Spring 2007

Fast sphere decoder for MIMO systems

Praveen G. Krishnan

Department:

FAST SPHERE DECODER FOR MIMO SYSTEMS

by

PRAVEEN G KRISHNAN

A THESIS

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI–ROLLA

in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

2007

Approved by

K. Kosbar, Advisor

S. Grant

G. K. Venayagamoorthy

ABSTRACT

This work focuses on variants of the conventional sphere decoding technique for Multi Input Multi Output (MIMO) systems. Space Time Block Codes (STBC) have emerged as a popular way of transmitting data over multiple antennas achieving the right balance between diversity and spatial multiplexing. The Maximum Likelihood (ML) technique is a conventional way of decoding the transmitted information from the received data, but at the cost of increased complexity. The sphere decoder algorithm is a sub-optimal decoding technique that is computationally efficient achieving a symbol error rate that is dependent on the initial radius of the sphere.

In this thesis, the decreasing rate of the radius of the sphere is increased by using a scaling factor of less than unity. This allows the algorithm to examine less number of vectors compared to the original algorithm making it much more computationally efficient. The sphere decoding algorithm is largely focussed on the Alamouti codes that have two antennas at the transmitter. This work extends the sphere decoding algorithm to other STBC having more than 2 transmit and receive antennas. The performance and the computational complexity of the fast sphere decoder is compared with that of the original sphere decoder and its variants.

ACKNOWLEDGMENTS

A masters program with thesis option can be a daunting task, as one looks to unravel the mysteries of research in a new and probing field. Many a time, like a bolt from the blue, one has to face this question, *Am I up to it?* It is then that you look up to some people for the right kind of guidance and positive approach. There have been numerous instances over the course of research when I was faced with puzzling and intriguing questions, and without any hesitation, I thank Dr. Kurt Kosbar, my academic advisor, from the bottom of my heart for getting me out of those tough situations. His constant advice and suggestions have helped me many a time to find solutions to the toughest of problems. I would like to thank my committee members, Dr. Steven L Grant and Dr. G. K. Venayagamoorthy for reviewing my thesis and offering helpful feedback and suggestions. I would like to express my thanks to Adam Panagos for guiding me in the right direction.

I take this opportunity to thank the most important people in my life, my parents, and my brother for having stood by me for all the decisions I have taken in my life. The masters degree program would be incomplete without their struggle and prayers. They are the undying source of inspiration in my pursuit of happiness. As an international student, it is not an easy task to leave your loved ones behind and travel miles beyond to follow your dream. It is, at this time, that I take extreme pleasure in thanking my roommates and friends Yadu, Rama and Venkatesh for putting up with me and supporting me during my lows and highs. It would have been impossible to have achieved all these things without their understanding and support. I would also like to express my thanks to Ashwin and Karthik Chandramouli for their constant encouragement and support. The Indian community here at Rolla is truly fabulous and I thank everyone for their support. Life at Rolla would have been totally boring without the many wonderful moments with some special friends. I would also like to thank the UMR faculty and non-faculty members, and Rolla community for their support.

I would also like to thank my friends Kavitha, Venkat and Prasad for their unflinching moral support during my stay here. I would like to thank my family friends Mr. Subramanian and Mr. Krishnamurthy for their absolute faith in me when they came forward with financial backing at a crucial juncture. It is certainly an unforgettable moment in my life. I would also like to thank my banker in India, Canara Bank, and the wonderful manager, Mr. Shivanna for easing the pressure off me and my family with an easy way out with the education loan.

All in all, I would like to thank everyone who have made it possible for me to climb up the ladder of the graduate degree program. I would like to apologize to anyone whose name I have missed out

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

1. INTRODUCTION

Multi Input Multi Output (MIMO) systems have replaced the conventional Single Input Single Output (SISO) systems in the last decade, and has emerged as a major area of research in the field of Wireless Communications for major applications. The multiple antennas at the transmitter and receiver can achieve a data rate that is much higher than that of the SISO system. The multiple antennas also aim to improve the performance of the system through various diversity techniques.

A MIMO system can be defined as a wireless communication system consisting of multiple antennas at the transmitter and multiple antennas at the receiver. The number of antennas at the transmitter and receiver are denoted by M and N respectively. For the case, when the transmit antennas equal the receive antennas, we denote the number by N.

The demand on higher data rates have always been of prime importance for wireless subscribers, as a result of which, new techniques in signal processing and coding need to be developed. There was a need to utilize the existing bandwidth to support a larger data rate with better quality that would combat the effects of multipath fading and Additive White Gaussian Noise (AWGN) prevalent in the channel. The quality of transmission can be improved by sending multiple copies of the transmitted signal over various paths to multiple receivers. This would ensure that, under the worst case circumstances, one received antenna would carry the desired signal. This technique is called as diversity. There are different types of diversity - time, frequency, space and polarization. Space diversity would ensure that the bandwidth utilization remains the same as that of SISO systems, while at the same time ensuring that the data rate and quality of transmission is increased [1].

Space-Time Block Codes (STBC) combine spatial and time diversity to increase the capacity of the system. Space-Time Block codes and Space-Time Trellis codes are examples of space time coding techniques. A general MIMO system that uses Space-Time Block Codes in the wireless communications system is illustrated in Figure 1.1. The multiple antennas at the transmitter transmit a certain number of symbols within a set of time slots. This results in the receiver getting more than one copy of the data after experiencing the various channel effects. The received data can be decoded using various techniques - Zero Forcing (ZF), Minimum Mean Square Error (MMSE) and Maximum Likelihood Detection (ML). The complexity of the decoder increases as the number of antennas at the transmitter and receiver increases.

The ML decoding technique is the optimal decoding technique achieving the best performance in terms of Symbol Error Rate (SER). But the complexity of the ML detector increases

exponentially as the number of antennas at the transmitter and receiver increases. Consider a case of N symbols transmitted in T time blocks. The number of points to be evaluated for a K sized constellation to decode both the symbols would increase exponentially, and is given by K^N. This is highly complex considering a large system of multiple antennas and a large sized constellation.

The computational complexity of the ML detection algorithm can be reduced by getting the channel state information (CSI) at the transmitter. The transmit precoding and receive shaping techniques would decompose the MIMO system into several parallel independent paths. As a result, the decoder search is reduced to K points, and this results in huge saving in computation compared to the K^N search for an unknown channel. The CSI cannot be obtained at the transmitter under all conditions, and therefore, a sub-optimal way of decoding the symbols is realised through Sphere Decoding. Sphere Decoders offer an alternative approach at a lower complexity to improve the efficiency of the decoding technique.

There are many variants of the Sphere Decoder, depending on the initial radius, and the way the search points are located in the algorithm. This thesis explores the basic sphere

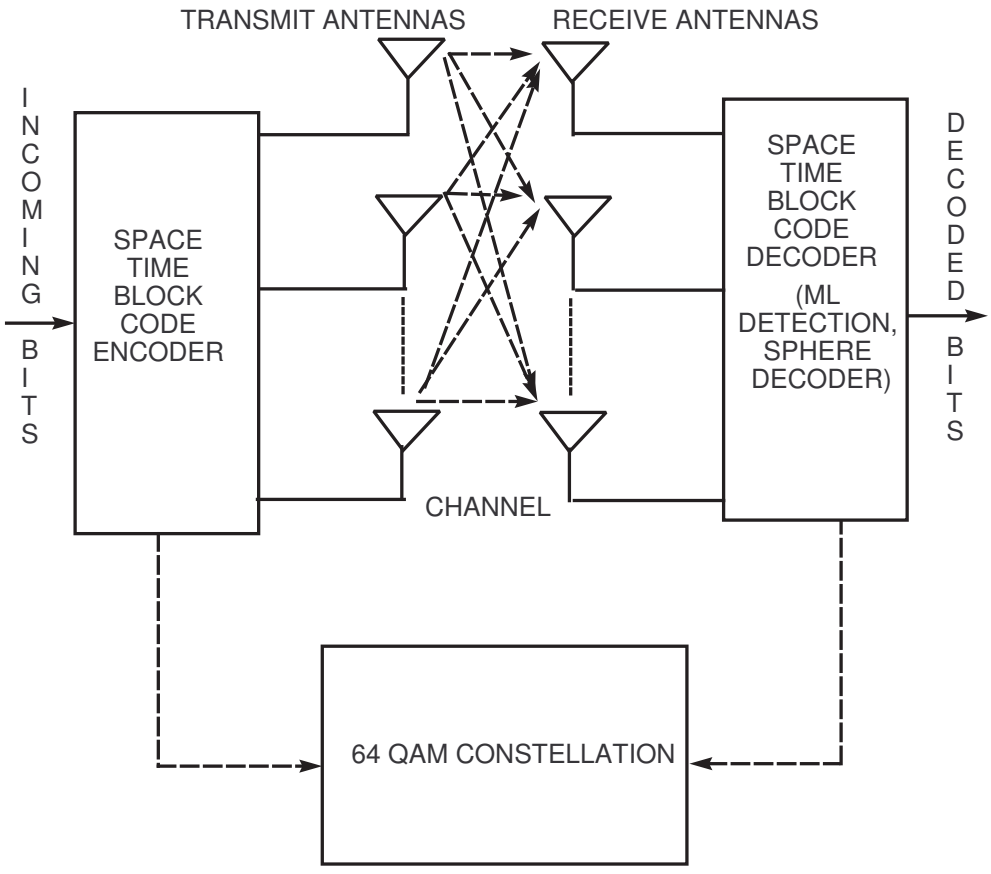

Figure 1.1 Diagram of a wireless MIMO communication system

decoding algorithm, and compares it to its variants in computational complexity and SER. The original sphere decoder, after the computation of the first point in the lattice, reduces the radius of the sphere to the value of the distance of this new point to the received point. The decreasing rate of the radius can be increased by multiplying with a scaling factor less than unity. This can be adjusted adaptively or by keeping a fixed scaling factor. A fixed scaling factor would reduce the complexity of the sphere decoder to a large extent at the expense of SER. This paper studies the effect of reducing the radius at a constant rate to examine complexity and performance. Sphere Decoders are clearly defined for Alamouti codes, and this paper explores the effect of sphere decoders on other space time block codes for multiple antennas.

In the following sections, the importance of STBC for MIMO systems along with the different decoding techniques - ML, sphere decoders and its variants is discussed. Simulations are carried out by transmitting a million STBCs over a flat fading channel, whose coefficients vary per block for different values of signal to noise ratio (SNR). The results are compared for SER and complexity. The complexity is determined by the number of floating point operations per second (FLOPS).

2. SPACE-TIME BLOCK CODES

2.1. INTRODUCTION

The space-time block codes provide a new dimension to the transmission of data using multiple transmit and receive antennas. Data is encoded using a space-time block code, and the encoded data is split into N streams which are simultaneously transmitted using N transmit antennas. The received signal at each antenna is a linear superposition of the N transmitted signals, perturbed by noise [1]. In most situations, the wireless channel suffers attenuation and path loss due to destructive addition of multipaths in the propagation media and to interference from other users. In a Rayleigh fading environment, it is often difficult for the receiver to determine the exact message transmitted due to severe attenuation and multipath fading characteristics, unless some less attenuated component of the signal is provided to the receiver. In many situations, since the wireless channel is neither significantly time-variant nor highly frequency selective, the possibility of deploying multiple antennas at the transmitter and receiver is considered to achieve spatial diversity.

2.2. SPACE-TIME BLOCK CODES GENERATION

The Space-Time Block codes are matrices that are designed to make use of transmit and receive diversity. The elements in the matrix are symbols taken from any particular constellation. The constellation may be as simple as a Binary Phase Shift Keying (BPSK)constellation with only real symbols, or the symbols may also be complex as can be seen in a Quadrature Amplitude Modulation (QAM) or Quadrature Phase Shift Keying (QPSK) constellations. The constellation decides the real or complex symbols transmitted as a part of the matrix.

2.3. THE SYSTEM MODEL

The Space-Time Block code [2] for a 2 transmit antenna system is given by the Alamouti [3] Code.

$$G_2 = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \tag{2.1}$$

The columns specify the number of antennas, while the rows specify the time slots. The time taken to transmit each entry in the matrix is T_{Symbol}, the symbol time. The time taken to transmit each row of the matrix is T_{Slot}, the slot time. The time taken to transmit each bit is T_{Bit}, the bit time. If K is the size of the constellation and M is the number of transmit

antennas, the bit time, symbol time and the slot time can be systematically represented by the following equations

$$T_{Bit} = T_{Symbol}/log_2(K) \tag{2.2}$$

$$T_{Slot} = T_{Symbol} \cdot M \tag{2.3}$$

$$T_{Block} = (Number\ of\ rows\ in\ the\ STBC) \cdot T_{Slot} \tag{2.4}$$

The waveform transmitting a given symbol can be represented by [4]

$$x_l(t) = \sum_{i=1}^{\infty} x_i g(t - iT_{Symbol}) \tag{2.5}$$

where x_i denotes the sequence of amplitudes obtained by mapping k-bit blocks of binary digits from the information sequence x_n into the amplitude levels ± 1, ± 3, $\pm(M-1)$, M = $2^{k/2} - 1$ depending on the type of QAM constellation selected, g(t) is the real pulse over interval $(0, T_{symbol})$.

The waveform transmitting the information in a given slot time can be represented by

$$x(t) = \sum_{j=1}^{\infty} \sum_{i=0}^{N_{slot}} d_j g(t - jT_{Slot} - iT_{Symbol}) \tag{2.6}$$

where, N_{slot} refers to the number of symbols transmitted in a given slot time, $d_j = [x_{j1}, x_{j2}]$, $x_{ji} \in \{x_i\}$ and x_i form the entries of the STBC matrices.

From the block code G_2, we notice that two symbols are transmitted in two time slots. So, the rate of this code is 1.

The representation of a band pass signal in the desired form is given by

$$s(t) = Re\{[x_l(t)]e^{j2\pi ft}\} \tag{2.7}$$

The signal $x_l(t)$ is the equivalent low pass signal and is the complex envelope of the real signal s(t) represented by a(t)+jb(t). a(t) and b(t) can be viewed as information bearing signal amplitudes of the quadrature carriers cos(2πft) and sin(2πft), respectively. The real valued signal s(t) has a frequency content centered in the vicinity of f.

$$s(t) = Re\{[a(t) + jb(t)]e^{j2\pi ft}\} \tag{2.8}$$

The signal amplitudes take a set of discrete values for a 4, 16 and 64 QAM signal constellation. The set of definite amplitude levels for a(t)+jb(t) constitute the complex signals x_1, x_2.... These complex values are the entries of the space time block code matrices.

The STBC transmitted depends on the number of transmit and receive antennas in the system. The STBC considered here are Orthogonal Space Time Block Codes that have the following unitary property

$$GG^H = \sum_{i=1}^{n_s} |x_n|^2 \cdot I = c \cdot |I| \tag{2.9}$$

The orthogonality has the following significance

- Achieves full diversity order provided by the transmit and receive antennas - these codes exploit the 2η dimensions, where $\eta = min(M, N)$.
- Simple ML decoding algorithms can be obtained by Linear processing at the receiver

2.4. THE TRANSMISSION MODEL

Consider a wireless communication system with M transmit antennas and N receive antennas, transmitting the symbols at different time slots. In each time slot, a codeword is transmitted, corresponding to the number of transmit antennas. The channel is assumed to be a flat fading channel and the path gain from transmit antenna i to the receive antenna j is defined to be $\alpha_{i,j}$. The path gains are modeled as samples of independent complex Gaussian random variables with a variance of 0.5 per real dimension [5]. The channel is assumed to be quasi-static over a number of time slots, or it is at least assumed to be static for time slots that correspond to transmission of one block code.

At time t, the signal r_t^i received at antenna j, is given by [6]

$$r_t^i = \sum_{i=1}^{n} \alpha_{i,j} c_t^i + \eta_t^j \tag{2.10}$$

where the noise samples η_t^j are independent samples of a zero-mean complex Gaussain random variable with variance $n/(2SNR)$ per complex dimension. The average energy of the symbols transmitted from each antenna is normalized to be one.

Assuming perfect channel state information is available, the receiver computes the decision metric

$$\sum_{t=1}^{l} \sum_{j=1}^{m} \left| r_t^j - \sum_{i=1}^{n} \alpha_{i,j} c_t^i \right|^2 \tag{2.11}$$

over all code words

$$c_1^1 c_1^2 \cdots c_1^n c_2^2 \cdots c_2^n \cdots c_l^1 c_2^l \cdots c_l^n$$

The codeword that minimizes the above sum is the code word that is favored for transmission. This is computationally intense since all possible combinations of the symbols are verified in the metric.

2.5. ALAMOUTI CODE

Alamouti space-time block codes are a special class of orthogonal block codes achieving a code rate of 1. Space-time block codes are represented by a $m \times l$ matrix, where m represents the number of transmit antennas and l represents the number of time slots required for transmission. Alamouti codes have a special significance in that the columns of the matrix are orthogonal achieving a code rate of 1. The code rate is determined by the number of symbols transmitted in a given number of time slots. For example, for a two transmit two receive antenna system, two symbols are transmitted in two time slots, thus getting a code rate of 1. The Alamouti code for a two transmit two receive antenna system is given by (2.12).

$$G_2 = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \tag{2.12}$$

The columns specify the number of transmitters, while the rows specify the time slots. The symbols are taken from a complex constellation, with each symbol represented by b bits. At time slot 1, kb bits arrive at the encoder and select constellation signals $s_1, \cdots, s_k$. The symbols obtained from mapping the bits to a particular constellation gives rise to a matrix C, which contains specific constellation symbols. The symbols are transmitted from M transmit antennas for each kb bits. If c_t^i represents the element in the tth row and ith column of C, the entries c_t^i, $i = 1, 2, \cdots, m$ are transmitted simultaneously from transmit antennas $1, 2, \cdots, m$ at each time slot $t = 1, 2, \cdots, p$. So, the ith column of C represents the symbol transmitted from the ith antenna and the tth row of C represents the transmitted symbols at time slot t. Exploiting the basic feature of C, the orthogonal columns of C provide a linear decoding scheme and give rise to a simplified ML decoding scheme. Since k symbols are transmitted in p time slots, the rate R of the code is defined to be $R = k/p$. Therefore, for the Alamouti code, the code rate is 1.

2.6. OTHER SPACE-TIME BLOCK CODES

The STBCs are realised with the goal of finding a tradeoff between diversity and multiplexing. There is the tendency to increase the data transmission rate by sending unique symbols at different time instants. This would save the redundancy of sending the same symbols, which in turn would increase the data rate. But, in a multipath environment, there is

a possibility that the channel cannot support the data rate decided by the Space Time Block Code. The multipath fading characteristics most often results in lossy transmission, which would lead to incorrect decoding at the receiver. In such cases, the multiplexing rate has to be compromised by reducing the code rate. It is absolutely essential to keep the matrices of the Space-Time Block Codes orthogonal in order to have remarkably simple decoding algorithms based on linear processing at the receiver. Apart from the most popular Alamouti code which has a code rate of 1, there are several other codes, which maintain the orthogonal structure of the code, but with lower code rates. The following are the half rate space-time block codes for a 3-Tx and a 4-Tx system.

$$G_3 = \begin{bmatrix} x_1 & x_2 & x_3 \\ -x_2 & x_1 & -x_4 \\ -x_3 & x_4 & x_1 \\ -x_4 & -x_3 & x_2 \\ x_1^* & x_2^* & x_3^* \\ -x_2^* & x_1^* & -x_4^* \\ -x_3^* & x_4^* & x_1^* \\ -x_4^* & -x_3^* & x_2^* \end{bmatrix} \tag{2.13}$$

In this case, four symbols are transmitted in eight time slots. The rate of this code is therefore 1/2.

$$G_4 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \\ x_1^* & x_2^* & x_3^* & x_4^* \\ -x_2^* & x_1^* & -x_4^* & x_3^* \\ -x_3^* & x_4^* & x_1^* & -x_2^* \\ -x_4^* & -x_3^* & x_2^* & x_1^* \end{bmatrix} \tag{2.14}$$

As in the previous case, four symbols are transmitted in eight time slots for this four transmitter system. So, the rate of this code is also 1/2.

The code rate can be increased by reducing the time slots in transmitting the same information. This can be clearly observed in the following codes which have a higher code rate.

$$H_3 = \begin{bmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & \frac{x_3}{\sqrt{2}} & \frac{-x_1-x_1^*+x_2-x_2^*}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & -\frac{x_3}{\sqrt{2}} & \frac{x_2+x_2^*+x_1-x_1^*}{\sqrt{2}} \end{bmatrix} \tag{2.15}$$

In this case, three symbols are transmitted in four time slots, which results in a rate of 3/4. For a four transmit antenna system, the STBC is given by

$$H_4 = \begin{bmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} & -\frac{x_3}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & \frac{x_3}{\sqrt{2}} & \frac{-x_1-x_1^*+x_2-x_2^*}{2} & \frac{-x_1-x_2^*+x_1-x_1^*}{2} \\ \frac{x_3}{\sqrt{2}} & -\frac{x_3}{\sqrt{2}} & \frac{x_2+x_2^*+x_1-x_1^*}{2} & -\frac{x_1+x_1^*+x_2-x_2^*}{2} \end{bmatrix} \tag{2.16}$$

It is not possible to design space-time block codes having a code rate of 1 for any given number of transmit antennas. Alamouti code is a special case for a 2 transmit antenna system that maintains the orthogonal structure with a code rate of unity. The research on space-time block codes have led to many results demonstrating the highest code rate possible for different number of transmit antenna systems. There was a challenge to come up with complex orthogonal space-time block codes with rate greater than 1/2 apart from the codes discussed above having a code rate of 3/4. Such a code is obtained for a 5 transmit antenna system and is as shown in the matrix below [7].

$$G_5 = \begin{bmatrix} x_1 & x_2^* & x_3^* & x_4^* & 0 \\ x_2 & -x_1^* & 0 & 0 & x_5^* \\ x_3 & 0 & -x_1^* & 0 & -x_6^* \\ 0 & x_3 & -x_2 & 0 & x_7 \\ x_4 & 0 & 0 & -x_1^* & x_8^* \\ 0 & -x_4 & 0 & x_2 & -x_9 \\ 0 & 0 & -x_4 & x_3 & x_{10} \\ x_5 & 0 & -x_7^* & -x_9^* & -x_2^* \\ 0 & x_5 & -x_6 & x_8 & x_1 \\ x_6 & -x_7^* & 0 & -x_{10}^* & x_3^* \\ x_7 & x_6^* & x_5^* & 0 & 0 \\ x_8 & x_9^* & -x_{10}^* & 0 & -x_4^* \\ x_9 & -x_8^* & 0 & x_5^* & 0 \\ x_{10} & 0 & x_8^* & x_6^* & 0 \\ 0 & -x_{10} & -x_9 & x_7 & 0 \end{bmatrix} \tag{2.17}$$

From the above matrix, it is seen that 10 symbols are transmitted in 15 time slots. This would give a code rate of 2/3.

The simulations are carried out by transmitting the STBCs $G_2 - G_5$ and decoding them using ML detection and sphere decoding. These techniques are discussed in the following sections.

3. MAXIMUM LIKELIHOOD DETECTION

3.1. INTRODUCTION

Maximum likelihood decoder achieves the best performance in terms of SER among all the decoding techniques. The computational complexity of this technique is very high, but with the advent of orthogonal space-time block codes, the exponential complexity is reduced to linear processing at the receiver. The space-time block codes along with multiple antennas achieve significant performance gain at no extra processing expense. In a maximum likelihood decoder, the complexity of the decoder increases exponentially as the number of transmit antennas. The significance of orthogonal space time block codes arises from the fact that the exponential complexity is reduced to a linear complexity.

The important aspect of the maximum likelihood detector is to estimate the channel at the receiver. The channel can be estimated using different estimation strategies. The different methods may be blind, semi-blind and assuming perfect channel state information is available, the receiver computes the decision metric [6]

$$\sum_{t=1}^{l}\sum_{j=1}^{m}\left|r_t^j-\sum_{i=1}^{n}\alpha_{i,j}c_t^i\right|^2 \tag{3.1}$$

over all code words

$$c_1^1c_1^2\cdots c_1^nc_2^2\cdots c_2^n\cdots c_l^1c_2^l\cdots c_l^n$$

The codeword that minimizes the above sum is the code word that is favored for transmission.

3.2. TWO ANTENNA TRANSMIT DIVERSITY SCHEME

A simple transmit diversity scheme was proposed by Alamouti that improves the quality of the signal and the simple processing is discussed below [3].

It is assumed that the channel is constant across two consecutive symbols. The channel co-efficients are picked from an i.i.d distribution. If α_1 and α_2 are the channel co-efficients from the two transmitters to the receiver, then the received signal at the receiver at two different time instants can be given by

$$r_1=\alpha_1c_1+\alpha_2c_2+\eta_1 \tag{3.2}$$

$$r_2=-\alpha_1c_2^*+\alpha_2c_1^*+\eta_2 \tag{3.3}$$

The codewords can be recovered from the received signal by the following equations

$$\widetilde{c_1} = \alpha_1^* r_1 + \alpha_2 r_2^* = (\alpha_1^2 + \alpha_2^2)c_1 + \alpha_1^* \eta_1 + \alpha_2 \eta_2^* \tag{3.4}$$

$$\widetilde{c_2} = \alpha_2^* r_1^* - \alpha_1 r_2^* = (\alpha_1^2 + \alpha_2^2)c_2 - \alpha_1 \eta_2^* + \alpha_2^* \eta_1 \tag{3.5}$$

The combiner sends the decoded symbols to the maximum likelihood detector, which minimises the following decision metric

$$|r_1 - \alpha_1 c_1 - \alpha_2 c_2|^2 + |r_2 + \alpha_1 c_2^* - \alpha_2 c_1^*|^2 \tag{3.6}$$

over all possible values of c_1 and c_2. Expanding the above equation and deleting terms that are independent of the codewords, the above minimization leads to separately minimizing

$$|\alpha_1^* r_1 + \alpha_2 r_2^* - c_1|^2 + (\alpha_1^2 + \alpha_2^2 - 1)|c_1^2| \tag{3.7}$$

for detecting c_1 and

$$|\alpha_2^* r_1 - \alpha_2 r_2^* - c_2|^2 + (\alpha_1^2 + \alpha_2^2 - 1)|c_2^2| \tag{3.8}$$

for decoding c_2.

3.3. TWO ANTENNA TRANSMIT RECEIVE DIVERSITY SYSTEM

The maximum likelihood decoder for a two antenna transmit receive diversity system is similar to the one derived in the previous section. The decision metric for such a system is given by

$$\sum_{j=1}^{m} (|r_1^j - \alpha_{1,j} s_1 - \alpha_{2,j} s_2|^2 + |r_2^j + \alpha_{1,j} s_2^* - \alpha_{2,j} s_1^*|^2) \tag{3.9}$$

over all possible values of s_1 and s_2, respectively. The quasi-static nature of the channel assumes that the path gains are constant over two transmissions. The above metric is expanded and those terms that are independent of the codewords are deleted. The final minimization metric reduces to the equivalent expression.

$$\begin{aligned} -\sum_{j=1}^{m} [&r_1^j \alpha_{1,j}^* s_1^* + (r_1^j)^* \alpha_{1,j} s_1 + r_1^j \alpha_{2,j}^* s_2^* + (r_1^j)^* \alpha_{2,j} s_2 \\ &- r_2^j \alpha_{1,j}^* s_2 - (r_2^j)^* \alpha_{1,j} s_2^* + r_2^j \alpha_{2,j}^* s_1 + (r_2^j)^* \alpha_{2,j} s_1^*] \\ &+ (|s_1|^2 + |s_2|^2) \sum_{j=1}^{m} \sum_{i=1}^{2} |\alpha_{i,j}|^2 \end{aligned} \tag{3.10}$$

The above metric decomposes into two parts, each of them being a function of only one variable. In this case, the first part is a function of s_1 and the other a function of s_2. The first part is given by

$$-\sum_{j=1}^{m}[r_1^j\alpha_{1,j}^* s_1^* + (r_1^j)^*\alpha_{1,j}s_1 + r_2^j\alpha_{2,j}^* s_1 + (r_2^j)^*\alpha_{2,j}s_1^*] \\ +(|s_1|^2)\sum_{j=1}^{m}\sum_{i=1}^{2}|\alpha_{i,j}|^2 \tag{3.11}$$

which is a function of s_1 only and is independent of s_2.

The second part of the expansion is given by

$$-\sum_{j=1}^{m}[r_2^j\alpha_{2,j}^* s_2^* + (r_1^j)^*\alpha_{2,j}s_2 - r_2^j\alpha_{1,j}^* s_2 + (r_2^j)^*\alpha_{1,j}s_2^*] \\ +(|s_2|^2)\sum_{j=1}^{m}\sum_{i=1}^{2}|\alpha_{i,j}|^2 \tag{3.12}$$

which is a function of s_2 only and is independent of s_1.

The minimization of (3.9) is equivalent to minimizing these two parts separately. This is turn is equivalent to minimizing the decision metric

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{1,j}^* + (r_2^j)^*\alpha_{2,j})\right] - s_1\right|^2 \\ +\left(-1+\sum_{j=1}^{m}\sum_{i=1}^{2}|\alpha_{i,j}|^2\right)|s_1|^2 \tag{3.13}$$

for detecting s_1 and the decision metric

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{2,j}^* - (r_2^j)^*\alpha_{1,j})\right] - s_2\right|^2 \\ +\left(-1+\sum_{j=1}^{m}\sum_{i=1}^{2}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.14}$$

for detecting s_2.

From (3.13) and (3.14), it can be clearly seen that the symbols do not have to be obtained from exponential processing of the data. Simple linear processing is sufficient to obtain the symbols. The estimation of symbol s_1 is done by trying out all possible combinations for s_1. If s_1 had to be determined based on the values of multiple symbols in the equation, then the computation would vary exponentially. If there were two symbols in the equation, then

the estimation of s_1 would have had to be dependent on $(Constellationsize)^N$ number of iterations, where N is the number of dependent symbols. It can be clearly seen that there is no performance sacrifice in using it.

3.4. MAXIMUM LIKELIHOOD DECODING - OTHER STBC

The decoders for G_3, G_4, H_3 and H_4 can be similarly derived. The specific formulae for decoding G_3 is as given below. The decision metric to decode s_1 is given by

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{1,j}^* + r_2^j\alpha_{2,j}^* + r_3^j\alpha_{3,j}^* + (r_5^j)^*\alpha_{1,j} + (r_6^j)^*\alpha_{2,j} + (r_7^j)^*\alpha_{3,j})\right] - s_1\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_1|^2 \tag{3.15}$$

The decision metric to decode s_2 is given by

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{2,j}^* - r_2^j\alpha_{1,j}^* + r_4^j\alpha_{3,j}^* + (r_5^j)^*\alpha_{2,j} - (r_6^j)^*\alpha_{1,j} + (r_8^j)^*\alpha_{3,j})\right] - s_2\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.16}$$

The decision metric to decode s_3 is given by

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{3,j}^* - r_3^j\alpha_{1,j}^* - r_4^j\alpha_{2,j}^* + (r_5^j)^*\alpha_{3,j} - (r_7^j)^*\alpha_{1,j} - (r_8^j)^*\alpha_{2,j})\right] - s_3\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_3|^2 \tag{3.17}$$

The decision metric to decode s_4 is given by

$$\left|\left[\sum_{j=1}^{m}(-r_2^j\alpha_{3,j}^* + r_3^j\alpha_{2,j}^* - r_4^j\alpha_{1,j}^* - (r_6^j)^*\alpha_{3,j} + (r_7^j)^*\alpha_{2,j} - (r_8^j)^*\alpha_{1,j})\right] - s_4\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_4|^2 \tag{3.18}$$

For decoding G_4, the decoder minimizing the decision metric is given by the following expression for decoding s_1

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{1,j}^* + r_2^j\alpha_{2,j}^* + r_3^j\alpha_{3,j}^* + r_4^j\alpha_{4,j}^* + (r_5^j)^*\alpha_{1,j} + (r_6^j)^*\alpha_{2,j} + (r_7^j)^*\alpha_{3,j} + (r_8^j)^*\alpha_{4,j})\right] - s_1\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_1|^2 \tag{3.19}$$

For decoding s_2, the decision metric is given by

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{2,j}^* - r_2^j\alpha_{1,j}^* - r_3^j\alpha_{4,j}^* + r_4^j\alpha_{3,j}^* + (r_5^j)^*\alpha_{2,j} - (r_6^j)^*\alpha_{1,j} - (r_7^j)^*\alpha_{4,j} + (r_8^j)^*\alpha_{3,j})\right] - s_2\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.20}$$

For decoding s_3, the decision metric is given by

$$\left|\left[\sum_{j=1}^{m}(r_1^j\alpha_{3,j}^* + r_2^j\alpha_{4,j}^* - r_3^j\alpha_{1,j}^* - r_4^j\alpha_{2,j}^* + (r_5^j)^*\alpha_{3,j} + (r_6^j)^*\alpha_{4,j} - (r_7^j)^*\alpha_{1,j} - (r_8^j)^*\alpha_{2,j})\right] - s_3\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_3|^2 \tag{3.21}$$

For decoding s_4, the decision metric is given by

$$\left|\left[\sum_{j=1}^{m}(-r_1^j\alpha_{4,j}^* - r_2^j\alpha_{3,j}^* + r_3^j\alpha_{2,j}^* - r_4^j\alpha_{1,j}^* - (r_5^j)^*\alpha_{4,j} - (r_6^j)^*\alpha_{3,j} + (r_7^j)^*\alpha_{3,j} + (r_8^j)^*\alpha_{4,j})\right] - s_1\right|^2 + \left(-1 + 2\sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_1|^2 \tag{3.22}$$

The rate 3/4 code H_3 is decoded in a similar way by using the following decision metric for s_1

$$\left|\left[\sum_{j=1}^{m}\left(r_1^j\alpha_{1,j}^* - (r_2^j)^*\alpha_{2,j} + \frac{(r_4^j - r_3^j)\alpha_{3,j}^*}{2} - \frac{(r_3^j + r_4^j)^*\alpha_{3,j}}{2}\right)\right] - s_1\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.23}$$

For decoding s_2, the decision metric is given by

$$\left|\left[\sum_{j=1}^{m}\left(r_1^j\alpha_{2,j}^* - (r_2^j)^*\alpha_{1,j} + \frac{(r_4^j + r_3^j)\alpha_{3,j}^*}{2} + \frac{(-r_3^j + r_4^j)^*\alpha_{3,j}}{2}\right)\right] - s_2\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.24}$$

For decoding s_3, the decision metric is given by

$$\left|\left[\sum_{j=1}^{m}\left(\frac{(r_1^j + r_2^j)\alpha_{3,j}^*}{\sqrt{2}} + \frac{(r_3^j)^*(\alpha_{1,j} + \alpha_{2,j})}{\sqrt{2}} + \frac{(r_4^j)^*(\alpha_{1,j} - \alpha_{2,j})}{\sqrt{2}}\right)\right] - s_3\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{3}|\alpha_{i,j}|^2\right)|s_3|^2 \tag{3.25}$$

Similarly, to decode the rate 3/4 code H_4, the decision metric to estimate s_1 is given by

$$\left|\left[\sum_{j=1}^{m}\left(r_1^j\alpha_{1,j}^* + (r_2^j)^*\alpha_{2,j} + \frac{(r_4^j - r_3^j)(\alpha_{3,j}^* - \alpha_{4,j}^*)}{2} - \frac{(r_3^j + r_4^j)^*(\alpha_{3,j} + \alpha_{4,j})}{2}\right)\right] - s_1\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_1|^2 \tag{3.26}$$

The decision metric to decode s_2 is given by

$$\left|\left[\sum_{j=1}^{m}\left(r_1^j\alpha_{2,j}^* - (r_2^j)^*\alpha_{1,j} + \frac{(r_4^j + r_3^j)(\alpha_{3,j}^* - \alpha_{4,j}^*)}{2} + \frac{(-r_3^j + r_4^j)^*(\alpha_{3,j} + \alpha_{4,j})}{2}\right)\right] - s_2\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_2|^2 \tag{3.27}$$

The decision metric to decode s_3 is given by

$$\left|\left[\sum_{j=1}^{m}\left(\frac{(r_1^j + r_2^j)\alpha_{3,j}^*}{\sqrt{2}} + \frac{(r_1^j - r_2^j)\alpha_{4,j}^*}{\sqrt{2}} + \frac{(r_3^j)^*(\alpha_{1,j} + \alpha_{2,j})}{\sqrt{2}} + \frac{(r_4^j)^*(\alpha_{1,j} - \alpha_{2,j})}{\sqrt{2}}\right)\right] - s_3\right|^2 + \left(-1 + \sum_{j=1}^{m}\sum_{i=1}^{4}|\alpha_{i,j}|^2\right)|s_3|^2 \tag{3.28}$$

4. THE SPHERE DECODER

4.1. INTRODUCTION

ML decoder uses an exhaustive search algorithm where all possible codewords are checked, and the one with minimum distance is selected as the final codeword. The search algorithm is carried out by comparing the received vector with all possible codewords over an additive white Gaussian noise channel. The ML decoding technique is certainly an optimum decoding technique but with the increase in the number of transmit and receive antennas along with a complex constellation adds to the computational complexity of the decode mechanism.

Sphere Decoding is a new type of decoding technique which aims to reduce the computational complexity of the decoding technique. In case of a sphere decoder, the received signal is compared to the closest lattice point, since each codeword is represented by a lattice point [8]. The number of lattice points scanned in a sphere decoder depends on the initial radius of the sphere. The correctness of the codeword is in turn dependent on the SNR of the system. The search can easily be restricted by drawing a circle around the received signal in such a way so as to encompass a fixed number of lattice points. This entails a search of not all the codewords in the constellation and allows only those codewords to be checked that happen to fall within the sphere. All codewords outside the sphere are not taken into consideration for decoding.

Considerable research has gone into sphere decoding in the last decade. This has resulted in the emergence of quite a few sphere decoders with various variants to facilitate the decoding process. The conventional sphere decoders have been replaced by sphere decoders where the search proceeds independent of the initial radius. There are also list sphere decoders where more than one solution can be found. The size of the list can be as long as the constellation size, in which case the complexity is the same as the ML decoding technique or it can be of much lesser size, in which case the number of points scanned would reduce. The sphere decoder is investigated for each case and SER for different types of space time block codes is determined and its complexity is compared with the conventional ML decoder.

4.2. ORIGINAL SPHERE DECODER

In order to make use of sphere decoding, a lattice representation of the multiple antenna system is obtained. Each Space Time Block Code transmits a vector of symbols, and these symbols are aggregated to obtain a lattice point. When dealing with lattice codes for the fading channel, there is the problem of decoding a totally arbitrary lattice given its generator matrix. In the case of independent fading channels, with perfect CSI given to the receiver,

the ML decoding of the received vector has to satisfy the performance metric given by the following equation

$$min\|\mathbf{r}-\mathbf{x}\| = min\|\mathbf{w}\| \tag{4.1}$$

where $\boldsymbol{r}$ is the received vector. $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is one of the transmitted lattice code points. The lattice points can be written as the set $\mathbf{x} = \mathbf{uM}$, where M is the lattice generator matrix corresponding to the basis $(\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n})$ and $\mathbf{u} = (u_1, u_2, \cdots, u_n)$ is the integer component vector to which the information bits are easily mapped. The noise vector $\mathbf{n} = (n_1, n_2, \cdots, n_n)$ has real, Gaussian distributed independent random variable components, with zero mean and unit variance.

The lattice decoding algorithm [9] searches the lattice through a given set of points that is bounded by the initial radius of the sphere with the received point as center. The complexity of the algorithm is controlled by the initial radius of the sphere. The received vector and the space time block code are translated to lattice points, and the difference is represented by an integer component vector ξ of the cubic lattice. The flowchart in Figure 4.1 discusses the necessary steps in sphere decoding. We write $\mathbf{x} = \mathbf{uM}$, $\mathbf{r} = \rho M$ with $\rho = (\rho_1, \rho_2, \cdots, \rho_n)$, and $\mathbf{w} = \xi\mathbf{M}$ with $\xi = (\xi_1, \xi_2, \cdots, \xi_n)$. ρ and ξ are real vectors. Then, $\mathbf{w} = \sum_{i=1}^{n} \xi_i \mathbf{v_i}$, where $\xi_i = \rho_i - u_i, i = 1, \cdots, n$ define the translated coordinate axes in the space of the integer component vectors.

The Cholesky factorization of the Gram matrix $G = MM^T$ yields $G = R^T R$, where R is an upper triangular matrix. Then,

$$Q(\xi) = \sum_{i=1}^{n} q_{ii} \left(\xi_i + \sum_{j=i+1}^{n} q_{ii}\xi_j \right)^2 \leq C \tag{4.2}$$

The main function of the sphere decoder is to determine the vector by finding the equations of the border of the ellipsoid. The upper and lower bounds of the components u_n and u_{n-1} are

$$\left\lceil -\sqrt{\frac{C}{q_{nn}}} + \rho_n \right\rceil \leq u_n \leq \left\lfloor \sqrt{\frac{C}{q_{nn}}} + \rho_n \right\rfloor \tag{4.3}$$

Similarly, for the $n-1$th term, the equation is given by

$$\left\lceil -\sqrt{\frac{C - q_{nn}\xi_n^2}{q_{n-1,n-1}}} + \rho_{n-1} + q_{n-1,n}\xi_n \right\rceil \leq u_{n-1} \leq \left\lfloor \sqrt{\frac{C - q_{nn}\xi_n^2}{q_{n-1,n-1}}} + \rho_{n-1} + q_{n-1,n}\xi_n \right\rfloor \tag{4.4}$$

This equation can be extended to determine the intermediate points of the lattice. The ith integer component is determined using the same set of equations. The bounds are thus obtained iteratively from one integer component to the other. The following set of equations

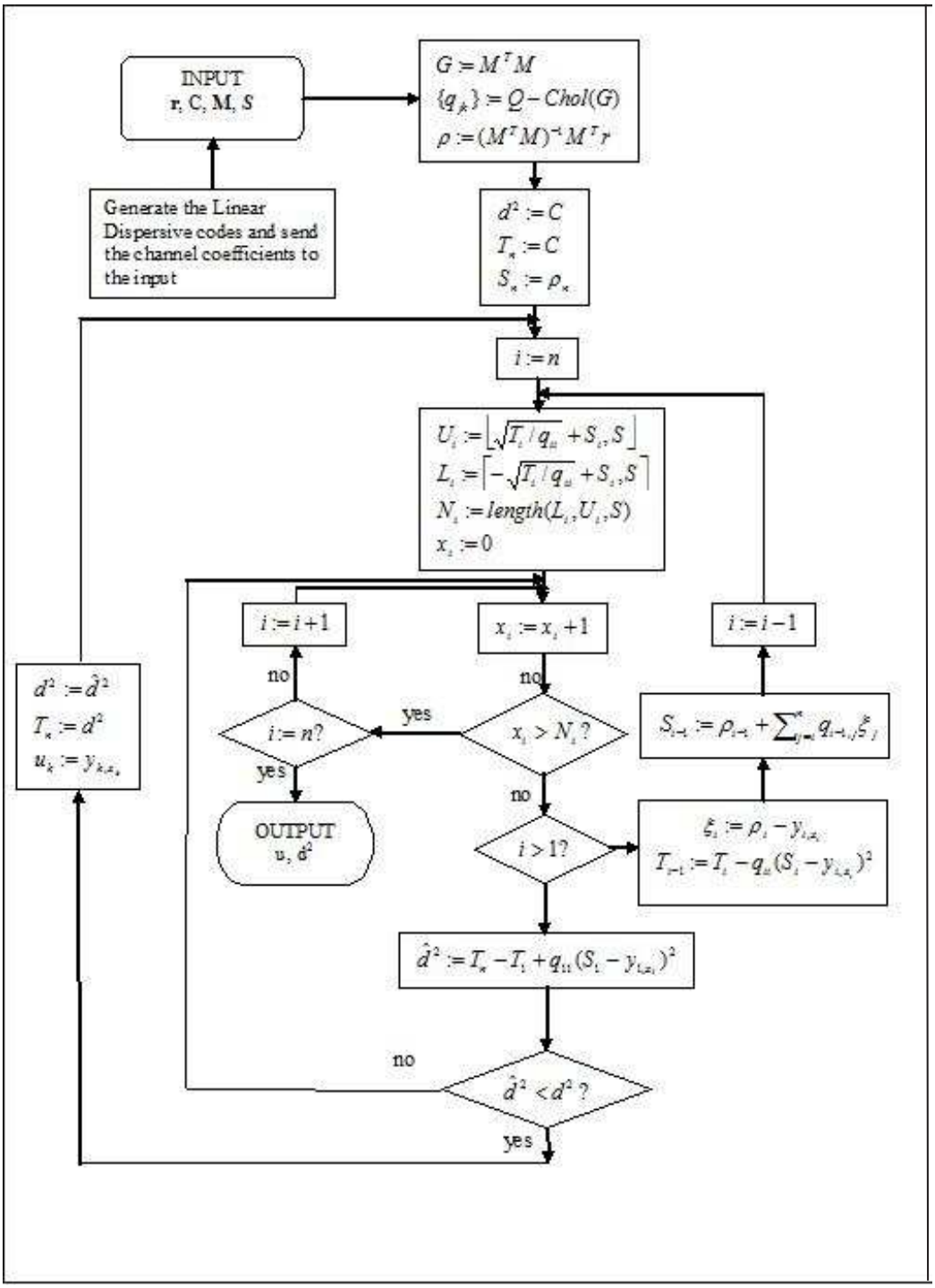

Figure 4.1 Flowchart of the original sphere decoder

are used to obtain the bounds

$$S_i = S_i(\xi_{i+1}, \cdots \xi_n) = \rho_i + \sum_{l=i+1}^{n} q_{il}\xi_l \tag{4.5}$$

As the ith point in the lattice is determined, the location of the $(i-1)$th point in the system is governed by the equation

$$T_{i-1} = T_{i-1}(\xi_i, \cdots \xi_n) = C - \sum_{l=i}^{n} q_{ll} \left(\xi_l + \sum_{j=l+1}^{n} q_{lj}\xi_j \right)^2 = T_i - q_{ii}(S_i - u_i)^2 \tag{4.6}$$

When a vector is found inside the sphere, the radius is computed accordingly with respect to the newly found point in the lattice. The updated radius is used for further calculations to check for remaining vectors in the lattice space.

The radius must be chosen in such a way that the value selected must equal the covering radius of the lattice. The initial radius selected plays a critical role in identifying the correct point in the lattice. Ideally, the noise variance of the system is found and the initial radius of the sphere is adjusted according to the Signal to Noise Ratio. This entails the sphere decoder to find at least a single point inside the sphere and prevents the condition of erasure, which is a result of decoding failure. In case of an erasure, the radius of the sphere is increased and the sphere decoding is again repeated to obtain favorable results.

The flowchart of the conventional sphere decoder is slightly modified by obtaining the zero-forcing solution, instead of just computing the pseudo-inverse of the channel. The sphere decoder is then applied to the received vector that is obtained from the zero-forcing decoder to determine the accurate solution.

4.3. MODIFIED SPHERE DECODING ALGORITHM

The proposed sphere decoding algorithm computes the actual point from the surface of the sphere towards the center as it searches for the optimum solution. This is inefficient since the ML solution is associated with the lattice point closest to the surface of the sphere, and there is an additional complexity in starting the search from the surface of the sphere.

The modified sphere decoder takes care of this problem by examining the points close to the received point at the center of the sphere. As a new point is reached, the search is restricted by computing the radius of the sphere to be equal to the distance of the newly discovered lattice point from the sphere center.

The flowchart representing the modified sphere decoder highlights the differences as compared to the original sphere decoder. As the lower and upper bounds are computed, in order to determine the ith coordinate, the components of $\mathbf{y_i}$ are sorted in ascending order according to the metric

$$|y_{ij} - S_i|^2 \tag{4.7}$$

for $1 \leq j \leq N_i$, with the output stored in the row vector z_i. Whenever a valid lattice point is found, the radius is updated and the lower and upper bounds are again recomputed. The algorithm is explained in detail in the [8]. The computational savings with the modified sphere decoder is a lot as compared to the original sphere decoder. This is discussed in detail with reference to the simulation results.

4.4. FAST SPHERE DECODER USING AN ADAPTIVE RADIUS

A sphere decoder is proposed in which the initial radius of the sphere is not important. In the original sphere decoder, whenever a point is found inside the sphere, the radius of search sphere reduces to the value of the distance of this new point to the received point. In the proposed method, this radius decreases at a faster rate. The rate of decrease of the radius is governed by equation (4.8) [10]

$$d^2 = k * \hat{d}^2 \tag{4.8}$$

where d is the new radius and $\hat{d}^2$ is the squared distance of the previous founded point to the center and the scale parameter k is bounded by $0 \leq k \leq 1$

The question that arises is the way the scale parameter k has to be chosen. The radius is inversely proportional to the value of k. The algorithm speeds up as the value of k gets smaller, but keeping the value too small may cause some problems as the algorithm approaches the received point. So, the value of k has to be changed adaptively. The possible values for a sphere in the mth dimension determines the points in the $m+1$th dimension, and it is the number of these possible values, t, that is used for calculating the value of k. In general, if this number is large, then the number of points inside the radius C will be large and the coefficient k tends to the lower side. On the other hand, if the sphere decoder has to search less number of points, then the value of k tends to one. The function that gets these desired properties is given by [10]

$$k(t) = exp\left[-(\frac{1}{\alpha}t)^{\beta}\right] \tag{4.9}$$

where t is the number of points in the nth dimension of lattice and α, β are the parameters that control this function.

4.5. FAST SPHERE DECODER - CONSTANT SCALING FACTOR

The algorithm of the original sphere decoder is slightly modified to include a scaling factor less than unity as shown in Figure 4.2. The original sphere decoder changes the radius of the sphere once the first point inside the sphere is determined. There is a scaling factor that is associated with the radius that has to be determined. The scaling factor k is fixed at one for the original sphere decoder, but this radius can be reduced by allowing it to decrease

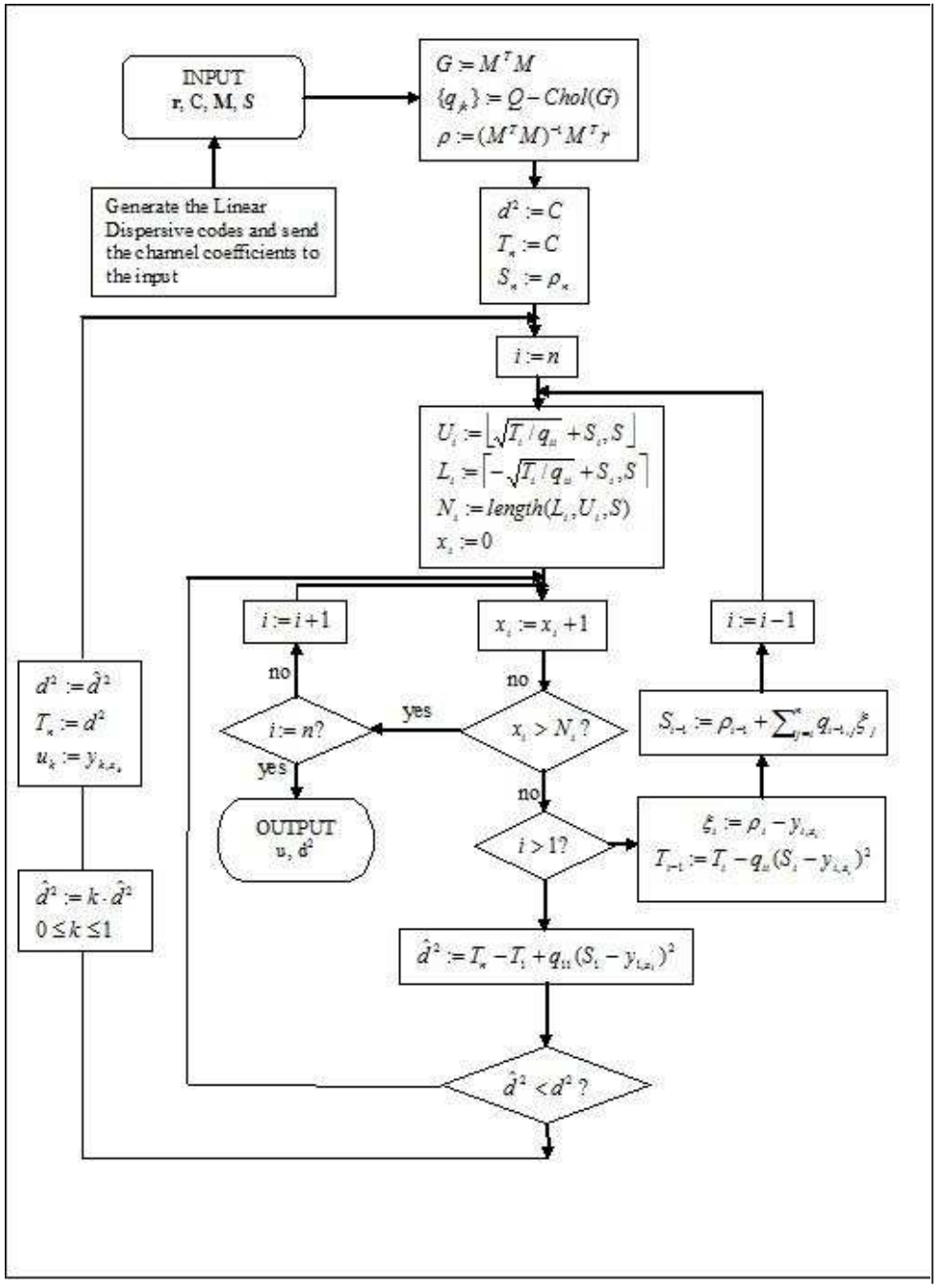

Figure 4.2 Flowchart of the sphere decoder with a scaling factor less than unity

at a faster rate. This would have the advantage of reducing the number of computations as compared to the modified sphere decoder and the fast sphere decoder when the initial radius of the sphere is kept low. At high initial radius, the modified sphere decoder has the advantage over other algorithms, since it is not entirely dependent on the initial radius of the sphere. The savings in computational complexity due to a constantly varying radius using a scaling factor even slightly lesser than unity is much more than the original sphere decoder. The

performance of such a system over an adaptively varying radius is slightly inferior in terms of the Symbol Error Rate, but the computational savings outweighs small changes in SER.

The performance of the fast sphere decoder is checked for different values of the scaling factor. At very low values of the scaling factor, there is bound to be significant savings in the computational complexity, but at the cost of high SER. The computational complexity is directly proportional to the scaling factor, and as k tends to 1, the performance of the fast sphere decoder approaches that of the original sphere decoder in terms of the symbol error rate as well as complexity. So, an optimum value of k must be chosen in such a way that there is not a large compromise with respect to the symbol error rate at a reduced complexity.

The radius can be altered as per the following equation

$$d^2 = k * \hat{d}^2 \tag{4.10}$$

where d is the new radius and $\hat{d}^2$ is the squared distance of the previous founded point to the center and the scale parameter k is bounded by $0 \leq k \leq 1$.

As compared to the previous method, k is set a value between 0 and 1. This value is not determined adaptively, and thus significant savings in computation is achieved. A low value of k is chosen when the SNR of the system is low, and a high value of k is chosen when the SNR of the system is high. This is because the received vector is translated to a point in the lattice which is the transmitted vector at high values of the SNR. A good performance can be achieved using this sphere decoder if there is an idea of the SNR of the system. Thus, depending on the SNR, a low or high value of k is set to obtain good performance in terms of SER and complexity.

4.6. LIST SPHERE DECODER

The list sphere decoder is a generalization of the sphere decoder [11]. The sphere decoder finds a single vector that sub-optimally solves the Maximum Likelihood detection problem. In case of a list sphere decoder, it finds the L best solutions depending on the value taken for the list size. It stores each of these vectors u_k, along with their corresponding radius C_k in a list $L_k = u_k, C_k$ for $k = 1, \cdots, L$. The list sphere decoder is obtained by modifying the sphere decoder algorithm. Each time, it finds a point in the sphere, the list sphere decoder checks to see if it is better than any of the other arguments in the list, and if so, exchanges them.

The list sphere decoder requires an order of L comparisons. In effect, the selection sort algorithm is directly dependent on the number of entries in the list. The complexity of the algorithm increases as the list size increases. The list size is fixed in such a way that $2^{M_c.M} > N_{cand} \geq 1$, where M_c is the bits per symbol, M is the number of transmit antennas

and N_{cand} is the candidate set of all suitable points that maximizes the minimum distance metric. The sphere decoder algorithm is modified to a list sphere decoder using the following steps. Every time, the search finds a point inside a sphere of radius r [12],

- the radius does not decrease to correspond to the new radius r
- if the list is not full, this point is updated to the list; or if the list is full, the list is updated with the new point if the new point has a smaller radius than the largest radius in the list.

Hence, L contains the maximum likelihood estimate and N_{cand-1} neighboring points for which the minimum metric is satisfied. A large r generally allows for large N_{cand}, which amounts to large number of points being evaluated, yielding a reliable list. There is also a tradeoff between the accuracy and the speed of the sphere decoder. It is computationally intensive to find N_{cand} points than to find a single point since the radius does not decrease adaptively with every point that is found. The number of points N_{cand} in the list depends on the initial radius of the sphere. If the initial radius of the sphere is kept very small, no matter how large N_{cand} is, it is insufficient to determine such a large number. On the other hand, if the initial radius of the sphere is very large, it would take a large number of iterations to find the best suitable N_{cand} number of points. A large radius would actually search through many candidate points before finding the best N_{cand} number of points.

5. SIMULATION RESULTS

The space time block codes are decoded using Maximum Likelihood technique for different set of codes. The original sphere decoder is examined for different values of the radius to determine the Symbol Error Rate. The number of floating point operations is determined for each type of space time block code for different values of the radius.

5.1. ALAMOUTI CODE

The simulation is carried out for G_2 by keeping the radius as $\sqrt{11}$, and the graph in Figure 5.1 denotes the Symbol Error Rate for about a million symbols. It can be clearly

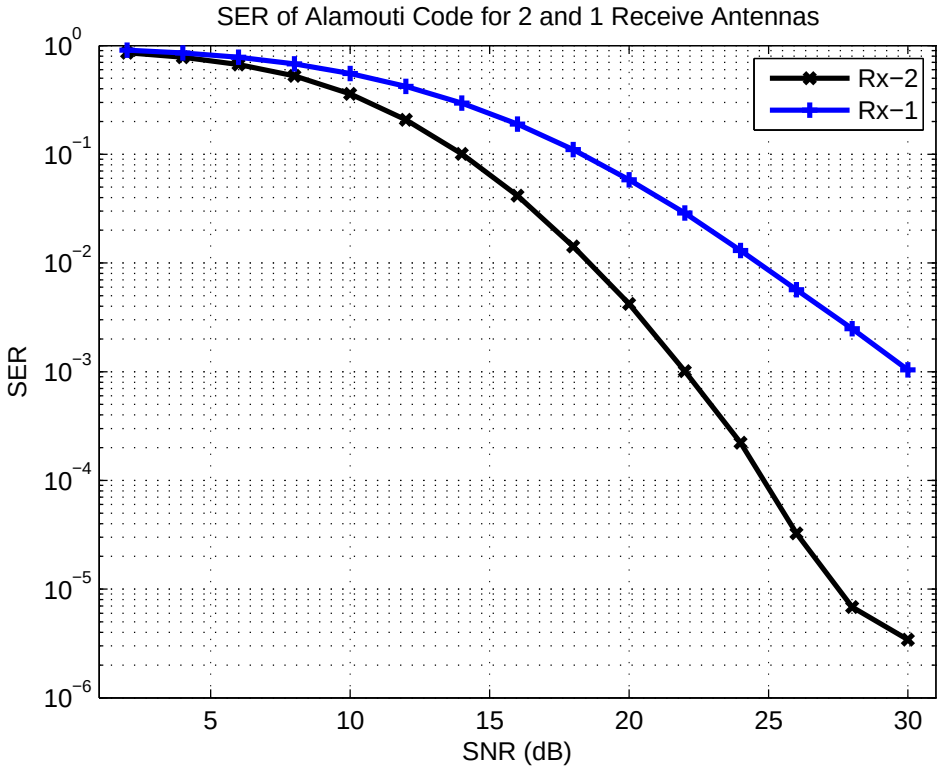

Figure 5.1 SER of Alamouti Code for 2 and 1 receive antennas - k=1,r=$\sqrt{11}$

noticed from the figure that the SER improves with the increase in number of receive antennas. In both curves, the number of transmit antennas is kept constant, and only the number of receive antennas is changed. The number of receive antennas is clearly a deciding factor in determining the performance of the MIMO system.

It can be seen from Figure 5.2 that as the initial radius of the sphere increases, the SER improves drastically. At a radius of 1, 7dB of extra SNR is required to achieve the same SER as obtained with a radius of $\sqrt{11}$. As the scaling factor is reduced from 1 for a radius of $\sqrt{11}$, the performance of the sphere decoder degrades as can be noticed from the curves for a scaling factor of 0.3 and 0.5 respectively. It can also be observed that for a radius of 10, there is no significant improvement in SER as what is obtained for a radius of $\sqrt{11}$. This is because, a radius of $\sqrt{11}$ encompasses all the points in the lattice, and so an increase in radius does not increase the number of lattice points examined. The complexity analysis for these curves can

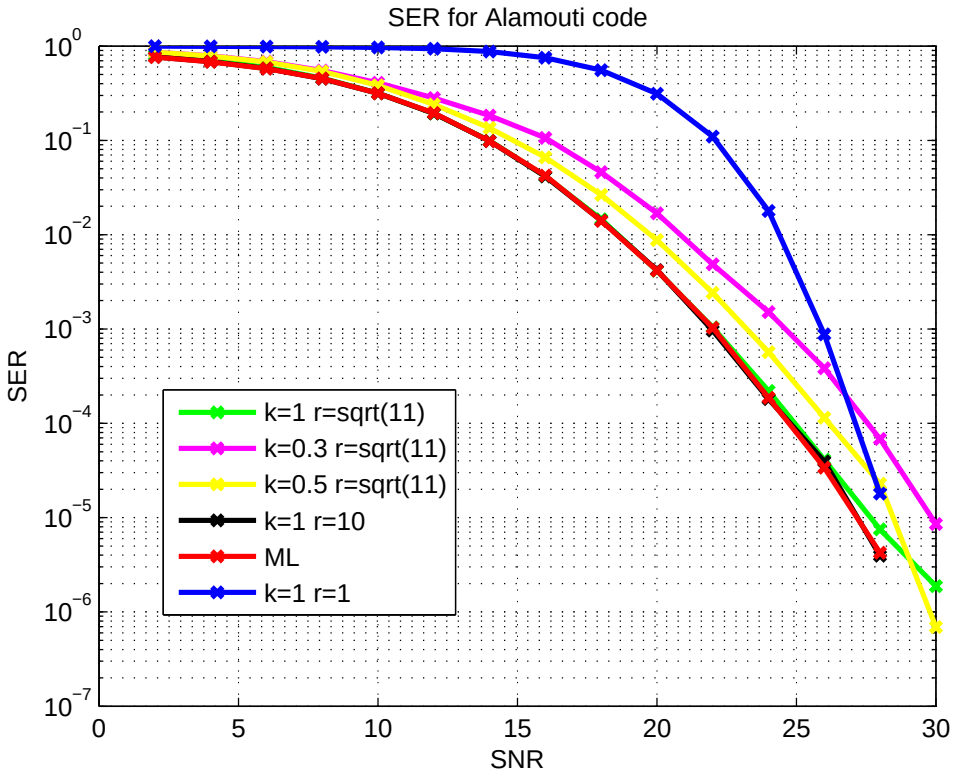

Figure 5.2 SER of Alamouti Code

be examined in Figure 5.3. It can be observed that ML decoding has the highest complexity. The number of FLOPS per symbol is significantly higher than that of the other techniques. The computational complexity of the original sphere decoder can be significantly reduced by reducing the scaling factor to 0.5 and 0.3. The modified sphere decoder has the least complexity compared to the other decoding techniques for a given radius. The computational complexity of the fast sphere decoder lies between that of the original sphere decoder and the modified sphere decoder. Thus, Alamouti code has an SER that improves with increasing

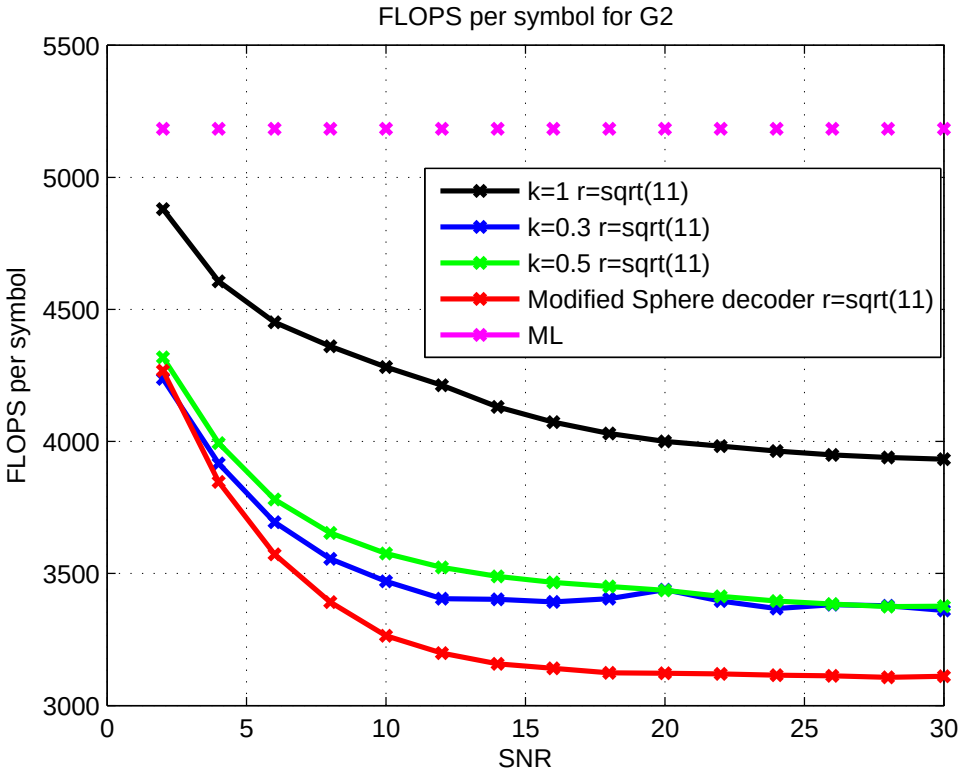

Figure 5.3 Complexity of Alamouti Code

radius. This improvement cannot be noticed when the radius increases beyond $\sqrt{11}$ as all the lattice points are covered for this radius.

5.2. G3 STBC

In case of G_3, there are three transmit antennas with a code rate of 1/2. The SER of the system for G_3 improves with increase in radius. At a radius of 1, there is a big degradation in performance of the sphere decoder as far as the SER is concerned. It can be seen that from Figure 5.4 that there is a good improvement in performance with increase in radius. At a radius of $\sqrt{11}$, it can be noticed that the reduction in scaling factor does not impact the performance much. A radius of $\sqrt{11}$ is not sufficient to examine all the points in the lattice. So, an increase in radius to a value of 5 gives the sphere decoder the option of exploring more points in the lattice, and thus arriving at a better decision. There is a limiting value of radius beyond which there is no significant improvement in SER. Thus, a large radius does not guarantee a better SER. It can also be seen that for very low values of radius, the SER is one for values of SNR upto 20dB.

The complexity of G_3 is analysed in Figures 5.5 and 5.6. It can be observed that at very low values of the radius, the number of FLOPS per symbol is very low for the original

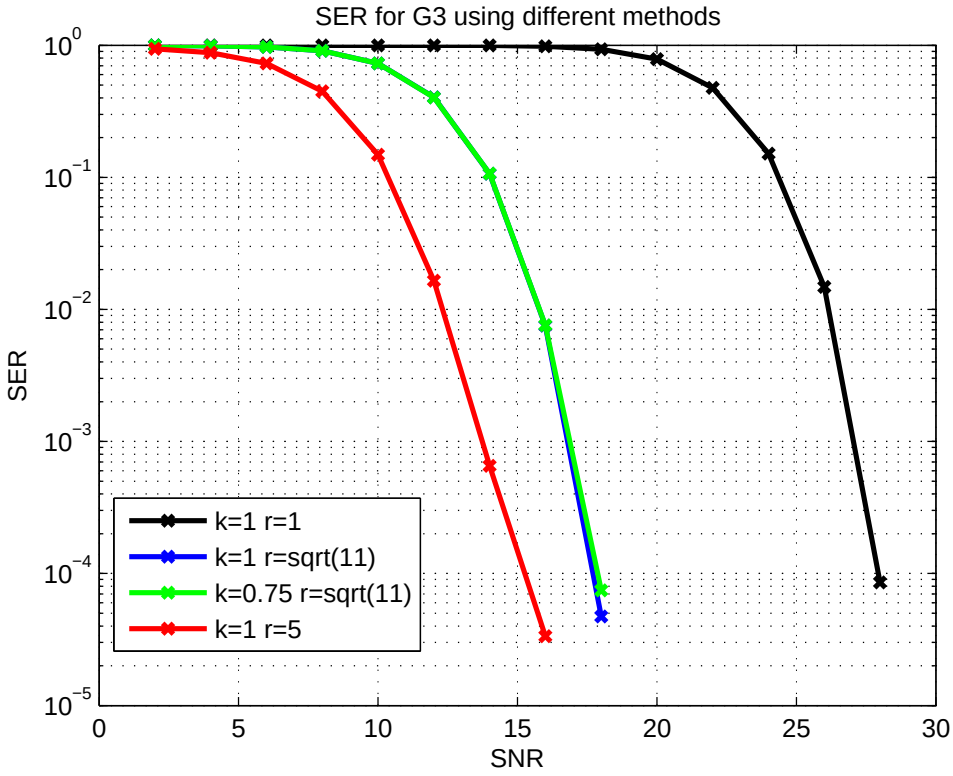

Figure 5.4 SER of G3

sphere decoder. It is only after 20dB that the FLOPS per symbol begins to increase. This is because at low values of SNR, the sphere decoder declares an erasure as it is not able to find sufficient points within the lattice. But as the SNR increases, the number of FLOPS per symbol becomes more consistent. It can also be noticed that with a scaling factor of less than unity, the number of FLOPS per symbol reduces significantly.

From the figures, it can be observed that increasing the radius to 5 does not increase the complexity to a large extent. But since the complexity is considered per symbol, small savings in complexity will result in large savings when transmitting a large quantity of symbols. It can also be observed that at a radius of $\sqrt{11}$, there is large savings in complexity when the sphere decoder changes its complexity from 1 to 0.9, but remains almost consistent when the scaling factor changes from 0.9 to 0.75. This is because the number of points examined does not change by a large amount when the scaling factor is reduced from 0.9 to 0.75 as compared to the change from 1 to 0.9.

5.3. G4 STBC

The SER and complexity of G_4 are analysed in Figures 5.7 and 5.8, respectively.

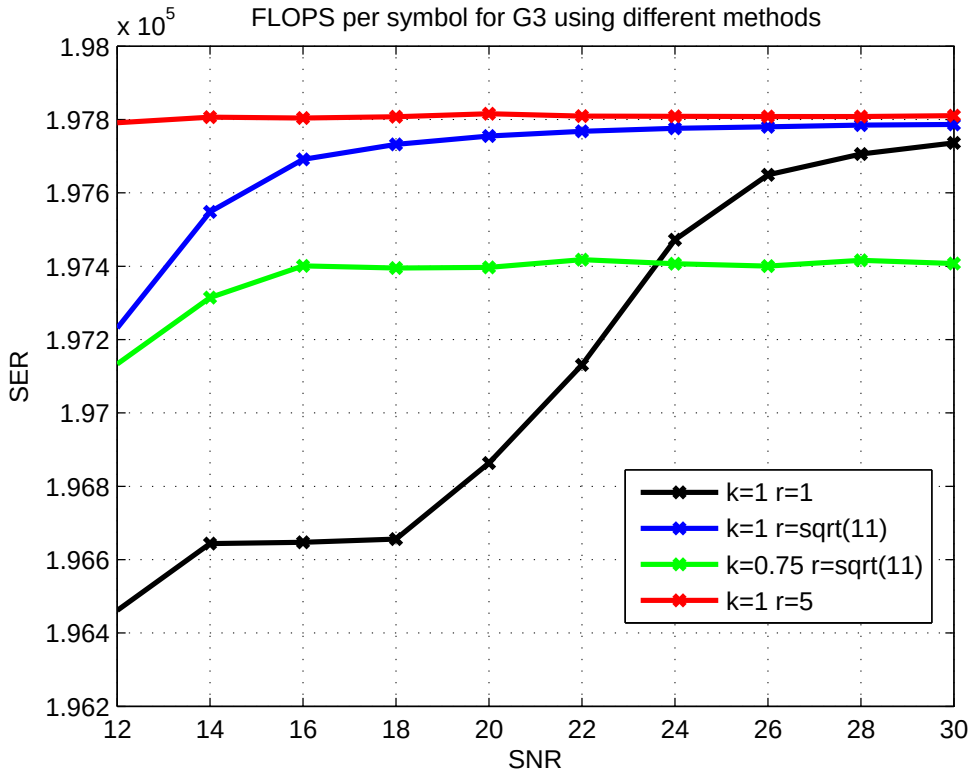

Figure 5.5 Complexity of G3

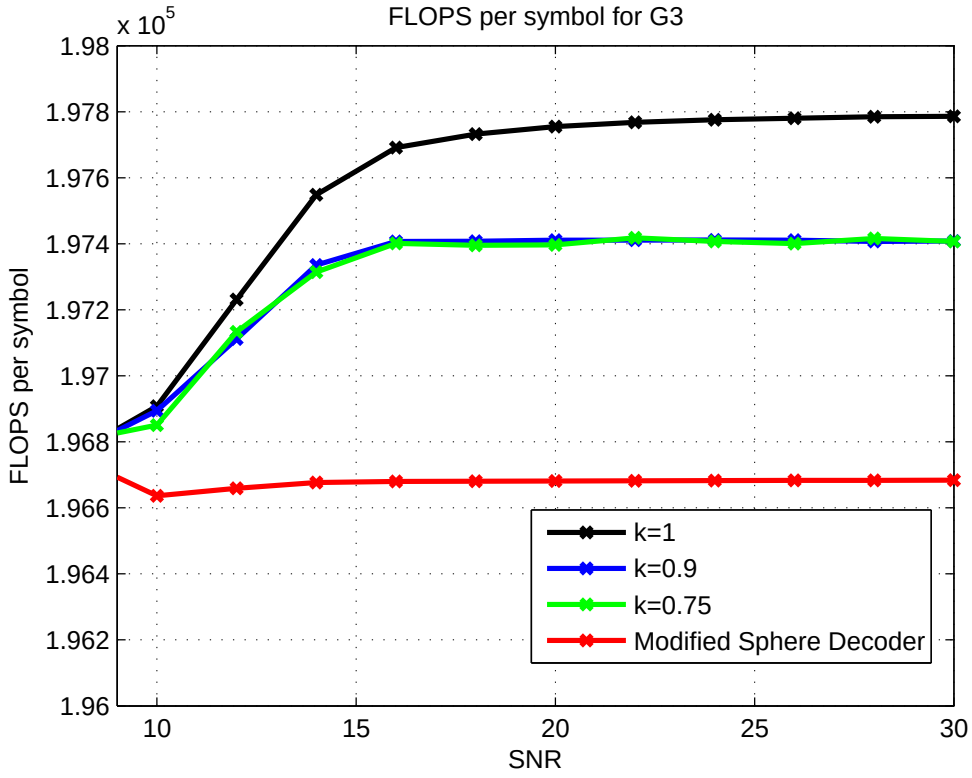

Figure 5.6 Complexity of G3 - Radius $\sqrt{11}$

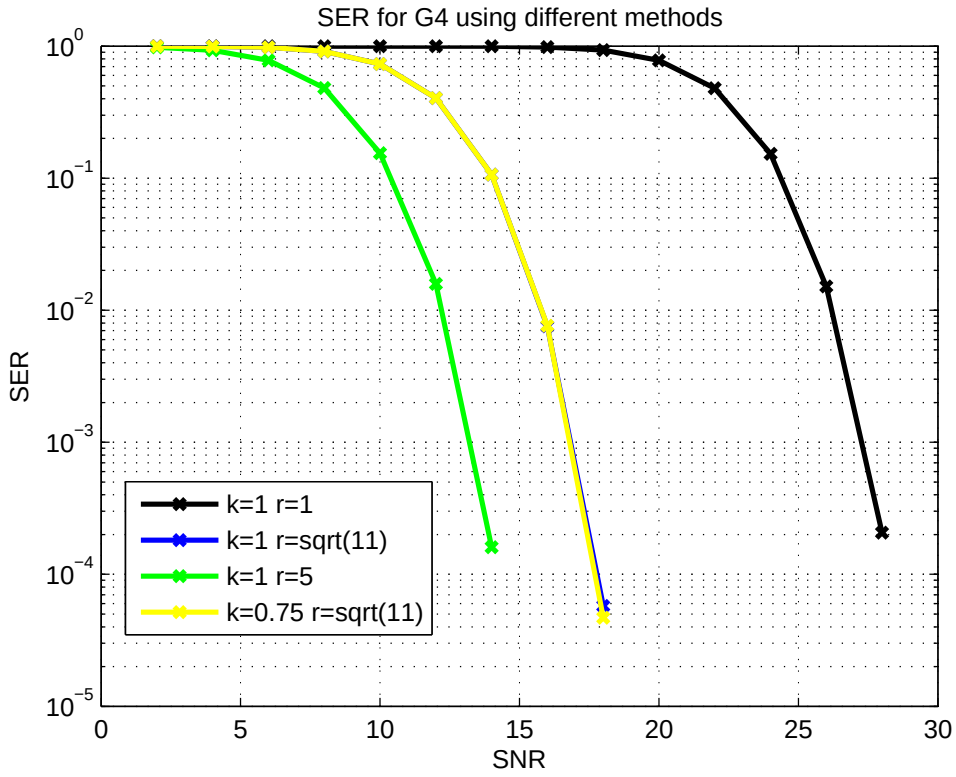

Figure 5.7 SER of G4

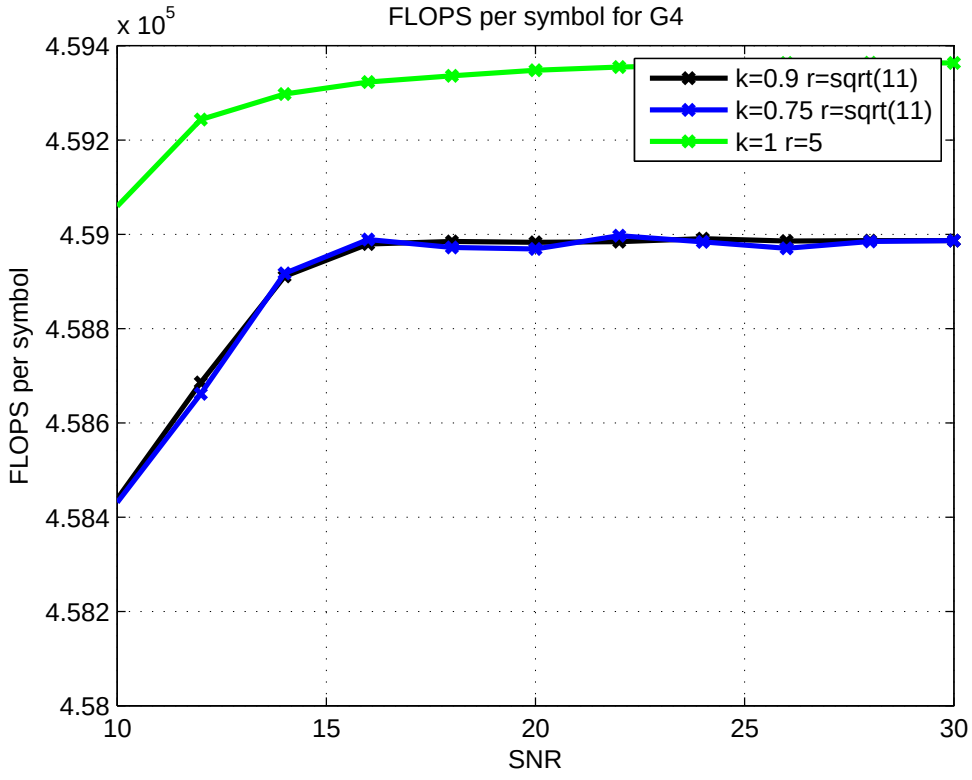

Figure 5.8 Complexity of G4

The complexity graph indicates that as the scaling factor is reduced from 0.9 to 0.75, there is not much difference in the computational complexity. This is because the number of points examined within the lattice does not change by a large amount.

5.4. H3 STBC

The SER and complexity of H_3 are analysed in Figures 5.9 and 5.10, respectively. As in the previous cases, it can be seen that at low values of radius, the SER suffers a degradation in performance. As the radius is increased the SER improves, but it should be noticed that the waterfall curve is not as prominent as the other STBC SER curves. The SER for H_3 has more of linear variations when decoded using sphere decoding. But, the ML curve still exhibits a good fall at higher values of SNR.

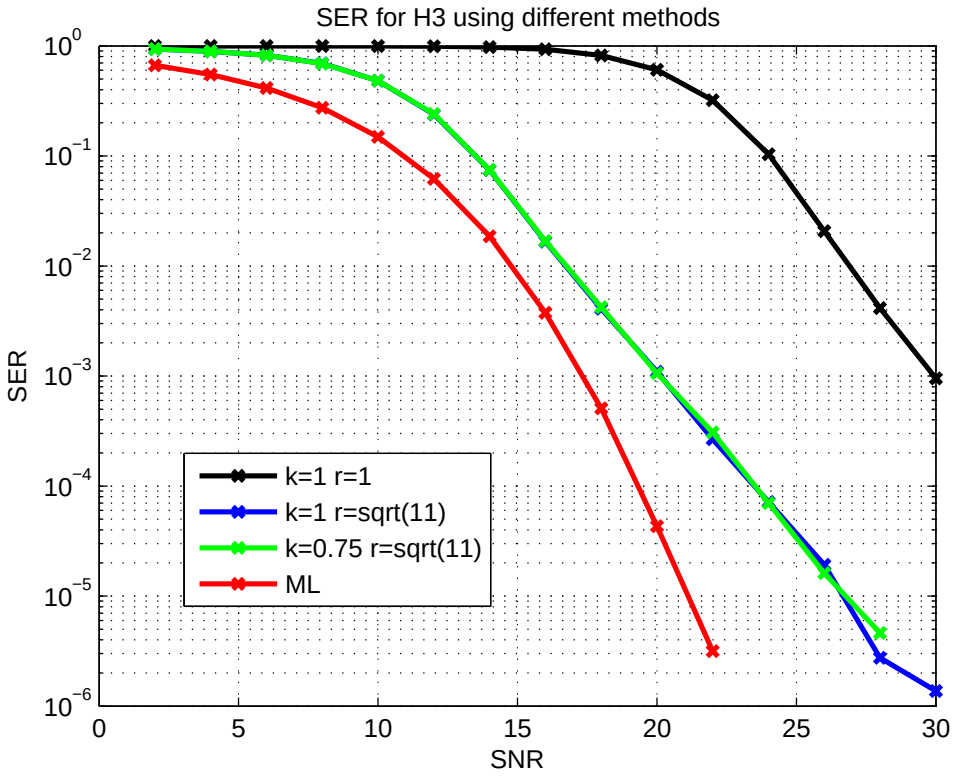

Figure 5.9 SER of H3

From the complexity curve, it can be observed that for a radius of $\sqrt{11}$, a scaling factor of 0.75 has lesser FLOPS per symbol compared to that of the original sphere decoder. It can also be seen that the modified sphere decoder has the least complexity but achieving the same performance as that of the original sphere decoder.

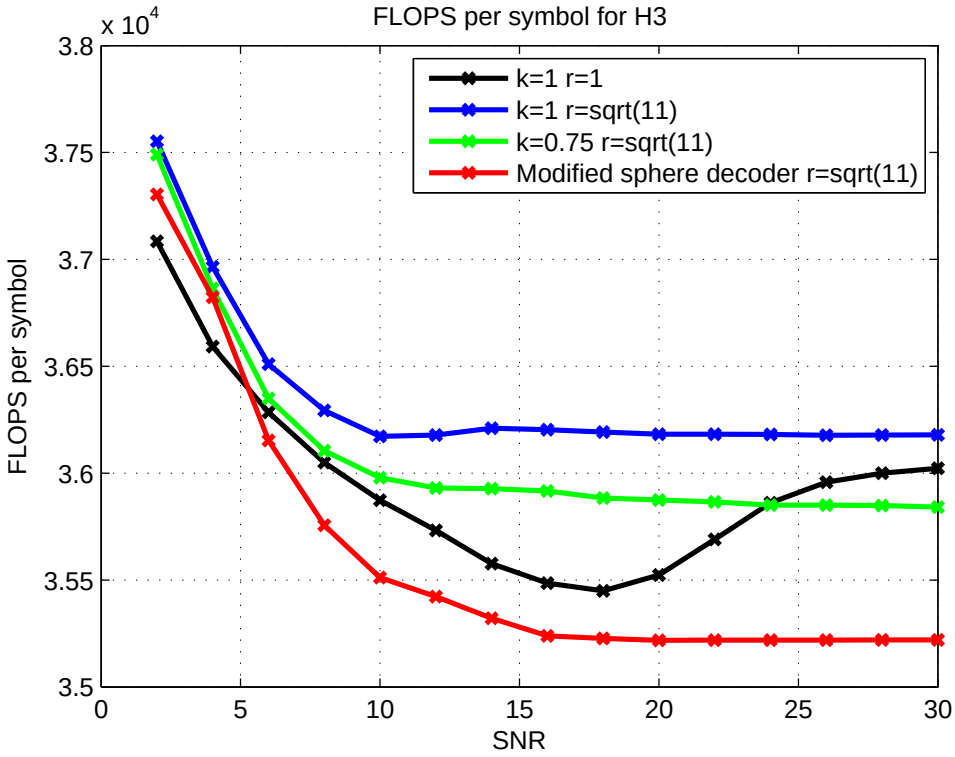

Figure 5.10 Complexity of H3

The linearity of the SER curve suggests that the sphere decoder fails to decode the symbols correctly at high values of SNR. In general practice, the SER of the system falls at a very high rate as the SNR of the system is increased. This could be due to the complex orientation of the lattice.

5.5. H4 STBC

The SER and complexity of H_4 are analysed in Figures 5.11 and 5.12, respectively. From the SER figure, it can be easily observed that reduction of scaling factor facilitates a reduction in complexity. For a radius of 10, it can be seen that the SER is influenced only by the scaling factor. At a radius of $\sqrt{11}$, it can be observed that the SER is high at low values of SNR, but at large values of SNR, it overtakes the performance when k is 0.1 and radius is 10. So, scaling factor is a critical factor when the radius of the sphere is high.

The FLOPS per symbol exhibit significant reduction as the scaling factor is reduced from 1 to 0.75 at a radius of $\sqrt{11}$. The modified sphere decoder clearly achieves the best complexity for a given radius. It can also be observed that there is no reduction in FLOPS per symbol when the scaling factor is reduced from 1 to 0.9.

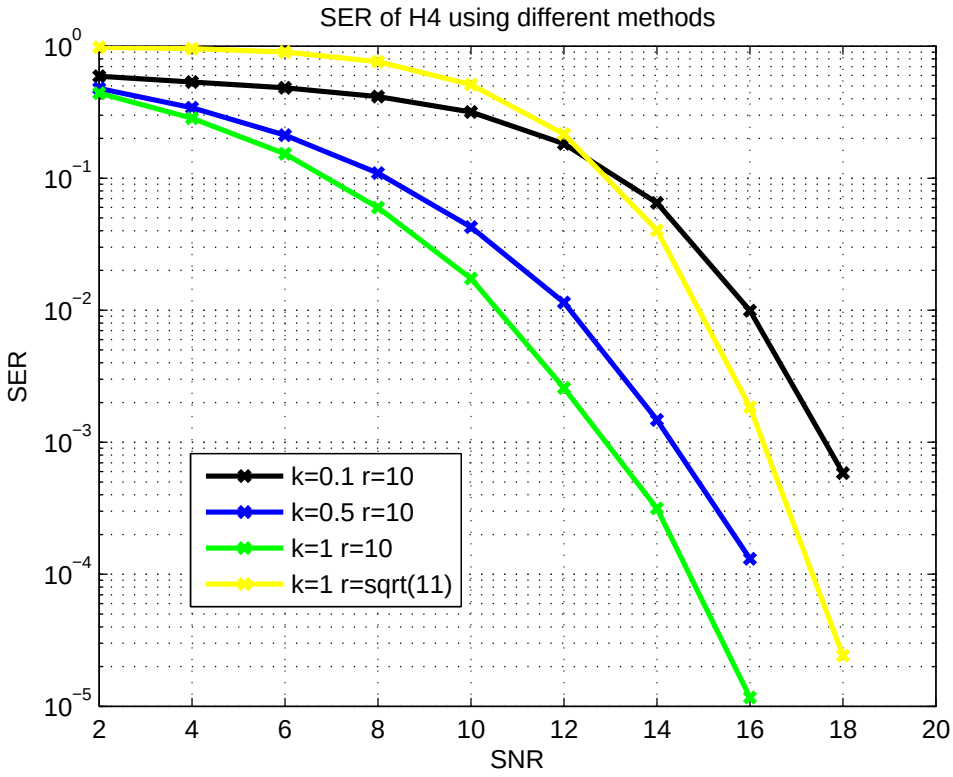

Figure 5.11 SER of H4

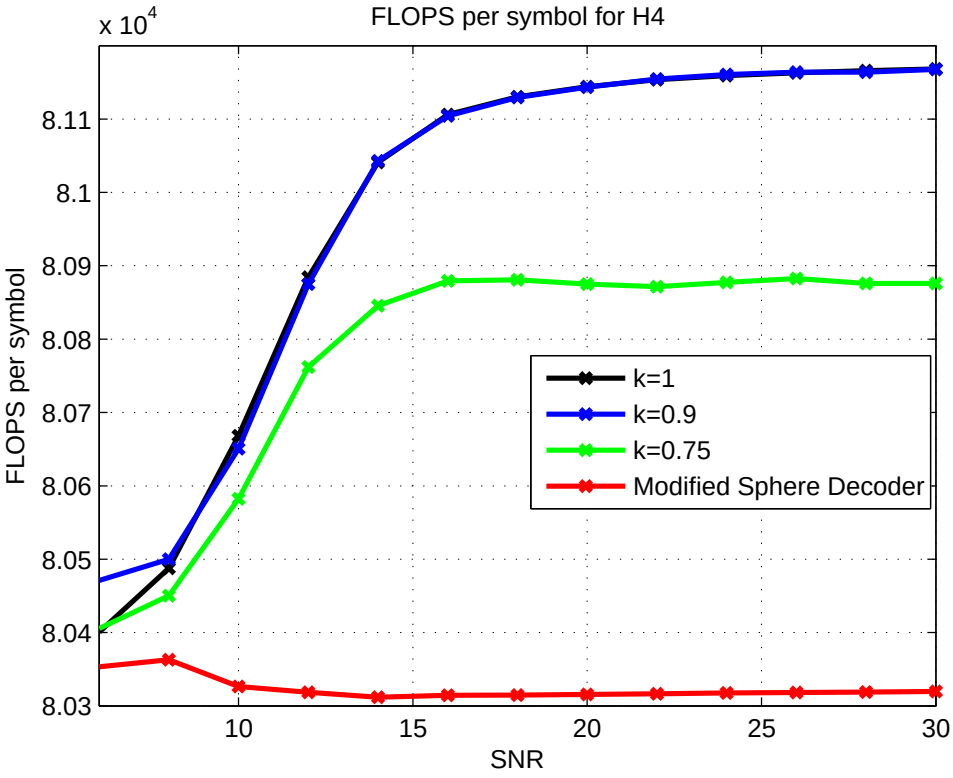

Figure 5.12 Complexity of H4

The graphs clearly indicate the significant changes in computations for different values of the scaling factor. It also shows the importance of determining the right value of radius to achieve a good SER.

5.6. G5 STBC

The SER and complexity of G_5 are analysed in Figures 5.13 and 5.14, respectively. The SER curve clearly shows the difference in SER for different values of radius and scaling factors. At a radius of 10, different scaling factors give different performance graphs. The FLOPS per

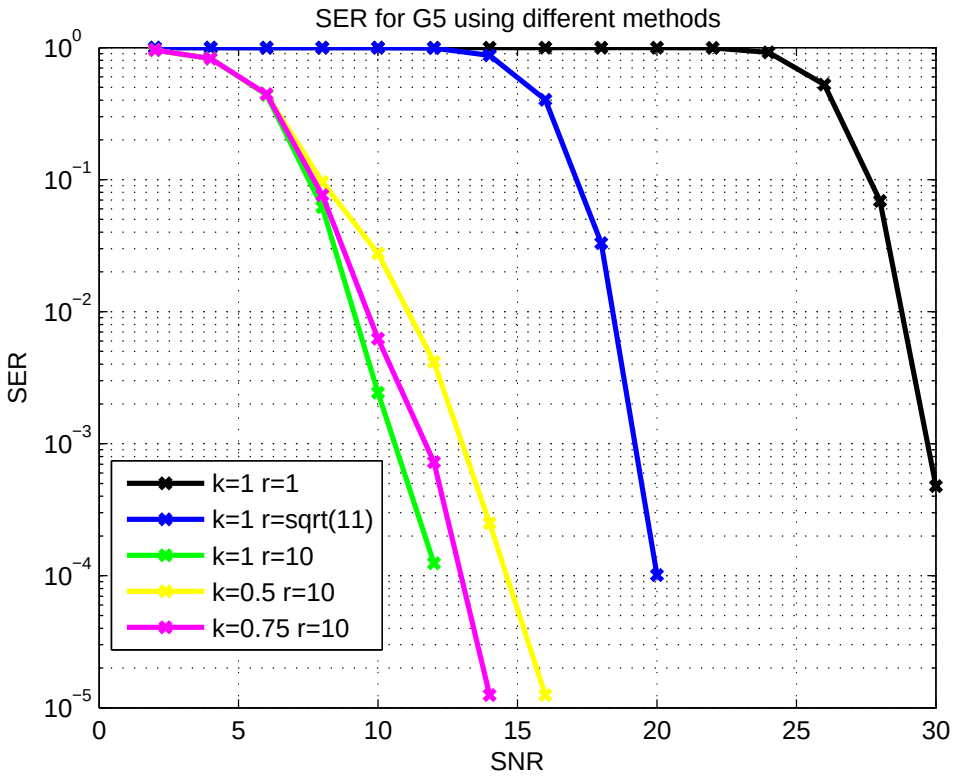

Figure 5.13 SER of G5

symbol for different conditions is obtained in Figure 5.14. It can be seen that at such a high value of radius, when the scaling factor is reduced, the computational complexity reduces considerably. The number of FLOPS per symbol at a scaling factor of 1 for a radius of $\sqrt{11}$ is close to what is obtained when the scaling factor is reduced at considerably large values of radius. The modified sphere decoder achieves large reduction in number of FLOPS per symbol. An overall comparison of SER for different STBCs is as shown in Figure 5.15.

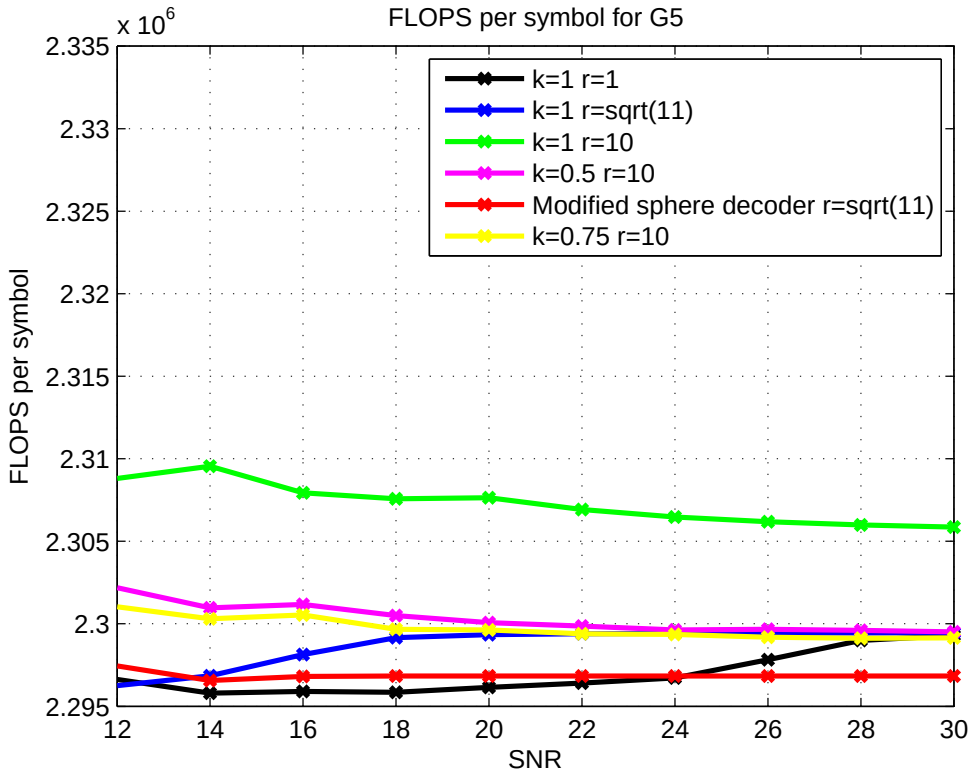

Figure 5.14 Complexity of G5

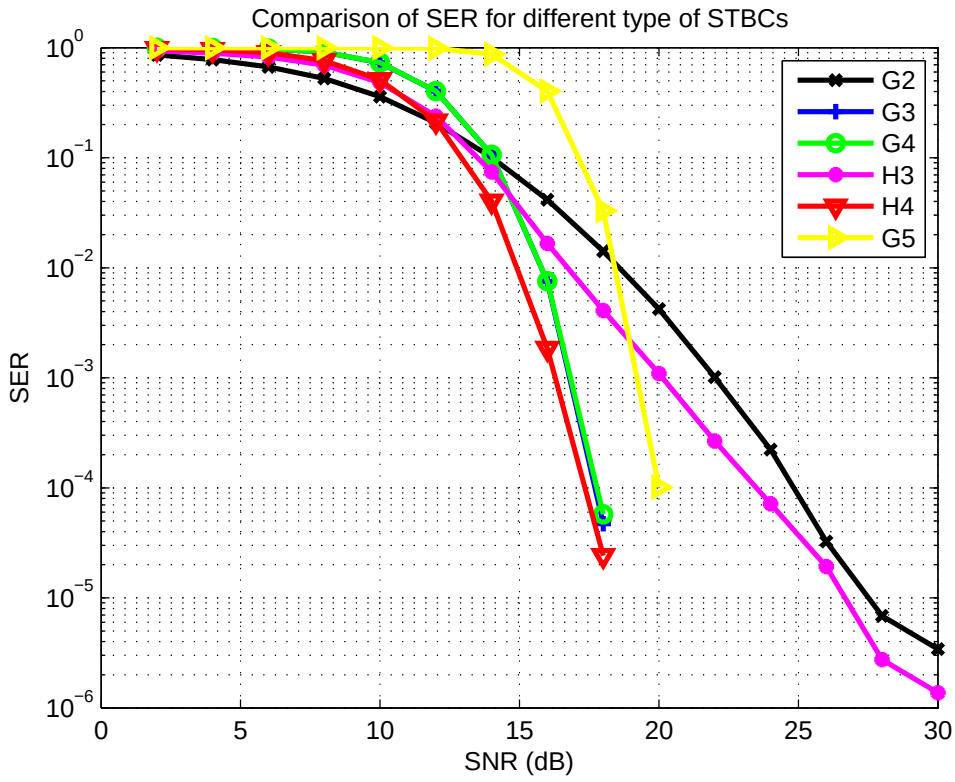

Figure 5.15 Comparison of Different STBCs

6. CONCLUSION

Space Time Block Codes have emerged as popular means of transmitting information over MIMO systems achieving the right balance between diversity and spatial multiplexing. Alamouti code is a well known space time block code that makes use of two transmit antennas. There are other space time block codes for multiple transmit antennas. These codes do not have a code rate of unity, but there is the trade-off of good performance due to better diversity and more antennas at the transmitter and receiver. In this paper, space time block codes for three, four and five transmit antennas are studied and their performance is compared with that of the Alamouti code.

In the first section, we examined the Maximum Likelihood (ML) Detector to decode the different space time block codes. The following result was observed. Increasing(Decreasing) SNR increased(decreased) the performance of the system in terms of the Symbol Error Rate (SER). The FLOPS per symbol is independent of the SNR, since increasing or decreasing the SNR does not change the number of points visited in the constellation. ML decoding is a brute force algorithm with large overhead in the computational complexity.

In the second section, the sphere decoding algorithm is studied along with its different variants. The original sphere decoder algorithm proposed by [9] is carried out for the different space time block codes discussed. The SER for all these codes is determined and the computational complexity of the decoder is significantly less compared to the ML detection. The modified sphere decoding algorithm [8] is an improvement over the original sphere decoding algorithm with significant savings in computation achieving the same SER. The second variant of sphere decoding is when the radius of the sphere decreases at a faster rate than the original sphere decoder. The rate of decrease of the radius is determined by different values of the scaling factor (in this case 0.9 and 0.75). It is observed that as the scaling factor decreases, the computational complexity reduces considerably. The SER is not affected at very large values of the SNR, but at low values of SNR, the decoder suffers a degradation in SER. Thus, it can be concluded that ML decoding is the optimal decoder for achieving the best performance in terms of SER, but at high computational complexity. The original sphere decoder achieves close to ML performance, but a significant reduction in complexity. The modified sphere decoder achieves the same performance as the original sphere decoder in terms of SER, but at a much lesser complexity. Employing scaling factors less than unity achieves a good performance in terms of computational savings, but at a slight degradation in SER at low values of SNR.

Future work includes investigating the use of error correction codes to the Space Time Block Codes to achieve better performance. Channel estimation techniques can be employed to determine the channel at the receiver. Blind and semi-blind approaches can be used to estimate the channel and the effect of SER with imperfect channel estimation can be studied. The sphere decoding algorithms should also be tried for larger constellations for large number of transmit antennas. Since, in real time scenario, the rate of variation of the channel is too fast, it should be investigated to see if there is any technique to track the channel variations per time slot. Further techniques can be explored to achieve a method that combines the features of ML decoding and sphere decoding.

BIBLIOGRAPHY

[1] A. Goldsmith, *Wireless Communications*. United Kingodom: Cambridge University Press, 2005.

[2] H. J. Tarokh, V. and A. R. Calderbank, "Spacetime block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, 1999.

[3] S. M. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[4] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.

[5] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when using Multiple Antennas," *International Journal on Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[6] H. J. Tarokh, V. and A. R. Calderbank, "Spacetime Block Coding for Wireless Communications: Performance Results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, 1999.

[7] X.-B. Liang, "A High-Rate Orthogonal Space-Time Block Code," *IEEE Communications Letters*, vol. 7, pp. 222–223, May 2003.

[8] A. M. Chan and I. Lee, "A New Reduced Complexity Sphere Decoder for Multiple Antenna Systems," *IEEE International Conference on Communications*, vol. 1, pp. 460–464, 2002.

[9] E. Viterbo and J. Boutros, "A Universal Lattice Code Decoder for Fading Channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, 1999.

[10] V. T. V. Razavizadeh, S. M. and P. Azmi, "A New Faster Sphere Decoder for MIMO Systems," *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology*, pp. 86–89, 2003.

[11] V. T. V. Razavizadeh, S. M. and P. Azmi, "Efficient Implementation of Sphere Demodulation," *IEEE Workshop on SPAWC*, pp. 36–40, June 2003.

[12] B. M. Hochwald and S. T. Brink, "Achieving near Capacity on a Multiple-Antenna Channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.

VITA

Praveen G Krishnan was born on 2nd October, 1981 in the southern part of India. He received his Bachelors in Engineering from PESIT, Bangalore, India in July 2003. He has worked for two years with Robert Bosch India Limited, Bangalore, India. He received his MS degree from the University of Missouri-Rolla in May 2007. His topics of interest include signal processing and wireless communications with a specialization in MIMO systems and Space Time Block Codes.