semidefinite structure. Clearly in equation 1.11 capacity of the MIMO channel is sum of capacities of multiple SISO channels. If channel is known at the transmitter, then we can use the "water filling algorithm" [12] to optimally allocate power among the eigenmodes of the channel to maximize the transmission rate.

1.5 Capacity of Time varying MIMO Channel

When considering the capacity of time varying MIMO channel, it is customary to look at two extreme situations. The first situation is when the channel is fast varying and is characterized by ergodic capacity. If the channel is slowly varying, it is modeled as quasi-state, i-e constant in the transmit block and characterized by outage capacity.

Ergodic Capacity: When channel is sufficiently fast varying then we can expect a codeword to experience enough channel states so that by the law of large numbers, we can expect the average rate of information transfer to be

$$C = \max_{R_s} E_H[B\log_2 \det(\mathrm{I}_{mr} + \mathbf{H}\mathrm{R}_{ss}\mathbf{H}^H)] \quad 1.14$$

where the expectation is applied on channel matrix's distribution.

Outage capacity: The Ergodic Capacity has been used as a measure for the spectral efficiency of the MIMO channel. The capacity under channel ergodicity that defined as the average of the maximal value of the mutual information between the transmitted and the received signal, where the maximization was carried out with respect to all possible transmitter statistical distributions. Another measure of channel capacity that is frequently used is *outage capacity*. With outage capacity, the channel capacity is associated to an outage probability. Capacity is treated as a random variable which depends on the channel instantaneous response and remains constant during the transmission of a finite-length coded block of information. If the channel capacity falls below the outage capacity, there is no possibility that the transmitted block of information can be decoded with no errors, whichever coding scheme is employed. The probability that the capacity is less than the outage capacity denoted by C_{outage} is q. This can be expressed in mathematical terms by Prob $\{C \leq Coutage\} = q$.

The ergodic capacity can be understood as the mean of this random variable while the outage capacity is a measure of its spread. If we are willing to accept an outage probability of p_{outage}, then the outage capacity is the rate R such that $p\left(R > log_2 \det(I+HH^H)\right) < p_{outage}$.

1.6 Multiuser MIMO

The mathematical model for a MIMO system is quite broad in the sense that the very same model also applies to a multiuser setting with multiple transmitters / receivers. The basic Multiuser MIMO (MU-MIMO) is shown in figure 7. Some simple examples of multiuser systems include the uplink in a cellular setting, and the downlink in a cellular system [12]. Our main emphasis in this thesis will be on the single user setting.

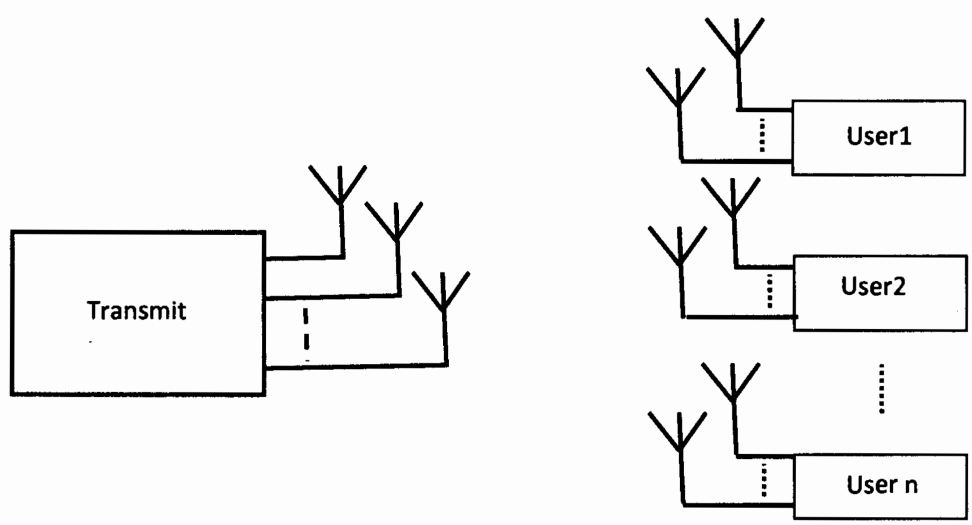

Figure 7: Multiuser MIMO

1.7 Conclusion

We have summarized on MIMO system model and system equation. Performance of system can be improved by using the MIMO system. It enhances the Diversity Gain, Beam Forming, Spatial Multiplexing Gain and Array Gain. Some special techniques are used for multiplexing like serial coding and parallel coding. V-BLAST and D-BLAST also are used for coding and decoding. Capacity of MIMO wireless channel can be improved by using multiple antennas at the transmitter and receiver. Capacity of MIMO channel is the sum of SISO channel.

Chapter 2

MIMO Receivers

2.1 Introduction

This chapter is devoted to MIMO receivers; with special focus on single-user systems. Our focus is to discuss the MIMO receiver like linear receiver and nonlinear receiver. Linear receivers are Zero Forcing (ZF) and Minimum Mean Square Error (MMSE), while linear receivers are Maximum Likelihood (ML) and Sphere Decoder (SD). Since some of them may exhibit a complexity that makes them unpractical. So it is important to seek receivers that achieve a close-to-optimum performance while keeping a moderate complexity. A basic assumption in this chapter is that channel state information is available at the receiver.

This chapter organized as follows. Section 2.2 describes receiver diversity. Section 2.3 describes MIMO equalizers (ZF, MMSE, SIC, ML and SD) and their performance.

2.2 Receiver diversity

Receiver diversity can be achieved by combining the different fading path. These fading paths are achieved from multiple receive antennas and are combined to obtain the desired signal. To achieve the desire signal it will pass through a demodulator. Demodulator separates the desire signal.

In the receiver there are two types of gain: diversity gain and array gain. The diversity gain is the coherent combining of multipath received signals [14] while Array gain is collecting the maximum energy by using the directional antennas. It has been already discussed in section 1.3.3.

The signal combining can be performed in different ways as in Figure 8, which is called combining techniques.

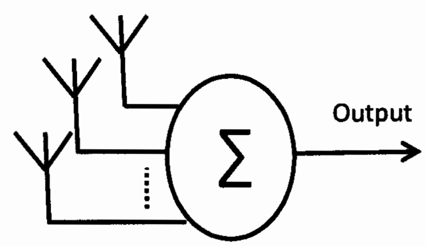

Figure 8: Diversity combining

Different combining techniques are [4] are selection combining, threshold combining, maximal ratio combining, and equal gain combining. Using these techniques equalizers are designed.

2.3 Equalizers

Delay spread in channels causes Inter-symbol interference (ISI). ISI creates a problem for the receiver because the received signal does not resemble the transmitted signal. To mitigate the effect of ISI, equalizers are used. Both the noise and signal pass through the equalizer which can enhance the noise power. Linear equalizers tend to enhance the noise whereas nonlinear equalizers do not do that reasons have been discussed in the next sections. A second issue relating to receiver design is the complexity. To mitigate the ISI at the receiver, the equalizer must have an estimate of impulse response of the channel. But the wireless channel changes with time, so the equalizer must track the channel by updating its estimates periodically as the channel varies. Training and tracking is called adaptive equalization. Since the equalizer must adapt with the changes of the channel. But the problem is when the channel is fast fading. In addition to dealing with channel variations, the equalizer must also contend with noise enhancement as it seeks to remove interference.

Equalizers are broadly classified into linear equalizers and non linear equalizers. Linear equalizers are Zero Forcing (ZF) and Minimum Mean Square Error (MMSE), and non linear equalizers are Maximum Likelihood and decision feedback equalizers.

2.3.1 Zero Forcing (ZF)

MIMO system can be interpreted as a system of equations. In these equations, the unknown variables are transmitted symbols. A system of equations can be solved if the number of unknowns is equal to the number of equations. If the system of equations is tall, an approximate solution can be found by invoking least squares theory. This involves pre-multiplying the system of equations by the inverse of the tall matrix **H.** The resulting solution is known as the Zero Forcing (ZF) solution. Zero Forcing is a linear MIMO detection technique because it involves a linear operation of pre-multiplying by a matrix. The received signal at the receiver is

$$y = \mathbf{H}\mathbf{s} + n \qquad 2.1$$

H is channel matrix, ***s*** is transmitted vectors, ***y*** is out put vector and ***n*** is noise added by the channel. To achieve the estimate of s multiply the inverse of channel matrix

$$\hat{\mathbf{s}} = \mathbf{H}^{-1}\mathbf{y} \qquad 2.2$$

where $\hat{\mathbf{s}}$ is estimation of $\mathbf{s}$. For detection of each stream via zero forcing, all other streams are considered as interference [13]. To eliminate the interference we have to perform nulling of the interferers by linearly weighting the received signals. For nulling the interferers, we choose weight vectors $\mathbf{w}_i^T$ $(i = 1,2, \ldots M_T)$ referred to nulling vectors [13]

$$\mathbf{w}_i^T \mathrm{h_k} = \begin{Bmatrix} 0 & i \neq k \\ \mathbf{1}^T & i = k \end{Bmatrix} = \begin{bmatrix} h_{11} & 0 & 0 \\ 0 & h_{22} & 0 \\ 0 & 0 & h_{33} \end{bmatrix}$$

where h_k is k^{th} column of **H**. Let $\mathbf{w}_\mathrm{i}^\mathrm{T}$ is the i^{th} row of a **W** matrix, then

$$\mathbf{WH} = \mathrm{I}_{M_T} \qquad 2.3$$

where **W** is the weight matrix that cancel out the effect the channel matrix. So by forcing the interference to zero, **s** can be estimated. If **H** is rectangular matrix then **W** becomes equal to the pseudo inverse of **H**. It denoted by $\mathbf{H}^\dagger$

$$\mathbf{W} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H = \mathbf{H}^\dagger \qquad 2.4$$

If **H** is to be assumed iid (independent identically distributed), then pseudo inverse exists, so equation 2.1 become

$$\mathbf{H}^\dagger\mathbf{y} = \mathbf{H}^\dagger\mathbf{Hs} + \mathbf{H}^\dagger\boldsymbol{n} \qquad 2.5$$

we know that $\mathbf{H}^\dagger\mathbf{H} = \mathrm{I}$, so

$$\hat{\mathbf{s}} = \mathbf{s} + \mathbf{H}^\dagger\mathbf{n} \qquad 2.6$$

From the equation 2.6, the major disadvantage of the zero forcing is to cause the noise to be enhanced if **H** is ill conditioned.

2.3.1a Performance and Complexity

The performance of Zero Forcing equalizer is shown Figure 9 in the form of bit error rate (BER). The zero forcing equalizer requires computation of the pseudo inverse of a tall matrix. If the channel does not change significantly, the previous calculated pseudo inverse can be used for next symbol calculation. On the other hand if channel is fast fading then we have to calculate pseudo for almost every symbol.

2.3.2 Minimum Mean Square Error (MMSE)

In this approach a vector **s** is to be estimated on the basis of received vector by minimizing the mean square error between the detected vector and the transmit vector. It would seem strange that the mean square error can be minimized at the receiver but the receiver is not aware of what was transmitted. However, the expectation operation is used to calculate the mean square error over all possible transmission combinations [13].

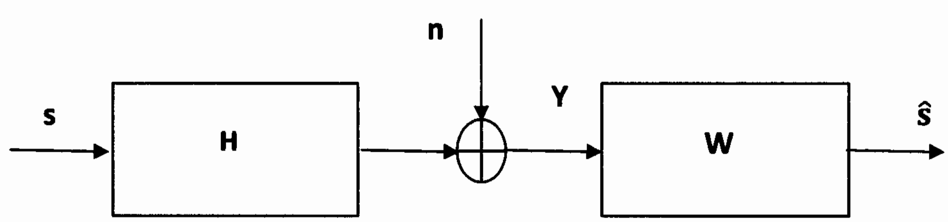

Figure 9: MMSE Architecture

To overcome the noise enhancement caused by zero forcing, we have to design such an equalizer that maintains a balance between ISI suppression and noise enhancement. Figure 9 shows a MIMO receiver employing a linear equalizer. We have

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n} \qquad 2.11$$

where $\boldsymbol{s}$ is transmitted vector that is equal to

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_k \\ s_{k+1} \\ \vdots \\ s_n \end{bmatrix}$$

$\mathbf{W}$ is a linear processing that is applied on received signal vector to estimate the transmitted signal. To obtain the MMSE solution, we choose $\mathbf{W}$ to minimize the mean square error. The estimate of $\boldsymbol{s}$ is equal to

$$\hat{\mathbf{s}} = \mathbf{Wy} \qquad 2.12$$

where $\hat{\mathbf{s}}_k = \mathbf{w}_k^{\mathrm{T}}\mathbf{y}$, this shows that k^{th} row of $\mathbf{W}$. The MMSE cost function is

$$J = E\{(\hat{\mathbf{s}}_k - \mathbf{s}_k)^2\} \qquad 2.13$$

$$J = E\left\{\left(\mathbf{w}_k^{\mathrm{T}}\mathbf{y} - \mathbf{s}_k\right)^2\right\} \qquad 2.14$$

$$J = E\{(\mathbf{w}_k^T(\mathbf{Hs} + \mathbf{n}) - \mathbf{s}_k)^2\} \qquad 2.15$$

$$J = E\{(\mathbf{w}_k^T\mathbf{Hs} + \mathbf{w}_k^T\mathbf{n} - \mathbf{s}_k)(\mathbf{s}^T\mathbf{H}^T\mathbf{w}_k + \mathbf{n}^T\mathbf{w}_k - \mathbf{s}_k)\} \qquad 2.16$$

$$J = E\{(\mathbf{w}_k^T\mathbf{Hss}^T\mathbf{H}^T\mathbf{w}_k + \mathbf{w}_k^T\mathbf{Hsn}^T\mathbf{w}_k - \mathbf{w}_k^T\mathbf{Hss}_k + \mathbf{w}_k^T\mathbf{ns}^T\mathbf{H}^T\mathbf{w}_k + \mathbf{w}_k^T\mathbf{nn}^T\mathbf{w}_k - \mathbf{w}_k^T\mathbf{ns}_k - \mathbf{s}_k\mathbf{s}^T\mathbf{H}^T\mathbf{w}_k - \mathbf{s}_k\mathbf{n}^T\mathbf{w}_k + \mathbf{s}_k^2)\} \qquad 2.17$$

By applying the expectation on cost function

$$J = (\mathbf{w}_k^T \mathbf{H} E\{\mathbf{s}\mathbf{s}^T\}\mathbf{H}^T\mathbf{w}_k + \mathbf{w}_k^T \mathbf{H} E\{\mathbf{s}\mathbf{n}^T\}\mathbf{w}_k - \mathbf{w}_k^T \mathbf{H} \mathrm{E}\{\mathbf{s}\mathbf{s}_k\} + \mathbf{w}_k^T E\{\mathbf{n}\mathbf{s}^T\}\mathbf{H}^T\mathbf{w}_k + \mathbf{w}_k^T E\{\mathbf{n}\mathbf{n}^T\}\mathbf{w}_k - \mathbf{w}_k^T E\{\mathbf{n}\mathbf{s}_k\} - E\{\mathbf{s}_k\mathbf{s}^T\}\mathbf{H}^T\mathbf{w}_k - E\{\mathbf{s}_k\mathbf{n}^T\}\mathbf{w}_k + E\{\mathbf{s}_k^2\}) \quad 2.18$$

$E\{\mathbf{s}\mathbf{s}^T\} = \mathrm{I}$ and $E\{\mathbf{s}\mathbf{n}^T\} = 0$ because signal and noise are uncorrelated. The equation 2.18 become

$$J = \mathbf{w}_k^T \mathbf{H}\mathrm{I}\mathbf{H}^{\mathrm{T}}\mathbf{w}_{\mathrm{k}} + 0 - 2\mathbf{w}_k^T \mathbf{H}\mathbf{1}_{\mathrm{k}} + 0 + \mathbf{w}_{\mathrm{k}}^{\mathrm{T}}\sigma_n^2 \mathrm{I}\mathbf{w}_k + 0 + 0 + 1 \qquad 2.19$$

$$J = \mathbf{w}_k^T \mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{w}_{\mathrm{k}} - 2\mathbf{w}_k^T \mathbf{H}1_{\mathrm{k}} + \mathbf{w}_{\mathrm{k}}^{\mathrm{T}}\sigma_n^2 \mathrm{I}\mathbf{w}_k + 1 \qquad 2.20$$

Taking the derivative of cost function J with respect to weight vector and put equals to zero

$$\frac{\partial J}{\partial w_k} = 2\mathbf{H}\mathbf{H}^T\mathbf{w}_k - 2\mathbf{H}1_k + 2\sigma_n^2\mathbf{w}_k = 0 \qquad 2.21$$

$$(\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})\mathbf{w}_k = \mathbf{H}\mathbf{1}_k \qquad 2.22$$

$$\mathbf{w}_k = (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})^{-1}\mathbf{H}\mathbf{1}_k \qquad 2.23$$

From the equation 2.17

$$\mathbf{w}_1 = (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})^{-1}\mathbf{H}\mathbf{1}_1$$

$$\mathbf{w}_2 = (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})^{-1}\mathbf{H}\mathbf{1}_2$$

and

$$\mathbf{w}_n = (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})^{-1}\mathbf{H}\mathbf{1}_n$$

$$\mathbf{W}_{MMSE}^T = [\mathrm{w}_1 \ \mathrm{w}_2 \dots \mathrm{w}_n] = (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \sigma_n^2\mathrm{I})^{-1}\mathbf{H} \qquad 2.24$$

Intuitively we can say MMSE solution reduces noise enhancement. By minimizing the mean square error between the transmitted and detected symbols, it seeks to balance the effect of the interference with the noise enhancement to provide. In the extreme situation with no noise, this equation reduces to the zero forcing solution (if $\sigma_n = 0$).

2.3.2a Complexity and performance

The complexity of the MMSE approach is identical to the ZF approach. The performance is better, especially at lower signal to noise ratios. From equation 2.24, it has been shown that when $\sigma_n = 0$ MMSE becomes ZF.

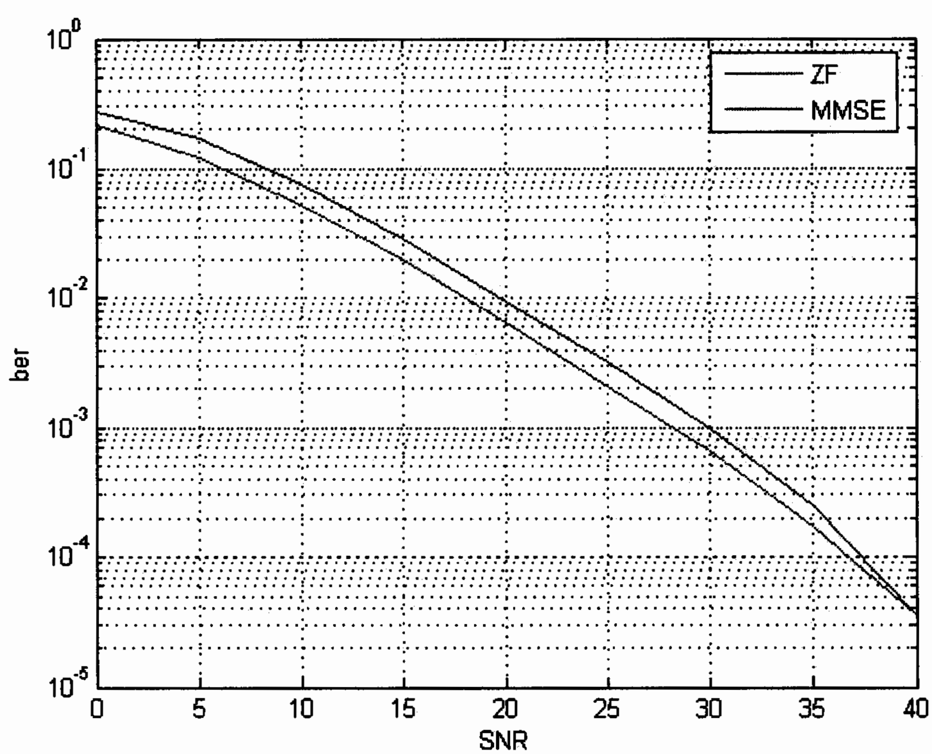

Figure 10: Performance graph of MMSE and ZF

2.3.3 Successive interference cancellation (SIC)

The performance of the linear receivers can be improved by using detected data to cancel out interference. This does require the detection process could be ordered. The idea in SIC is to subtract the effect of estimated symbols from the received signal vector $\boldsymbol{y}$. As

$$\hat{\mathbf{s}}_k = \mathbf{W}_k^T\mathbf{y} \qquad 2.25$$

$$\hat{\mathbf{s}}_1 = \mathbf{W}_1^T\mathbf{y} = \mathbf{W}_1^T(\mathbf{Hs}+\mathbf{n}) \qquad 2.26$$

$$= \mathbf{w}_1^T\mathbf{Hs} + \mathbf{w}_1^T\mathbf{n} = \mathbf{s} + \underbrace{\mathbf{w}_1^T\mathbf{n}}_{noise} \qquad 2.27$$

and

$$\hat{\mathbf{s}}_2 = \mathbf{W}_2^T(\mathbf{y} - \mathrm{h}_1\hat{\mathrm{s}}_1) = \mathbf{w}_2^T(\mathbf{Hs}+\mathbf{n}) - \mathbf{w}_2^T\mathrm{h}_1\hat{\mathrm{s}}_1) \qquad 2.28$$

$$= \mathbf{s} + \mathbf{w}_2^T\mathbf{n} - \mathbf{w}_2^T\mathrm{h}_1\hat{\mathrm{s}}_1 \qquad 2.29$$

and so on. This subtraction process is repeated until all symbols have been detected. Initially, the first symbol is detected by treating everything else as interference. Then the contribution of this estimated symbol to the received signal is subtracted out. The process repeats in this manner so that in detecting a bit, previously detected bits are canceled out while subsequent bits are suppressed by an equalizer. This process is also called Layer Peeling [2]. This is a nonlinear receiver design because the process of taking decisions on detected symbols and canceling out their interference in the received signal is a nonlinear operation. It can potentially suffer from error propagation. To overcome the effects of error propagation, special attention has to be paid to the order in which bits are detected. The complexity of this approach is not much more than that of the linear equalizers.

2.3.3a ZF-SIC

Successive interference cancelation receivers are based on the linear equalizers already discussed and their aim is to improve the performance of linear equalizers without running up a significant bill on complexity. There is a very large performance gap between the performance of Maximum Likelihood (ML) and linear equalizers. However, Maximum Likelihood also has much more complexity associated with it. SIC based receivers are compromise between these two techniques.

For a SIC based receiver, the first step is to order the transmit streams from the strongest to the weakest. The strongest stream will be detected first treating all other streams as interference and its contribution to the received signal will be subtracted out. Then the next strongest stream is detected and so on. The process of determining the sequence to detect the streams is called ordering. It is necessary to minimize the impact of error propagation.

To determine detection order, the covariance matrix of estimated error $\mathbf{s} - \hat{\mathbf{s}}$ is used. The covariance matrix is

$$\mathrm{R}_{ss} = E[(\mathbf{s} - \hat{\mathbf{s}})(\mathbf{s} - \hat{\mathbf{s}})^H] = \sigma_n^2(\mathbf{H}^H\mathbf{H})^{-1} = \sigma_n^2\mathbf{P} \qquad 2.30$$

or by pseudo inverse

$$\mathbf{P} = \mathbf{H}^{\dagger}(\mathbf{H}^{\dagger})^{-1} \qquad 2.31$$

Let $\hat{s}_j$ is j^{th} entry of $\hat{s}$ then the best approximate is for which P_{jj} is the smallest (i-e the j^{th} diagonal element of **P**). The algorithm is summarized as follows

i. Ordering: find out the transmitted signals with lowest error variance.

ii. Interference nulling: detect the strongest remaining stream from the received signal while nulling out the all weaker signals.

iii. Interference cancellation: subtract the contribution of the detected stream to the received signal.

iv. Repeat steps 2 and 3 until all streams have been detected.

QR decomposition: From the equation 2.1, we know that

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n} \qquad 2.32$$

H is channel matrix. We use the **QR** decomposition of a channel matrix, so $\mathbf{H} = \mathbf{QR}$. **Q** is unitary matrix, i.e. $\mathbf{Q} = \mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{H}} = \mathrm{I}$ and **R** is upper triangular matrix.

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \cdots & r_{1n} \\ 0 & r_{22} \cdots & r_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_{nn} \end{bmatrix}$$

put the value of Q and R in equation 2.21

$$\mathbf{y} = \mathbf{QRs} + \mathbf{n} \qquad 2.33$$

by pre-multiplying with $\mathbf{Q}^{\mathrm{H}}$ and we will get

$$\mathbf{Q}^{\mathrm{H}}\mathbf{y} = \mathbf{Rs} + \mathbf{Q}^{\mathrm{H}}\mathbf{n}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \cdots & r_{1n} \\ 0 & r_{22} \cdots & r_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} + \mathrm{Q^H n} \qquad 2.34$$

From the equation 2.34

$$\mathbf{y}_n = r_{nn}\mathbf{s}_n + noise \qquad 2.35$$

And

$$\hat{\mathbf{s}}_n = sign\left(\frac{y_n}{r_{nn}}\right) \qquad 2.36$$

Similarly

$$\hat{\mathrm{s}}_{n-1} = y_{n-1} - \frac{r_{n-1,n}\hat{\mathrm{s}}_n}{r_{n-1,n-1}} \qquad 2.37$$

We can estimate $\hat{\mathrm{s}}_n$ to $\hat{\mathrm{s}}_1$ using these steps.

2.3.3b MMSE-SIC

MMSE-SIC is similar to zero forcing SIC with one small difference. Instead of nulling the un-canceled interfering streams, they are suppressed by using the MMSE criterion. Covariance matrix of estimation error $s - \hat{\mathrm{s}}$ will be used to find the ordering for better detection. Recall from the previous section the equation

$$\mathrm{R}_{ss} = E[(\mathbf{s} - \hat{\mathbf{s}})(\boldsymbol{s} - \hat{\mathbf{s}})^H] = \sigma_n^2\mathbf{P} \qquad 2.38$$

where $\mathbf{P} = (\mathbf{HH}^\mathrm{T} + \sigma_n^2\mathrm{I})^{-1}$.

2.3.4 Maximum Likelihood Equalizer (MLE)

Maximum Likelihood Equalizer searches the maximum likelihood sequence among all transmitted vectors **s.** The maximum likelihood transmitted vectors are found using the following equation

$$\mathbf{s_{ml}} = \arg \min_{s_i \in \{s_1,\dots,s_k\}} \|\mathbf{y} - \mathbf{Hs_i}\|^2 \qquad 2.39$$

where $\{s_1, \dots, s_k\}$ are the set of possible transmitted vectors. K is the number of transmitted vectors that is $k = \mathrm{M}^{M_T}$ where M is the constellation points [10]. Note that it is not necessary that $M_T \leq M_R$ for Maximum Likelihood (ML) method.

Our goal here is to find the maximum likely symbol $\mathbf{s_i}$ from the given ensemble space $\{s_1, \dots, s_k\}$. To minimize the bit error rate, we have to maximize the probability $\mathbf{Pr(s = s_i|y)}$. This method is called MAP (maximum A posteriori probability) decoding. To find $\mathbf{Pr(s_i|y)}$, by applying the Bayes' rule we may write [15]

$$\mathbf{Pr(s_i|y)} = \frac{\mathbf{Pr(y|s_i).Pr(s_i)}}{\mathbf{Pr(y)}} \qquad 2.40$$

In the above equation $\mathrm{Pr}(\mathbf{y}|\mathbf{s_i})$ is conditional probability density function of a vector that has been sent. $\mathrm{Pr(s_i)}$ is the i^{th} vector's probability that was sent. When there is no apriori knowledge available about the transmit signal, it can be assumed that all possible transmit vectors are equally probable. So in this case, MAP becomes ML detection. From the equation 2.40, it can be seen that denominator is independent from s_i so the decision based on maximizing $Pr(\mathbf{s}_i|\mathbf{y})$ is equivalent to finding $\mathbf{s}_i$ that maximize $Pr(\mathbf{y}|\mathbf{s}_i)$. So there will be following probability density function (pdf):

$$Pr(\mathbf{y}|\mathbf{H}, \mathbf{s_i}) = \det(\pi\mathbf{Q})^{-1} exp\left(-(\mathbf{y} - \mathbf{Hs_i})^{\mathrm{H}}\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{Hs_i})\right) \qquad 2.41$$

where **Q** is a covariance matrix and will be equals to

$$\mathbf{Q} = E[(\mathbf{y}-\mathbf{m})(\mathbf{y}-\mathbf{m})^{\mathbf{H}}] \qquad 2.42$$

$$= E[(\mathbf{y}-\mathbf{Hs_i})(\mathbf{y}-\mathbf{Hs_i})^{\mathbf{H}}] = E[\mathbf{nn^H}] = \sigma_n^2 \mathbf{I}_{M_r} \qquad 2.43$$

so the resultant conditional pdf is

$$Pr(\mathbf{y}|\mathbf{H}, \mathbf{s_i}) = \frac{1}{(\pi\sigma_n^2)^{M_r}} \exp\left(-\frac{1}{\sigma_n^2}(\mathbf{y}-\mathbf{Hs_i})^{\mathbf{H}}(\mathbf{y}-\mathbf{Hs_i})\right) \qquad 2.44$$

Consequently, the decision to find the value of maximum conditional probability $\Pr(\mathrm{s_i}|\mathrm{y})$ is equivalent to

$$\arg \max_{s_i \in \{s_1,\dots,s_k\}} Pr(\mathbf{y}|\mathbf{H}, \mathbf{s_i}) = \arg \min_{s_i \in \{s_1,\dots,s_k\}} ||\mathbf{y}-\mathbf{Hs_i}||^2 = \mathbf{s}_{\mathrm{ml}} \qquad 2.45$$

MLE is optimal solution in term of Bit Error Rate (BER) under the assumption is that all the transmit sequences are equally likely.

2.3.4 Complexity

In the MIMO receiver the complexity of Maximum Likelihood (ML) detection increases exponentially in the number of transmit antennas and the constellation size. This is because all possible transmitted vectors must be checked to determine which is most likely with the received signal vectors [13].

Chapter 3

Low complexity ML methods

3.1 Introduction

In the MIMO receiver the Maximum Likelihood (ML) decoder decodes the transmitted signal by comparing the received signal vector with each possible combination of transmitted vectors. The complexity of the decoder grows exponentially in the number of transmit antennas. The complexity of maximum likelihood receiver is NP hard [11].

The idea behind Sphere Decoder (SD) is to reduce the computational complexity of Maximum Likelihood (ML) receiver. The sphere decoder searches only those noiseless received vectors that lie in the hyper sphere whose radius R around the received signal as opposed to maximum likelihood which searches among all possible transmissions [11].

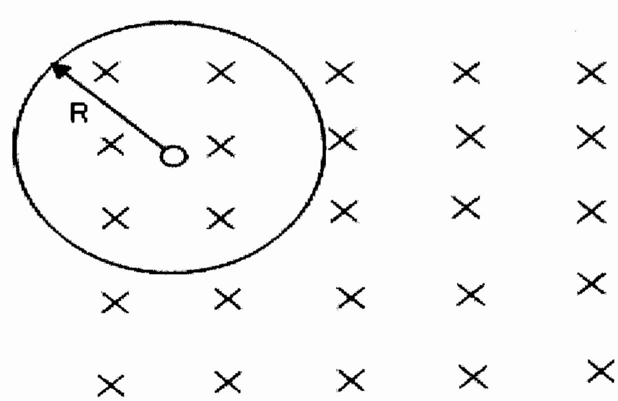

Figure 11: Lattice point Architecture

Unlike the brute force maximum likelihood receiver which has a fixed complexity that is exponential in the number of transmit antennas, the complexity of the sphere decoder depends on several different factors such as the received signal, the search ordering, the sphere radius. In contrast to the variable complexity of the sphere decoder, a modification called the Fixed Sphere Decoder has also been proposed which has a fixed decoding complexity.

3.2 Sphere decoder (SD)

The SD can achieve ML performance at reduced complexity by searching for the closest point among the possible Lattice points that lie within a hyper sphere of radius R around the received vector x [13]. Mathematically, the sphere decoder solves the problem [12]:

$$\hat{\mathbf{s}} = \arg \min_{s_i \in O^{M_T} : \|\mathbf{y}-\mathbf{Hs}\|^2 \leq R^2} \|\mathbf{y} - \mathbf{Hs}\|^2 \qquad 3.1$$

Where $\mathbf{H} \in \mathcal{R}^{m \times n}$, $x \in \mathcal{R}^{n \times 1}$, and O^{M_T} shows the m dimensional lattice points. O Shows the transmit constellation. The SD only searches for the ML transmit signal from among the lattice points that lie within a hyper sphere of radius R around the received signal. In contrast the ML decoder searches through all possible lattice points. A key question here is how does the SD know which lattice points are within a sphere centered at the received signal x. Calculating the distance of each lattice point to the received vector reduces to the exhaustive search carried out by the ML decoder. The SD can cleverly find the set of lattice points within the hyper sphere without the need for an exhaustive search thereby saving on complexity as compared to exhaustive search ML decoding. To understand this, first consider the simplest possible case. Finding the set of lattice points within a radius of the received signal in one dimensional space is simple. This set is simply the set of lattice points in an interval around the received signal. Similarly, if we know the set of lattice points in a k dimensional space that lie within a hyper sphere, then the possible values of the $(k+1)^{th}$ coordinate for the set of lattice points in the $k+1$dimensional space that lie within a hyper sphere of the same radius is simply an interval. We can thus recursively find and check all lattice points in the hyper sphere. Mathematically, by the Gram Schmidt orthogonalization

$$\mathbf{H} = \mathbf{Q}\begin{bmatrix} \mathbf{R}_{M_T \times M_T} \\ \mathbf{0}_{(M_R - M_T)\times M_T} \end{bmatrix} \quad 3.2$$

Where the unitary matrix is partitioned as $\mathbf{Q} = [\mathbf{Q_1} \quad \mathbf{Q_2}]$ such that $\mathbf{Q_1}$ contains the first M_T columns of $\mathbf{Q}$. With a little manipulation, it can be shown that the hyper sphere constraint becomes [16]

$$R^2 \geq \left\| \underbrace{\mathbf{Q}_1^H \mathbf{y}}_{\mathbf{x}} - \mathbf{Rs} \right\|^2 \quad 3.3$$

This also can be written as [16]

$$R^2 \geq \sum_{K=1}^{M_T} \left(\mathbf{y}_K - \sum_{l=K}^{M_T} \mathbf{r}_{K,l}\, \mathbf{s}_l \right)^2 \quad 3.4$$

Where $\mathbf{y}_K, \mathbf{s}_K$ are the k^{th} element of $\mathbf{y}$, $\mathbf{s}$. Also $\mathbf{r}_{k,l}$ is the $(k,l)^{th}$ element of the upper triangular matrix $\mathbf{R}$. In order to satisfy condition 3.4, a necessary but not sufficient condition is

$$R^2 \geq \left(\mathbf{y}_{M_T - r_{M_T}, M_T s_{M_T}} \right)^2 \quad 3.5$$

This implies the following condition on s_{M_T} [13]

$$\left\lceil \frac{y_{M_T}-R}{r_{M_T,M_T}} \right\rceil \leq s_{M_T} \leq \left\lfloor \frac{y_{M_T}+R}{r_{M_T,M_T}} \right\rfloor \quad 3.6$$

Likewise if we define $\mathbf{y}_{(M_T-1|m_T)} = \mathbf{y}_{M_T-1} - r_{M_T-1,M_T}\mathbf{s}_{M_T}$, we get the following necessary condition on $\mathbf{s}_{M_T-1}$ [13]

$$\left\lceil \frac{y_{(M_T-1|m_T)}-R_{M_T-1}}{r_{M_T-1,M_T-1}} \right\rceil \leq \mathbf{s}_{M_T} \leq \left\lfloor \frac{y_{(M_T-1|m_T)}+R_{M_T-1}}{r_{M_T-1,M_T-1}} \right\rfloor \quad 3.7$$

Where $\mathrm{R}^2_{M_T-1} = \mathrm{R}^2 - \left(\mathbf{y}_{M_T} - \mathbf{r}_{M_T,M_T}\mathbf{s}_{M_T}\right)^2$. The SD proceeds similarly to obtain the lattice points within the hyper sphere.

3.2.1 Complexity of Sphere Decoder

The worst case complexity of the SD is exponential in the dimension M_T just like the ML decoder. However, it has been shown [14] that the expected complexity of the sphere decoder is polynomial. The complexity of the SD fluctuates with the channel conditions and the noise in addition to the ordering employed [14]. The search radius can also be dynamically modified. The search radius may be initialized at ∞ and may be reduced to the distance of the received vector to a possible transmit vector every time a transmit vector is found within the sphere as the ML codeword must have distance less than or equal to this.

The questions we would like to tackle in this work are how to deal with the channel uncertainty in a SD/ML receivers.

3.3 Fixed Complexity sphere decoder (FSD)

The fixed complexity sphere decoder (FSD) is a clever modification to the sphere decoder. Its complexity is fixed – unlike the sphere decoder which has variable complexity that is exponential in the worst case. Like the sphere decoder, the FSD can be considered in terms of tree search. While the sphere decoder searches all points in a hyper sphere, i.e. visits all leaves of the tree which lie within a predetermined radius, the FSD reduces this complexity even further by searching only a fixed number of the points. This is cleverly achieved by solving equalization problems using the zero forcing equalizer we have already discussed. This enables the FSD to overcome the two main problems associated with the sphere decoder, i.e. its variable complexity and its sequential nature. The FSD is immediately parallelizable while the sphere decoder's sequential nature makes it difficult to share the complexity using parallel processing.

Chapter 4

Low Complexity Robust Maximum Likelihood Techniques

4.1 Introduction

If the channel is unknown at the receiver then there are many possibilities of channel. So there is much difficult to maintain the perfect channel information at the receiver. A channel is estimated by iteratively and estimates the transmitted symbol. In this chapter section 4.2 covers channel uncertainty. Section 4.3 is devoted for robust techniques like ML and SD. In the 4.4 low complexities approach is discussed.

4.2 Channel uncertainty

The MIMO channel mixes the input signals. To decouple the input signal at the receiver side requires MIMO detectors. For detection of signal there is need of knowledge of channel at the receiver that is called CSI at receiver. There are many types of receivers that already have been discussed in chapter 2. In the zero forcing equalizer uses pseudo inverse of the channel matrix **H** to obtained the transmitted signal. If **H** is is ill-conditioned, then $\mathbf{H}^{\dagger}$ enhances the noise at the receiver. The MMSE receiver $(\mathbf{H}^{T}\mathbf{H} + \sigma_n^2 I)^{-1}\mathbf{H}^{T}$ can be used instead. We consider the case where the channel matrix **H** is uncertain, i.e.

$$\mathbf{H} = \tilde{\mathbf{H}} + \bar{\mathbf{H}} \qquad 4.1$$

where $\tilde{\mathbf{H}}$ is unknown to the receiver, while $\bar{\mathbf{H}}$ is the CSI available to the receiver. We consider the channel uncertainty to be modeled as an ellipsoid [19] defined by **P,** i.e. the channel is modeled as

$$\mathbf{H} = \bar{\mathbf{H}} + \mathbf{P}[\mathbf{u}_1 \ \dots \ \mathbf{u}_{M_T}] \ \mid \ \|\mathbf{u}_K\|_2 \leq 1 \qquad 4.2$$

Where **P** is a $M_R \times M_P$ matrix and u_K shows the uncertainty from the K^{th} transmit antennas. This is a general model for the channel uncertainty. M_P gives the dimensions of the ellipsoid in which the channel variations lie. It can be anything from zero to M_T.

4.3 Robust Techniques

Now we consider a few simple examples to demonstrate the basic ideas of robust ML techniques, before we investigate the details. We first consider a simple model of the channel uncertainty such that in [19]

$$\mathbf{H} = \begin{cases} \mathbf{H}_1, \text{with probability } 1/2 \\ \mathbf{H}_2, \text{with probability } 1/2 \end{cases} \qquad 4.3$$

The receiver knows channel matrix $\mathbf{H_1}$ and $\mathbf{H_2}$ with the probability $\frac{1}{2}$. However, it is uncertain that which of these two values the channel assumes. A receiver has to be designed according to some criterion like ZF, MMSE, SIC and ML/SD. But which of these realizations $\mathbf{H_1}$ or $\mathbf{H_2}$ does the channel assume? For example if the receiver assumes that the channel is $\mathbf{H_1}$ it is correct half the time but the other half the time its assumption is inconsistent with the actual channel. Such a receiver would exhibit an early error floor which shows that channel uncertainty cannot just be ignored in receiver design especially if it is significant.

A robust approach which takes into account the channel uncertainty involves the receiver decoding the received signal for both possible channel realizations and choosing the best one

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}_K\mathbf{s}\| \quad 4.4$$

$$\mathrm{k} = \arg\min_{l} \left\{\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}_l\mathbf{s}\|\right\} \quad 4.5$$

where the minimization with respect to s is carried out by using a sphere decoder (SD).

4.4 Robust SD technique

For above example, this receiver is equivalent to the two ordinary sphere decoder receivers. There may be N possibilities as in 4.4 and 4.5. So this approach would have a complexity of N sphere decoders. This complexity can be reduced by using the SD receiver as follows in [19] so that there are N-1 linear equalizations and only one sphere decoding

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}_K\mathbf{s}\| \quad 4.6$$

$$\mathbf{H}_k = \arg\min_{\mathbf{H}_l} \|\mathbf{y} - \mathbf{H}_l\hat{\mathbf{s}}\| \quad 4.7$$

$$\hat{\mathbf{s}}_l = \mathbf{H}_l^{\dagger}\mathbf{y} \quad 4.8$$

However, how do we deal with the situation where the channel uncertainty has infinite possibilities? Clearly in 4.6 and 4.7, we cannot search every one of the infinite possibilities.

Now let us consider another simple model. Let the channel be

$$\mathbf{H} = \mathbf{H}_1 + t\mathbf{H}_2 \qquad 0 \le t \le 1 \qquad 4.9$$

This is very simple model for the channel uncertainty. There are infinite possibilities for the channel that it can take on. The uncertainty here is in the scale factor t. Perhaps the technique discussed in the previous section can be applied to this model using the bisection method – however it would be more appropriate to try to search for a lower complexity approach to deal with the uncertainty.

4.5 Low Complexity Techniques to deal with Channel Uncertainty

We consider the following approaches to dealing with the channel uncertainty.

4.5.1 Iterative detection and channel estimation

The most common approach to dealing with channel uncertainty is iterative detection and channel estimation. In the first step, uncertainty is ignored altogether and the data is detected with a receiver designed according to the uncertain channel. In the second step, the channel is re-estimated using detected data in the first step. Then the estimated channel is used to design a receiver to detect the data and so on.

This method is a local optimization algorithm [18, 19]. After a couple of iterations, this algorithm converges to a local solution. The quality of the solution depends on the starting point of the algorithm so that it performs well if there is little uncertainty but performs poorly as the uncertainty becomes significant. Mathematically, the received signal is

$$\underbrace{[x[1] \ \dots \ x[N]]}_{\mathbf{X}} = \mathbf{H} \underbrace{[s[1] \ \dots \ s[N]]}_{\mathbf{S}} + \underbrace{[n[1] \ \dots \ n[N]]}_{\mathbf{N}} \qquad 4.11$$

Estimation of the channel is a linear least squares problem and an estimate of **H** can be found using linear least squares method. The estimate uses the detected data $\hat{\mathbf{S}}$.

$$\hat{\mathbf{H}} = \mathbf{X}\hat{\mathbf{S}}^H(\hat{\mathbf{S}}\hat{\mathbf{S}}^H)^{-1} \qquad 4.12$$

$\hat{\mathbf{H}}$ is estimated by the receiver using its previous estimates of the transmit data $\hat{\mathbf{S}}$. An equalizer designed according to this channel estimate may then be used to form an estimate of the transmit signal for use in channel estimation in the next iteration. While linear least squares has a tendency to average out noise, in this case, we will not get perfect estimate of the channel because the estimation is carried out using detected data which is imperfectly known and may have errors [19]. The received signal may be expressed as

$$\mathrm{x} = \underbrace{\bar{\mathbf{H}}\mathbf{s}}_{\text{signal}} + \underbrace{\mathbf{PUs}}_{\text{interference}} + \mathbf{n} \qquad 4.13$$

where **P** is a matrix that defines an ellipsoid that models the channel uncertainty. Its dimensions are $M_R \times M_P$. In the above equation, **U** is an unknown quantity at the receiver and so the second term acts as interference in the detection process at a receiver that is designed on the assumption that the channel is $\bar{\mathbf{H}}$. So if the sphere decoder (SD) receiver or maximum likelihood (ML) is designed according to $\bar{\mathbf{H}}$, it may be expected that the second term will degrade the performance of receiver. It can be seen in equation 4.13 that the term 'interference' is dependent on the transmit signal **s**. Therefore the approach of treating this term as interference and trying to suppress it is strictly suboptimal. The redeeming feature of iterative detection and channel estimation is that it tries to incorporate this second term into the data detection process so that all received signal power is exploited for data detection. The second term in the equation 4.13 is used in the data detection process in consequent iterations when the channel is re-estimated

4.5.2 Linear Preprocessing

It is clear from the equation 4.13 that the second term initially constitutes interference to the data detection process and will result in poor detection if the term is not suppressed. So another approach to dealing with channel uncertainty would be to suppress the second term using a linear preprocessor. Like iterative detection and channel estimation, this approach is also a suboptimal local optimization algorithm because it tries to suppress a term that is actually data dependant and could aid the data detection process. We now demonstrate the idea. First consider the situation where $M_R \geq M_T + M_P$. Under this assumption, it is possible to choose the such a matrix **W**, whose rank should be equal to M_T and also lies in the left null space of **P**. So that

$$\mathbf{WP} = \mathbf{0} \qquad 4.14$$

When this preprocessor is applied at the received signal, the 'interference' term entirely vanishes and we get

$$\mathbf{Wy} = \underbrace{\mathbf{W\bar{H}}}_{\mathbf{G}}\,\mathbf{s} + \mathbf{Wn} \qquad 4.15$$

It is also noted by the equation 4.15, that intention of linear preprocessor is to just eliminate those terms that are created by the channel uncertainty. The purpose of the linear preprocessor is not to equalize or detect the data but just to suppress the uncertainty. Now we can apply a SD/ML receiver designed according to the effective channel **G** to detect the transmit data. This will make sure that the 'interference' term does not degrade the detection of the transmitted signal.

For example, the preprocessor can be applied for the simple model for channel uncertainty that has been shown in 4.1. We can rewrite that model as $\mathbf{H_2} = \mathbf{H_1} + \Delta\mathbf{H}$. If $\Delta\mathbf{H}$ has rank r that satisfies the constraint $M_R \geq M_T + r$, we could write $\Delta\mathbf{H}$ as in [14] as the product of the two rank r matrices $\Delta\mathbf{H} = \mathbf{AB}$. From this equality, that is possible to find a linear preprocessor $\mathbf{W}$ in the left null space of $\mathbf{A}$. By using this preprocessor to $\mathbf{H_2}$, we get

$$\mathbf{W(H_1 + AB) = WH_1} \qquad 4.16$$

As $\mathbf{WA} = 0$, the equation 4.16 shows, if the preprocessor is applied to $\mathbf{H_1}$ then it would give $\mathbf{WH_1}$. Therefore in 4.1 regardless of which realization the channel assumes, the sphere decoder (SD) can be designed for effective channel $\mathbf{WH_1}$ after applying the preprocessor.

It is not possible to always have $M_R \geq M_P + M_T$, so this approach will now be modified so this condition will not be longer necessary. It is to be noted that, this condition was mandatory to entirely eradicate the 'interference' that was caused through the channel uncertainty. In relaxing this situation, we will not able to eradicate this term completely. However, we expect to minimize this term as much as feasible.

When the linear preprocessor **W** is applied to the received signal vector **x**, we want the resulting signal to be close in the mean square error sense to some linear combination **G** of the transmitted signal vector **s**. We therefore try to find to minimize the cost function

$$J = E\left\{\left|\mathbf{w}_K^H \mathbf{y} - \mathrm{g}_K^H \mathbf{s}\right|^2\right\} \qquad 4.17$$

Where $\mathbf{w}_K^H$ is the k^{th} row of **W** and $\mathbf{g}_K$ is the effectual channel from the all transmitters toward the k^{th} receiver. Both $\mathbf{w}_K^H$ and $\mathbf{g}_K$ are design variables that are used to minimize the cost function of mean square error (MSE). It can be shown that this minimum cost function is achieved at

$$\mathbf{w}_K = \mathrm{R}_{\mathbf{yy}}^{-1}\mathrm{R}_{\mathbf{ys}}\mathbf{g}_K \qquad 4.18$$

Where $\mathrm{R}_{\mathbf{ys}} = E\{\mathbf{ys}^H\}$ and $\mathrm{R}_{\mathbf{yy}} = E\{\mathbf{yy}^H\}$. For this $\mathbf{w}_K$, the cost becomes

$$J = \mathrm{g}_K^H\left(\mathrm{R}_{\mathbf{ss}} - \mathrm{R}_{\mathbf{sy}}\mathrm{R}_{\mathbf{yy}}^{-1}\mathrm{R}_{\mathbf{ys}}\right)\mathbf{g}_K \qquad 4.19$$

The desired values of g_K are to be found by using the eigenvectors of the matrix

$$\mathrm{R}_{\mathbf{ss}} - \mathrm{R}_{\mathbf{sy}}\mathrm{R}_{\mathbf{yy}}^{-1}\mathrm{R}_{\mathbf{ys}} \qquad 4.20$$

Rows of the required preprocessor are to be obtained from the equation 4.18.

The MMSE criterion has been widely used in literature. To our knowledge, the application most similar to ours is in the channel shortening literature. More detail is available in [16] and references therein. In these references, the MMSE criterion is used to design channel shortening filters to reduce the delay spread of the channel. Therefore most of the multipath energy is restricted to a shorter delay spread. This is usually done to facilitate a smaller cyclic prefix in the discrete multitone, or is to allow a reduced state Viterbi decoder for enhanced data detection performance in single carrier communications. However, our work is different in the sense that here we use the MMSE criterion to design a linear preprocessor to suppress the channel uncertainty effects in contrast to reducing the span of the channel as in channel shortening.

Both approaches to dealing with channel uncertainty discussed above are local optimization algorithms. Iterative detection and channel estimation requires a good initialization to perform well but it incorporates the uncertainty term into the detection process. Applying a linear preprocessor will suppress the interference term due to the channel uncertainty to provide a better estimate but does not incorporate the interference term into the detection process. Both these local optimization algorithms complement one another. The linear preprocessor can be used to provide a good initialization to iterative detection and channel estimation while the iterative detection and channel estimation can be used to incorporate the uncertainty term into the data detection process that has been suppressed by the linear preprocessor. In short, we propose to concatenate these two local optimization algorithms to deal with the channel uncertainty.

4.6 Simulation Results

The simulated MIMO system has six transmit antennas($M_T = 6$) and six receive antennas ($M_R = 6$). The channel uncertainty is assumed to lie in a two dimensional ellipsoid that means the value of $\mathrm{M}_P = 2$. The channel uncertainty is modeled by the matrix $\mathbf{P} = \rho\widetilde{\mathbf{P}}$. In this equation $\widetilde{\mathbf{P}}$ has elements that are independent identically distributed (i.i.d), according to the unit normal distribution and the scaling factor ρ determines the amount of uncertainty. The actual channel is assumed to be uniformly distributed inside the uncertainty ellipsoid. By assuming the BPSK modulation scheme, the bit error rate plots (BER) are obtained. The bit error rate (BER) plots are given in FIG 4.2. This figure shows and compares the performance of two stages algorithm with the simple iterative detection and channel estimation over three iterations.

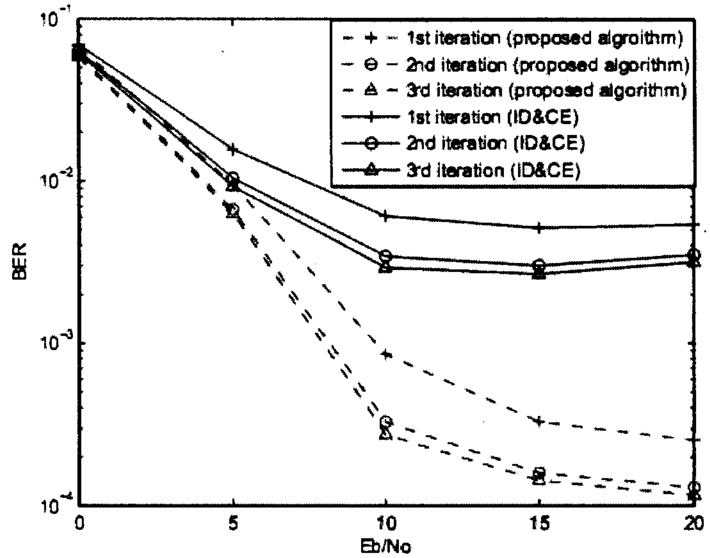

Figure 12: Performance of ID & CE E_b/N_o vs BER

The solid lines show the performance of iterative detection and channel estimation while dotted lines show the performance of iterative detection and channel estimation when the preprocessor is applied. In the first iteration the channel is assumed so the performance is poor. In the next iterations channel is estimated and reused for data detection and estimation. When the preprocessor is applied on the iterative detection and channel estimation iteratively, the performance is much better than using simple iterative detection and channel estimation.

Figure 4.3 plots the BER against the channel uncertainty ρ. Three iterations have been shown; the dotted plot shows the proposed model with preprocessing and solid plot shows the iterative detection and channel estimation without the preprocessing. It is observed that the performance of the scheme using a linear preprocessor is relatively stable to the effects of the channel uncertainty as compared to simply using iterative detection and channel estimation.

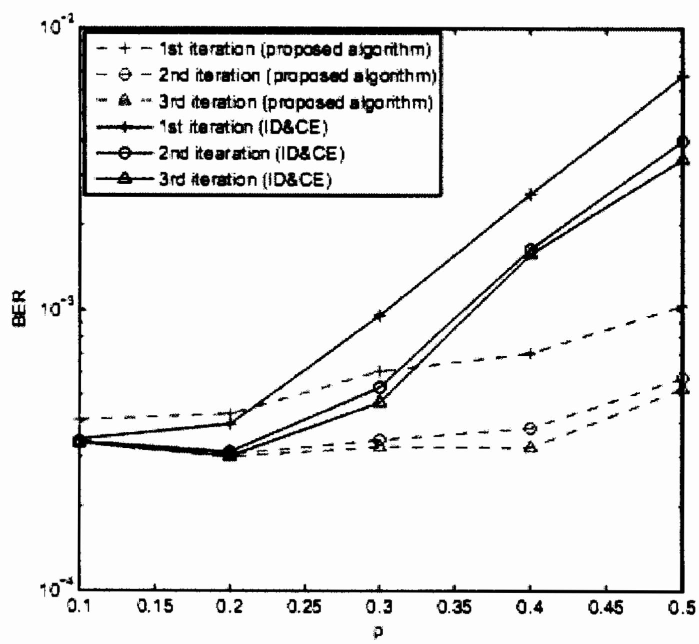

Figure 13: Performance in term of channel uncertainty

4.7 Conclusion

In summary, we have proposed a two stage algorithm by concatenating the two local optimization algorithms. Within a few iterations, the iterative detection and channel estimation algorithm converges to the local minimum. The linear preprocessor provides the iterative detection and channel estimation algorithm with a better initialization so that better overall performance is achieved. Simulations have verified the effectiveness of the approach and the performance gains are achieved without a very significant increase in computational complexity as the underlying idea is to apply a linear preprocessor which simply amounts to a matrix multiplication. We conclude with the note that the utility of the proposed approach is not just limited to a situation where the channel is uncertain, but can also be applied to a situation where the 'interference' term could be caused by say inter-cell interference from some co-channel cell.

References:-

[1] Kyung Seung Ahn Heath and R.W. Heung Ki Baik "Shannon Capacity and Symbol Error Rate of Space-Time Block Codes in MIMO Rayleigh Channels with Channel estimation Error", *wireless communication, IEEE Transaction on,* vol 7, pp 324-333, 2008.

[2] Di Renzo, Graziosi F and Santucci F, "Channel Capacity over Generalized Fading: A Novel MGF- Based Approach for Performance Analysis and Design of Wireless Communication Systems" *Vehicular Technology, IEEE Transactions on,* pp 127-149. 2009.

[3] Arogyaswami J.Pauleaj, Dhananjay A. Gore, Rohit U. Nabar and Helmut Bolcskei., "An Overview of MIMO Communication-A Key to Gigabit Wireless" *Proceeding of the IEEE,* pp 198-218, 2004.

[4] G.J.Foschini, "Layered space time architecture for wireless communication in fading environment when using multi-element antenna," *Bell labs Tech. J*, vol. 1, pp, 41-59, 1996.

[5] R. Gowaikar and B. Hassibi, "Statistical pruning for near–maximum likelihood decoding," *IEEE Trans. signal process*, vol. 55, no. 6, pp. 2661-2675,jun. 2007.

[6] D. Wubben, R. Bohnke, V. Kuhn and K. Kammeyer, "MMSE extension of V-BLAST based on shorted QR decomposition," *in proc, IEEE Vehicular Technology Conf (VTC)*, pp. 508-512, 2003.

[7] B. Hassibi and H. Vikalo. "On the expected complexity of sphere decoding," *in proc Asilomar Conf. signals. Systems and computers*, vol.2, pp, 1051-1055, 2001

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

[9] W. Zhao and G.B.Giannakis, "Sphere decoding algorithms with improved radius search", *IEEE Tans. Commun*, Vol. 53, pp 1104-1109, July. 2005.

[10] R. Gowaikar and B. Hassibi, "Statistical pruning for near maximum likelihood encoding," *IEEE Trans. Signal Process*, vol. 55, pp. 2661–2675, 2007.

[11] Mohamed Oussama Damen, Hesham El Gamal, and Giuseppe Caire, "On Maximum Likelihood Detection and the Search for Closest lattice Point", *IEEE Transaction on information theory*. vol. 49, no. X, 2003.

[12] Andrea Goldsmith, *Wireless Communication.* Stanford university, Cambridge University Press, 2005.

[13] G. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processes for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Area Commun.*, vol. 17, pp. 1841-1852, 1999.

[14] Omar Hayat, M.T.N. Qaisrani, and M. Akbar, "Sphere decoding in the presence of channel uncertainty," *IEEE ICET, FAST, 2010.*

[15] Albert van Zelst, *MIMO OFDM for wireless LANs*. PhD Theses.

[16] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans Signal Proces,* vol. 53, pp. 1474–1484, 2005.

[17] Luis G. Barbero and John S. Thompson, "Performance analysis of a fixed-complexity sphere decoder in high-dimensional MIMO systems" *Institute for Digital Communications University of Edinburgh.*

[18] L. G. Barbero and J. S. Thompson, "A fixed-complexity MIMO detector based on the complex sphere decoder," *submitted to IEEE Workshop on Signal Processing Advances for Wireless Communications (SPAWC '06).*

[19] B. Hochwald and S. Ten Brink, "Achieving near capacity on a multiple antenna channel," *IEEE Trans. On Communication,* vol. 51, pp. 389–399, 2003.

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

[21] B. Hassibi and H. Vikalo, "The expected complexity of sphere decoding part I: Theory," *IEEE Trans. Signal Process.,* vol. 53, pp. 2806–2818, 2005.

[22] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm. I. Expected complexity," *IEEE Trans .Signal Process,* vol. 53, no. 8, 2005.

[23] D. Daly, C. Hanegan, and A. D. Fagan, "Minimum mean-squared error impulse response shortening for discrete multi tone receivers," *IEEE Trans. Signal Process,* vol. 52, no. 1, pp. 301–306, 2004

[24] J. G. Proakis, *Digital Communications.* Englewood Cliffs, NJ: prentice Hall, 1995.

[25] H. Artes, D. Seethaler, and F. Hlawatsch, "Efficient detection algorithms for MIMO channels: a geometrical approach to approximate ML detection," *IEEE Trans. Signal Processing,* vol. 51, no. 11, pp. 2808-2820, 2003.

[26] J. Jalden, L. G. B. Ottersten, and J. S. Thompson, "The error probability of the fixed complexity sphere decoder," *in preparation. Analysis.* Boca Raton, FL: CRC Press, 2003

[27] D. Tse and P. Viswanath, 'Fundamentals of Wireless Communications', Cambridge University Press.

[28] A. Paulraj, D. Gore, R. Nabar, '*Introduction to Space time wireless communication'* Cambridge University Press.