Space-Time Layered Block Codes:
Bridging the Gap Between Maximum Rate and Full Diversity

Patrick Tooher

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Electrical Engineering) at

Concordia University

Montreal, Quebec, Canada

September 2004

Your file Votre référence
ISBN: 0-612-94713-0
Our file Notre référence
ISBN: 0-612-94713-0

ABSTRACT

Space-Time Layered Block Codes:
Bridging the Gap Between Maximum Rate and Full Diversity

by
Patrick Tooher

Impairments found in the wireless channel, such as destructive multipath fading, cannot be fully addressed by using coding alone, despite its recent advances. Recent results show that gains in capacity can be obtained by using multiple antenna elements at the transmitter and the receiver. Instead of mitigating the effects of the multipath fading, multiple-input multiple-output (MIMO) systems use the rich scattering channel to increase the capacity at no bandwidth cost. Methods include BLAST, which maximizes the rate, and space-time coding, which maximizes the diversity.

In this thesis, the fundamental trade-off between rate and diversity is derived for binary codes. Simple codes that can perform at any realizable rate/diversity are designed. These codes are referred to as Space-Time Layered Block Codes (STLBC), since they are in effect a 1-dimensional code layered into a space-time code. By selecting specific 1-D codes, the required diversity can be achieved at the maximal allowable rate.

In order to detect the new STLBC codes, an iterative MMSE detector is used jointly with a soft-input soft-output decoder. This detector/decoder uses a principle similar to Turbo decoding. STLBC codes introduce the concept of coding between layers, which requires a detector capable of dealing with possibly dependent layers. The classical multiuser detector is modified to allow for dependence.

Lastly, techniques to improve the performance of STLBC are presented. These are based on a new design criterion that ensures codebooks of high average rank, as well as making certain the information is spread out over the two available dimensions.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. M.R. Soleymani, whose guidance, timely advice and general support throughout my research, allowed me to complete my M.A.Sc.

I would also like to acknowledge the financial support of the National Sciences and Engineering Research Council of Canada (NSERC) grant, as well as that of the Electrical Engineering Department of Concordia University.

Lastly I would like to thank my family and friends for allowing me to use them to either vent my anger when progress was slow, or to celebrate when it was swift.

Dedicated to my late father, my mother, my sister and my friends.

TABLE OF CONTENTS

LIST OF FIGURES ix

LIST OF SYMBOLS xi

Chapter **Page**

1 INTRODUCTION ... 1

1.1 Temporal Processing ... 2

1.2 Space-Time Processing ... 2

1.2.1 BLAST ... 4

1.2.2 Space-Time Codes ... 5

1.2.3 Full Spatial Diversity ... 7

1.2.4 Full Rate ... 8

1.2.5 Comparison of BLAST with Space-Time Codes ... 9

1.2.6 Space-Time Layered Block Codes ... 9

1.3 Overview of Thesis ... 10

2 THEORY OF MIMO SYSTEMS ... 12

2.1 Signal Model ... 12

2.2 Capacity of MIMO Systems ... 14

2.3 BLAST Revisited ... 18

2.4 Simple Transmit Diversity Scheme ... 24

2.5 Space-Time Code Design Criteria ... 26

2.6 Binary Rank Criterion ... 30

2.7 Space-Time Codes ... 32

2.8 Adding Diversity to Layering 34

2.8.1 Combined Array Processing and Space-Time Coding 35

2.8.2 Threaded Space-Time Architecture 36

3 SPACE-TIME LAYERED BLOCK CODES 39

3.1 The Trade-Off Between Rate and Diversity 40

3.2 Space-Time Layered Block Code System Design 44

3.3 Detection and Decoding of STLBC 46

3.3.1 The Iterative Minimum Mean Square Error (MMSE) Detector .. 47

3.3.2 Decoding of STLBC 53

3.4 Performance of Iterative MMSE Detector and Decoder 58

3.4.1 Results 60

4 IMPROVEMENTS ON SPACE-TIME LAYERED BLOCK CODES 68

4.1 A New Criterion for Space-Time Codes 68

4.2 Maximizing the Use of Spatial Diversity on the information 71

4.3 Performance of Pseudo-Systematic STLBC 75

5 CONCLUSION AND FURTHER WORK 82

REFERENCES 85

LIST OF FIGURES

1.1 The Block Diagram of a BLAST system.......... 4

1.2 An example of a space-time coded system. (a) Block diagram, (b) Code trellis. . 6

1.3 Space-time codeword.......... 7

1.4 Sample space-time layered block code codeword matrix.......... 10

2.1 Capacity: Complementary Cumulative Distribution Function (in b/s/Hz/dim.).[4] 16

2.2 a) Capacity in b/s/Hz vs. number of antennas. b) Capacity in b/s/Hz/dimension vs. number of antennas [8]. 17

2.3 Transmission patterns in layered architecture. a) Diagonal BLAST (D-BLAST). b) Vertical BLAST (V-BLAST). 19

2.4 BLAST detection using nulling approach. 19

2.5 BLAST performance with error propagation. 23

2.6 Comparison of ZF and MMSE detection, with and without ordering of the signals, n=m=4, QPSK modulation, V-BLAST architecture. 24

2.7 Two-branch transmit diversity scheme with two receivers. 25

2.8 (a) Space-time trellis code with n=2 and full rate, (b) Space-time block code with n=3 and rate ½. 33

2.9 Combining layering with space-time codes.......... 35

2.10 Simple TST architecture. 37

3.1 Layered codeword matrix. 43

3.2 Diversity versus rate in a binary space-time layered block code with n=3. 44

3.3 Space-time layered block code (STLBC) transmitter block diagram.......... 45

3.4 Space-Time Layered Block Code (STLBC) encoder structure. 45

3.5 Iterative Multiuser detector for layered space-time signals.......... 47

3.6 The combined iterative MMSE detector and Pyndiah Decoder. 57

3.7 Performance updating a priori probabilities at the end of each block vs. at the end of each layer. 61

3.8 Performance without signal ordering. 63

3.9 Performance using signal ordering. 64

3.10 Performance after each iteration. 65

3.11 Performance for the iterative MMSE algorithm versus the zero-forcing algorithm. 66

3.12 Performance of simple coded scheme with non-coded scheme. 67

4.1 Example of a systematic codeword generator matrix. 71

4.2 (a) Layering of systematic 1-D codeword into codeword matrix. (b) Layering of modified 1-D codeword into codeword matrix. 72

4.3 Pseudo-systematic codeword matrix. 74

4.4 Translated codeword matrix. 74

4.5 Comparison of error performance between two codebooks achieving different *minrank*. 77

4.6 Comparison of error performance between two codebooks achieving different *avgrank*. 77

4.7 Comparison of error performance between two codes with different *dmin*. 79

4.8 Performance of high rate STLBC compared to non-coded BLAST. 80

4.9 Comparison of STLBC with simple layered coded scheme. 81

LIST OF SYMBOLS

n	number of transmit antennas
m	number of receive antennas
l	length of codeword
2^b	size of signaling constellation
i_j	information bit
p_j	parity bit
$\mathbf{s}$	transmitted signal matrix
$\hat{P}$	transmitted power
$\mathbf{r}$	received signal matrix
P	received power
$\mathbf{v}$	AWGN noise matrix
$N_o/2$	noise variance per dimension
ρ	SNR: signal-to-noise ratio
H	normalized matrix channel input response
C	capacity
G_i	nulling matrix
$\mathbf{e}$	erroneous codeword matrix
E_s	energy per symbol
λ_j	eigenvalue
d	diversity
$\underline{x}$	information vector
k	length of information vector
R	rate
γ	codeword matrix $\in \{0,1\}$
c	codeword vector
$\underline{w}_f$	feed-forward coefficient for detector
$\underline{w}_b$	feedback coefficient for detector

λ_t^i	extrinsic information
P_A	probability that binary event A=1
$\Lambda(\cdot)$	log likelihood ratio
y_j	detected signal
β	estimation scaling factor
$\alpha(j)$	scaling factor on extrinsic information
$\bar{e}$	mean number of errors
b	number of samples
$1-\alpha$	probability of statistical mean converging to the true mean
ε	tolerance on the mean
d_{min}	minimum Hamming distance

Chapter 1

Introduction

In order to achieve present and future goals in the development of wireless communications, wireless networks must be capable of high data rate applications rather than just current voice-based services. Barriers to the development of wireless communication services include low bit rates, high power consumption and high cost per bit. To reduce the error rate, the power requirement and the cost per bit, signal processing methods are applied. These methods can increase future wireless data rates without expanding the current bandwidth needs.

The two following categories of signal processing are of interest:

- Temporal processing: includes well-studied channel codes, which have recently been shown to achieve near capacity results.
- Space-time array processing: employs multiple antennas at the transmitter and the receiver.

In this chapter, these two methods of processing are explained in-depth. Systems using space-time processing are also introduced.

1.1 Temporal Processing

At its core, temporal processing is the use of redundancy in the data to ensure that even with the imperfections of the wireless channel, data can be fully recovered. The most basic of these methods, also very inefficient, would be to transmit the same information several times until the receiver can reconstruct the data. In 1948 C.E. Shannon [1] developed fundamental limits on the efficiency of telecommunications over noisy channels. This theorem states that there exist achievable codes with code rate close to that of the capacity of the channel, having a probability of error approaching zero. It was thought to be a purely theoretical limit, until almost fifty years later when codes were indeed developed to achieve capacity. Forney's concatenated codes [2] were the first step in finding a class of codes whose error probability decreases exponentially while the decoding complexity increases only algebraically. It was also showed that the optimal decoding method for cascaded codes is a soft-input operation. More recently Turbo codes [3], which are the culmination of this effort of constructing powerful codes from simple codes, have been developed.

1.2 Space-Time Processing

Of much greater interest in this thesis is the combination of temporal and spatial processing. The wireless channel suffers from attenuation due to the destructive effects of the addition of multipaths in the propagation media and due to the interference of other users. Error control coding, essentially temporal processing, does not address these impairments. Space-time processing, which includes space-time coding as a special case, flips the fading problem on its head. Instead of trying to mitigate the effects of fading and multiple access interference, the use of multiple antenna arrays at the transmit and receive sides of a wireless link in combination with signal processing and coding, optimizes the spectral efficiency by using the effects of fading to its benefit.

The basic model used to analyze a space-time processing system where multiple antennas are employed at both the transmitter and the receiver is known as a multiple-input multiple-output (MIMO) system; where the inputs are the transmitter antennas and the outputs are the receiver antennas.

A MIMO system can be seen as multiple SISO (single-input single-output) channels and thus its capacity is the sum of the individual capacities of these SISO sub-channels. Foschini [4] has shown that the capacity over the MIMO fading channel can grow linearly with the number of transmit or receive antennas. This increase in performance is achieved by exploiting both the inherent spatial and temporal diversity related to specific multipath wireless mobile channels.

Diversity at the receiver (a so-called SIMO channel) is a well-studied subject with a large body of work associated with it [5, 6, 7]. Adding an antenna array at the transmitter with sufficiently separated elements creates a system with transmit and receive diversity, producing a multiplicity of sub-channels.

Even though in this dissertation only a single user is considered, the problem of interference is still prevalent. Namely one must deal with the potentially detrimental effects of the unavoidable mutual interference among multiple SISO sub-channels of one MIMO channel. Originally, two main strategies were developed to deal with this interference: Bell Laboratories Layered Space Time Wireless Architecture (BLAST) and Space-Time Codes (STC). These two strategies basically exploit, rather than mitigate, the multipath channels in order to achieve heretofore unknowably feasible high spectral efficiencies (b/s/Hz).

1.2.1 BLAST

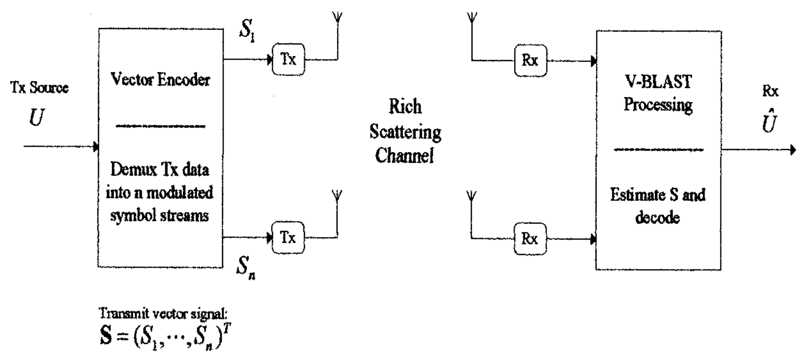


Figure 1.1. The Block Diagram of a BLAST system.

BLAST [8, 9, 10] takes advantage of the spatial dimension by transmitting and detecting independent co-channel data streams using multiple antennas. This results in a system that is bandwidth efficient. Figure 1.1 shows an example of a BLAST system, where the data from the transmitter is split into multiple substreams which are transmitted simultaneously via multiple antennas. Since all the substreams are transmitted in the same frequency band, there is a very efficient use of the spectrum. At the receiver end, multiple antennas pick up weighted additions of all of the various transmitted substreams. The differences in each sub-channel's characteristic (i.e. fading coefficient) allow for the data substreams to be separated at the receiver end by use of clever signal processing algorithms. In this way, BLAST systems act very much like multiple-user spread spectrum systems, where the transmit antennas are analogous to the users and the receive antennas are analogous to the spread coding gains. Therefore, much like in multiple-user detection, non-linear sub-optimal detection is used at the receiver end to separate incoming signals. As mentioned previously, the unavoidable multipaths in a wireless channel are accordingly exploited by using the spatial dimension to increase the data transmission rates.

1.2.2 Space-Time Codes

In BLAST systems, the advantage over using single antenna transmitter and receiver is that the data transmission rates can be increased dramatically without a need for higher bandwidth. At the other end of the MIMO spectrum lays space-time codes (STC). STCs are obtained when coding redundancy is introduced not only over the time domain, but also over the space domain. Coding thus becomes a two-dimensional problem. Low-complexity codes, such as the one shown in Figure 1.2, were developed at AT&T Research Labs by Tarokh *et al.* [11, 12, 13, 14, 15].

The simplest example available of a STC is a two antenna delay diversity scheme. This is where the data is first encoded with a channel code, which in this case is a repetition code of length 2. Then the output of the repetition code is sent to two parallel data streams which are transmitted with a symbol delay between them. A slightly more complex example of a space-time coding system is given in Figure 1.2 for a 4-PSK 2 b/s/Hz code. In this two transmit and two receive antenna system, the data is encoded using the 4 state trellis diagram shown in Figure 1.2 (b). The serial to parallel converter divides the encoded data into two substreams that are each modulated and transmitted via their own antenna. The decoder is a conventional Viterbi algorithm that computes the trellis transition path metrics and then makes a decision on the decoded data based on the minimum accumulated metric.

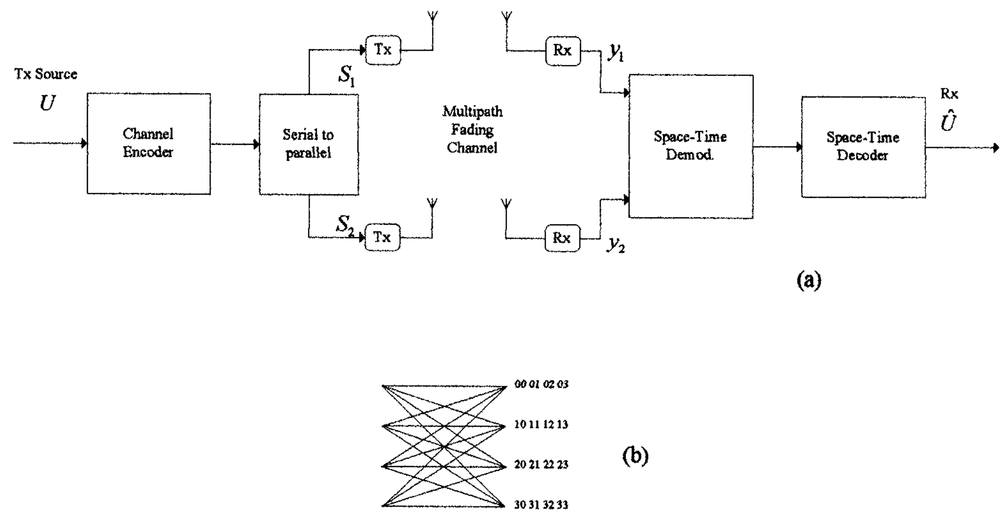


Figure 1.2. An example of a space-time coded system. (a) Block diagram, (b) Code trellis.

At the beginning of this section it was mentioned that space-time codes fall at the opposite end of the MIMO spectrum from BLAST. While BLAST employs the spatial dimension to increase the data transmission rates of codes, space-time codes employ the spatial dimension by using its diversity potential and thus improve the performance of codes at no bandwidth cost.

1.2.3 Full Spatial Diversity

Full spatial diversity is defined as the maximum achievable diversity advantage of using a MIMO space-time code over the use of a regular SISO temporal code. If a MISO system is assumed, the n transmit antennas provide n independent paths for the data to travel to the receiver. If appropriate signal processing is used, it is obvious that this MISO channels will have a much better error performance than a regular SISO channel. A full analysis of spatial diversity will be shown later, however a brief overview is now given. Each possible transmitted signal difference (or signal codeword if the code is linear) in a modulation is represented in matrix form, where the vertical axis represents the space dimension and the horizontal axis represents the time (Figure 1.3).

Time axis →

Space axis ↓	time=1	time=2	…	time = l
antenna 1	S_1	S_{n+1}	…	$S_{n(l-1)+1}$
antenna 2	S_2	S_{n+2}	…	$S_{n(l-1)+2}$
	…	…	…	…
antenna n	S_n	S_{n+n}	…	S_{nl}

Figure 1.3. Space-time codeword.

The rank of this set of complex-valued matrices determines the spatial diversity. An increase in this rank, that is to say an increase in diversity, improves the error performance of this code by an exponential decrease of decoding error rate versus signal-to-noise ratio (asymptotic slope of the performance curve on a log-log scale). In other words, as the diversity is increases, the asymptotic slope of the performance curve becomes steeper. If the codeword length l is greater than the number of transmit antennas

n, the maximum possible rank of that matrix is therefore n. In [11], Tarokh shows that full spatial diversity is obtained if and only if all of the codeword pair difference matrices exhibit a rank equal to n.

1.2.4 Full Rate

In basic temporal coding, a full rate code is one with no redundancy. That is to say, every symbol sent is an information symbol and thus the code achieves maximum data transfer for the allowed bandwidth. This is obviously not the optimal way of transferring data, as the lack of redundancy can make the number of errors reach detrimental levels depending at what signal-to-noise ratio the data is being transmitted. The concept of full rate in a space-time system is of more value. In a space-time system, the maximum allowable rate when making use of all the available spatial diversity is known as the full rate. It is interesting to note that regardless of the number of transmit antennas, the full rate of a space-time system is in fact the maximum data rate allowable if only one antenna was being used with no redundancy. For example, if a BPSK modulation scheme is used, the full rate of a space-time code with any number of transmit antennas n, is 1 b/s/Hz. Increasing the number of transmit antennas will increase the diversity of the code and not affect the data transmission rate.

On the other hand, BLAST does not try to maximize the spatial diversity potential obtained by increasing the number of antennas. Therefore, in the context of comparison, a system, such as BLAST, that does not achieve full spatial diversity can therefore achieve rate higher than the so-called full rate, up to maximum rate.

1.2.5 Comparison of BLAST with Space-Time Codes

The difference between BLAST and space-time coding is the correlation of signals across the spatial dimension (i.e. the transmit antennas). The decoding complexity and performance of space-time codes is similar to regular trellis codes for AWGN channels if the fading path gains are assumed known at the receiver. On the other hand, BLAST decoding, while of a similar complexity, is prone to error propagation. That is the presence of errors in an early stage of decoding is compounded as the decoding process carries on. In [16], Bevan setup a simulation experiment to compare the performance of BLAST and space-time codes. For frame size of 400 information bits and a frame error rate of $FER=10^{-2}$, the 32 state 4-PSK space-time code had an advantage of 2-3dB over the BLAST system.

1.2.6 Space-Time Layered Block Codes

In this thesis, space-time layered block codes (STLBC) are presented. These codes can operate at either end of the rate/diversity optimization curve and any point in between. They allow for coding between layers, and therefore maximize the spatial and temporal redundancy for each information bit. STLBC are in effect a superset of all available MIMO codes, tweaked differently they can provide BLAST codes or space-time codes. An example of a codeword in an STLBC is given in Figure 1.4.

i_1	p_1	i_8	i_4	p_1
i_6	i_2	p_1	p_1	i_5
p_1	i_7	i_3	p_1	p_1

i_j Information bit

p_1 Parity bit

Figure 1.4. Sample space-time layered block code codeword matrix.

1.3 Overview of Thesis

The first goal of this thesis is to produce a simple space-time code that truly utilizes all the spatial and temporal diversity available, for different rates and spatial diversity advantages. Unlike BLAST [8] and threaded space-time codes [17], the coding is not confined to a layer. The coding also serves two purposes, first to ensure a required diversity level and second to provide error control coding. Coined space-time layered block codes (STLBC), since it can be thought of as a 1-dimensional block code layered into a 2-dimensional codeword, this new design provides the freedom to increase the probability of error performance of BLAST codes, without greatly affecting their data rates.

The second goal of this thesis is to develop an iterative minimum mean square error (MMSE) detector based on that provided for multiuser detection by El Gamal [18]. Since in STLBC the coding can be done over layers, the iterative MMSE detector must therefore be adapted for dependent information.

The last goal of this thesis is to improve the performance the probability of error performance of STLBC. By better employing the available spatial and temporal diversity, the performance can be improved.

Outline

This thesis is organized as follows. In Chapter 2, an overview of the theory behind MIMO systems is given. A simple derivation of the increased capacity is given. In order to show a method of using this increased capacity, it is followed by a study of the detection of BLAST codes. Next space-time codes, employing full diversity, are studied. The pairwise probability of error from Tarokh [11] is given, as are the criteria to develop good codes. The binary rank criterion [19] is provided to help with the algebraic construction of space-time codes. Lastly, examples of codes combining the added rate with the added diversity are offered.

In Chapter 3, the trade-off between the diversity and the rate of binary codes is derived. Next, simple STLBC codes that combine the diversity and rate advantage are constructed. The detection and decoding of these codes via an iterative MMSE detector and Pyndiah decoder, is given. The detector is modified to allow for dependent layers, as is the case with STLBCs.

In Chapter 4, tools are provided to increase the performance of STLBCs and specific codes are generated and their performance is simulated. Finally, in Chapter 5, the conclusion is given, along with further work on STLBCs.

Chapter 2

Theory of MIMO Systems

In the introduction, the need for MIMO systems, as well as a brief overview of them, was presented. Most basic concepts of communications are well-studied for SISO systems; however some need to be adapted for MIMO. In this chapter several theoretic concepts about MIMO systems are presented in order to fill the gaps left from knowledge of SISO systems. The outage capacity for fading channels, as developed by Foschini [4], is firstly presented. Then more technical descriptions of BLAST and space-time codes are given. The binary rank criterion, as presented by Hammons [19], is then offered as a means of simplifying the task of code construction. Finally codes trying to utilize both diversity gain and rate improvements available in MIMO, such as the threaded architecture of El Gamal [17] are presented.

2.1 Signal Model

The MIMO system is viewed as a complex baseband model, using a fixed matrix channel (i.e. quasistatic) with additive white Gaussian noise (AWGN). The channel is fixed for any discrete time, although it is taken to be random. The following variables also need to be defined:

- There are n transmit antennas and m receive antennas.
- The transmitted signal, s(t), is an n-dimensional signal with fixed total power $\hat{P}$ regardless of n.
- The noise at the receiver, v(t), is a vector of m independent samples of a zero-mean complex Gaussian random variable with variance $N_0/2$ per dimension.
- The received signal, r(t), is an m-dimensional signal with average power P.
- The average signal-to-noise ratio at each receive branch is given by $\rho = \dfrac{P}{N_0}$.
- The matrix channel input response is given by g(t) and its Fourier transform is $G(f)$. The normalized matrix channel input response is given by h(t) and $H(f)$, where $\hat{P}^{1/2} \cdot G = P^{1/2} \cdot H$.

The basic vector equation describing the channel operating on the signal is thus given by

$$r(t) = g(t) * s(t) + v(t). \tag{2.1}$$

In normalized form and using the narrowband assumption, whereby the channel Fourier transform $G(f)$ is treated as a matrix constant; the normalized form of equation (2.1) is given as

$$r(t) = (P/(\hat{P} \cdot n))^{1/2} \cdot h \cdot s(t) + v(t). \tag{2.2}$$

If antenna elements are placed a distance of half a wavelength apart, the path losses tend to decorrelate [20]. Therefore when both the transmit and receive antenna arrays are properly spaced out, the Rayleigh model for the $m \times n$ matrix H is approximated by a matrix having the following independent identically distributed (iid) complex, zero-mean, unit-variance entries:

$H_{ij} = N(0, 1/\sqrt{2}) + \sqrt{-1} \cdot N(0, 1/\sqrt{2})$ and

$|H_{ij}|^2$ is a chi-squared variate with two degrees of freedom denoted by χ_2^2.

2.2 Capacity of MIMO Systems

Communication is thought of as being a series of bursts. These bursts are assumed short enough in duration so that the channel remains unchanged throughout their lengths. The channel is assumed unknown by the transmitter and tracked by the receiver. By unknown to the transmitter, it means that the realization of H during a burst is unidentified. The transmitter assumes that the communication is taking place with a receiver for which a certain m and ρ are available. These values ensure the transmitter that the channel offers a certain achievable capacity. However, due to the volatility of the channel, not all communication bursts are successful. There are some realizations of H during which the capacity value is too optimistic, and the required bit error rate cannot be achieved at the chosen transmission rate. Thus there is a channel outage and the channel is said to be in the OUT state. This leads one to be interested in the capacity that can be attained in a specific percentage of transmissions (for example, one may want the capacity to be achievable 99% of the time).

In [4] and again in [8], Foschini derives the outage capacity of a MIMO system in a fading environment. In order to provide the maximum capacity, it is assumed the transmitted signal vector s(t) is composed of n statistically independent components each with a Gaussian distribution. The capacity expression can therefore be derived from the general formula appearing in [21]

$$C = \log_2 \frac{\det A_s \cdot \det A_T}{\det A_u} \tag{2.3}$$

where $A_s = E(ss^*) = \hat{P}/n \cdot I_n$, $A_r = E(rr^*) = N_0 \cdot I_m + \hat{P}/n \cdot GG^*$, and $A_u = E(uu^*)$ where u is the $n+m$ dimensional vector $(s,r)^*$. Therefore A_u has A_s in the top left corner, A_r in the bottom right corner, $\hat{P}/n \cdot G^*$ in the top right corner and its complex conjugate in the remaining corner. The statistical independence between all the components of the signal vector s and noise vector v led to the simple computation of A_s, A_r and A_u.

The following identity

$$\det \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \det A \cdot \det(D - CA^{-1}B)$$

leads to the required expression for capacity [4], namely

$$C = \log_2 \det[I_{n_R} + (\rho / n_T) \cdot HH^*] \text{ b/s/Hz.} \tag{2.4}$$

This equation can be simplified depending on whether only receive diversity, transmit diversity or a combination of the two is used. For example, in the simple case of optimum ratio combining (n=1 and m=m), the capacity is given as

$$C = \log_2 \lfloor 1 + \rho \cdot \chi_{2n}^2 \rfloor \text{b/s/Hz.} \tag{2.5}$$

The case that is of interest in this dissertation is the case where there are multiple antenna arrays at both the transmitter and the receiver, and $n \geq m$. In such a combined transmit-receive diversity scheme, the capacity is lower bounded by

$$C > \sum_{k=n-(m-1)}^{n} \log_2\left[1 + (\rho/n) \cdot \chi_{2k}^2\right]. \tag{2.6}$$

In [4], Foschini shows that if the number of transmit and receive antennas are the same, namely m=n, as that number increases, so to does the lower bound given in (2.6) and it does so in a linear fashion. It can therefore be obviously seen that extra capacity is much easier to exploit using MIMO systems, than it is in SISO systems.

In Figure 2.1, the capacity is given for varying outage probabilities and at varying signal-to-noise ratios. It is seen that with high n and m, there is little gain in capacity using a percentile lower than .98. This means that the capacity can be guaranteed for as much as 98% of the time, while barely affecting its value.

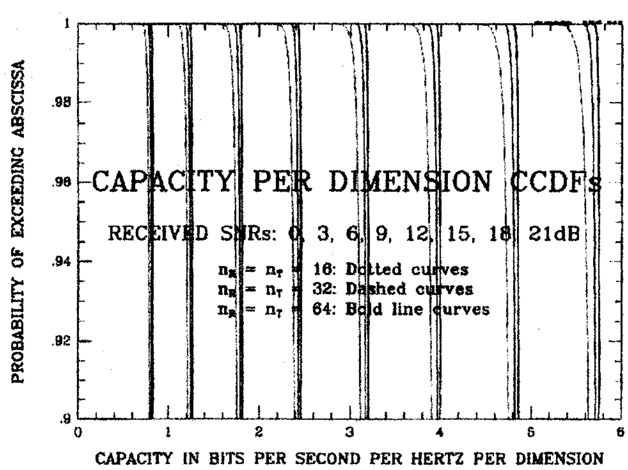
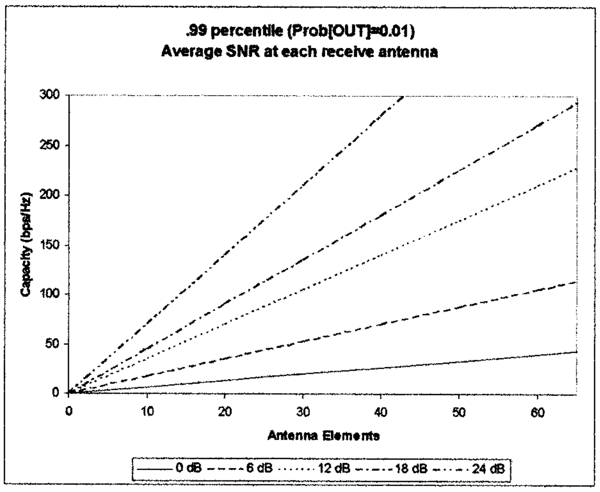
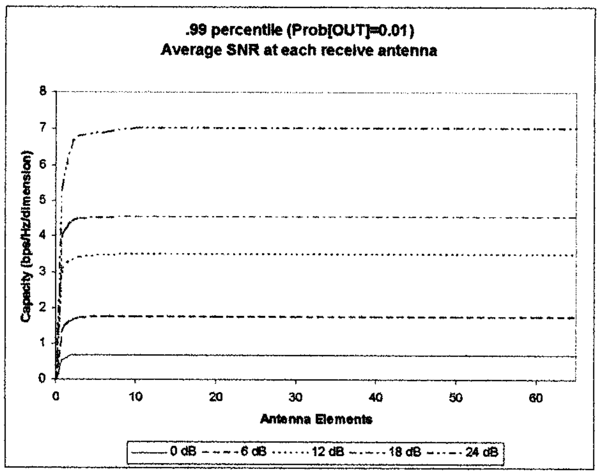


Figure 2.1. Capacity: Complementary Cumulative Distribution Function (in b/s/Hz/dim.).[4]

Figure 2.2 shows the improvement in capacity achieved when the number of antenna elements at both the transmitter and the receiver are increased. As expected from the capacity equation, there is a linear relationship between the capacity and number of transmit/receive antennas.



a)



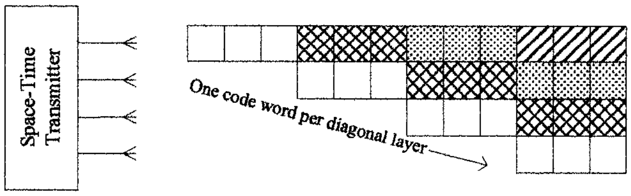
b)

Figure 2.2. a) Capacity in b/s/Hz vs. number of antennas. b) Capacity in b/s/Hz/dimension vs. number of antennas [8].

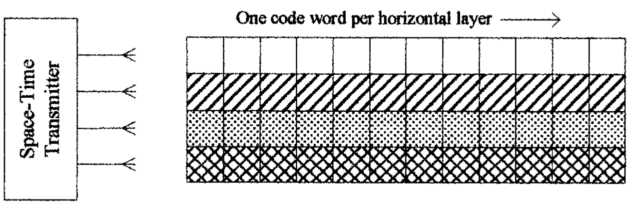
2.3 BLAST Revisited

In [8], Foschini demonstrates how the newfound capacity of MIMO systems can be exploited by using a layered architecture, called BLAST, for Bell Laboratories Layered Space-Time. This architecture, introduced in the previous chapter, uses the same number of transmit and receive antennas and it can achieve the lower bound on capacity values. In BLAST, the transmission is performed by demultiplexing a data stream into n data streams of equal rate. These data streams are encoded, and the encoder on each stream does not need to share any information with any encoder on another stream. At this point there are two ways to proceed: in D-BLAST, the bitstream/antenna association is periodically cycled, whereas in the simpler V-BLAST, each bit stream is assigned an antenna for the duration of the transmission. D-BLAST ensures that no substream is a victim to the worst of the n paths; however it also increases implementation complexities.

Theoretically, maximum-likelihood detection would be the optimum way of recovering the transmitted data at the receiver. But the complexity of such a system increases exponentially with the number of transmit antennas used. Therefore, the detection of BLAST is very similar to that encountered in the multiple access detection world. It is first assumed that the receiver knows the H matrix, due to some training phase that is considered already completed. The main idea behind the detection of BLAST systems is interference nulling. That is to say the interferers that have not yet been subtracted out must be nulled out when detecting each layer.



a)



b)

Figure 2.3. Transmission patterns in layered architecture. a) Diagonal BLAST (D-BLAST). b) Vertical BLAST (V-BLAST).

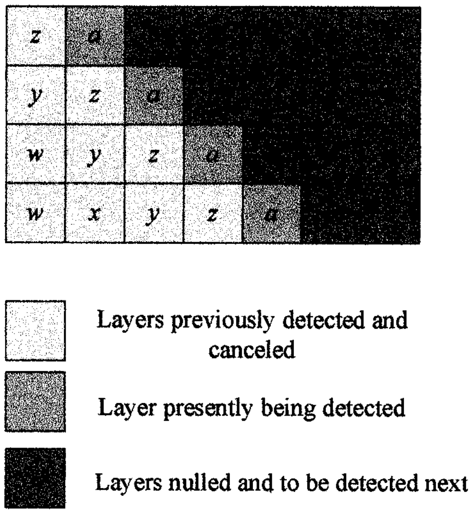


Figure 2.4. BLAST detection using nulling approach.

Figure 2.4 helps to understand the idea of the subtracting and nulling of layers when detecting BLAST. The layers before layer *a* have been previously detected, then reconstructed and subtracted from the received signal, those coming after a are thus nulled and a can be detected.

Using this detection method, it can now be seen through a simple example, as given in [8], that these codes achieve the lower bound on capacity given in Eqn. (2.6). Assuming n=m=4, in decoding the first layer, all three other layers interfere, therefore

$$C = \log_2\left[1 + (\rho / 4) \cdot \chi_2^2\right].$$

The first layer can now be reconstructed and removed, hence stopping it from causing interference on the other layers. The second layer is therefore interfered by two signals, thus

$$C = \log_2\left[1 + (\rho / 4) \cdot \chi_4^2\right].$$

The next layer has 1 interferer as the other two signals have been subtracted out,

$$C = \log_2\left[1 + (\rho / 4) \cdot \chi_6^2\right].$$

For a system that is cycling equally among these four conditions, the capacity would be

$$C = (1/4)\sum_{k=1}^{4} \log_2\left[1 + (\rho / 4) \cdot \chi_{2k}^2\right].$$

The ability to use all six systems simultaneously, as given by the use of multiple antenna elements, yields a capacity of

$$C = \sum_{k=1}^{4} \log_2\left[1 + (\rho / 4) \cdot \chi_{2k}^2\right],$$

which is the minimum outage capacity as provided by Eqn. (2.6).

Nulling and Canceling

Next, the concept of nulling and canceling are explained in-depth as they are of significant importance to this thesis. In [9, 22], the formal algorithm is given as the following.

At time k, in the first decoding step (i=1), let $\widetilde{H}_1 = H$ and $\widetilde{r}_1 = r_k$.

In each step i, the nulling matrix G_i is calculated as the pseudo-inverse of $\widetilde{H}_1$:

$$G_i = \widetilde{H}_i^{+} = (\widetilde{H}_i^{H} \widetilde{H}_i)^{-1} \widetilde{H}_i^{H}, \tag{2.7}$$

where $\widetilde{H}^{H}$ denotes the complex conjugate transpose of $\widetilde{H}$. Each line of G_i can be used to null all but one transmitted signal. Since any layer can be chosen as the first layer to detect, the idea arose that there may be a preferred ordering. It was shown in [22] that it is good to start with the layer showing the biggest post-detection signal-to-noise ratio. In order to do so, the detector chooses the row of G_i with minimum norm and defines the corresponding row as the nulling vector in this step:

$$k_i = \underset{j \in \{1,\ldots,n-i+1\}}{\arg\min} \left\| (G_i)_j \right\|^2, \tag{2.8}$$

$$w_{k_i} = (G_i)_{k_1}^{T}. \tag{2.9}$$

Multiplying w_{k_i} with the vector of the received signal $\widetilde{r}_i$ suppresses all layers but the one transmitted from antenna k_i and the scalar decision value is thus given by,

$$\bar{s}_k^{(k_i)} = w_{k_i}^{T} \widetilde{r}_i \tag{2.10}$$

Next the k_i-th layer is detected within the constellation C:

$$\hat{s}_k^{(k_i)} = \underset{\widetilde{s} \in C}{\arg\min} \left\| \widetilde{s} - \bar{s}_k^{(k_i)} \right\|^2. \tag{2.11}$$

Once one layer has been detected, the detection process can be improved for ensuing layers. By subtracting the part of the detected signal from the vector of received signals, the number of layers to be nulled out in the next step is decreased. The received vector becomes,

$$\tilde{r}_{i+1} = \tilde{r}_i - \hat{s}_k^{(k_i)} (\tilde{H}_i)^{k_i} \tag{2.12}$$

and within the channel, the k_i-th column $(\tilde{H}_i)^{k_i}$ is no longer necessary and thus eliminated,

$$\tilde{H}_{i+1} = \tilde{H}_i^{\bar{k}_i} . \tag{2.13}$$

The process is then restarted until all the layers have been detected, namely when $i=n$. Theoretically, at each step, the number of effective signals to be detected is reduced, while the number of receive antennas remains the same. The diversity level of the resulting system should then increase from layer to layer. However, one of the main impediments to the overall performance of BLAST is error propagation. As shown in Figure 2.5, errors that occur in the early stages of detection propagate into the later stages and negate the effect of the apparent increase in diversity.

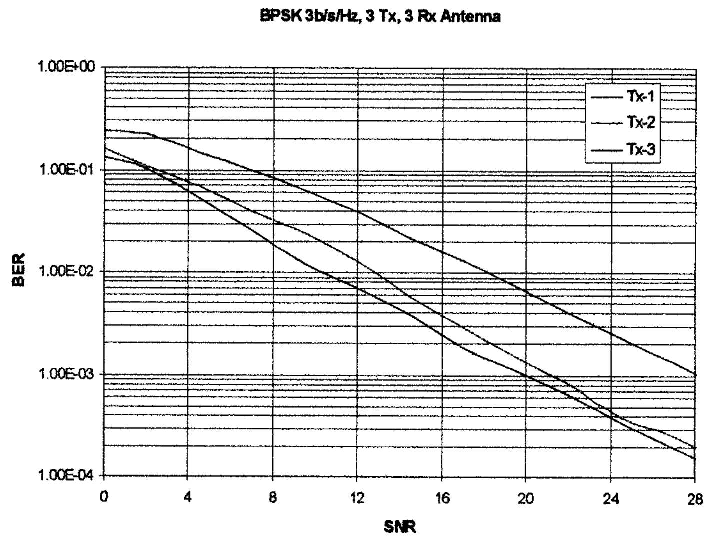


Figure 2.5. BLAST performance with error propagation.

In order to deal with this problem, Baro [23] proposes replacing the zero-forcing (ZF) nulling criterion by a more powerful minimum mean-square error (MMSE) algorithm. In MMSE the detector not only nulls out the interferers, it also takes into account the noise level in the channel. This is a double-edged sword, since the receiver must therefore estimate the SNR. The algorithm is the same as that explained above, save for modifying the cancellation matrix to

$$G_i = (H^H H + \frac{\sigma_n^2}{\sigma_d^2})^{-1} H^H, \tag{2.14}$$

where σ_n^2 / σ_d^2 denotes the signal-to-noise ratio. Figure 2.6 shows the improvement of BLAST detection when using an MMSE algorithm rather than the ZF algorithm. In the mid-range SNR it can be seen that the enhancement is at its premium, since there is an 8 dB gain at BER of 10^{-3} when comparing ZF to MMSE with both using signal ordering.

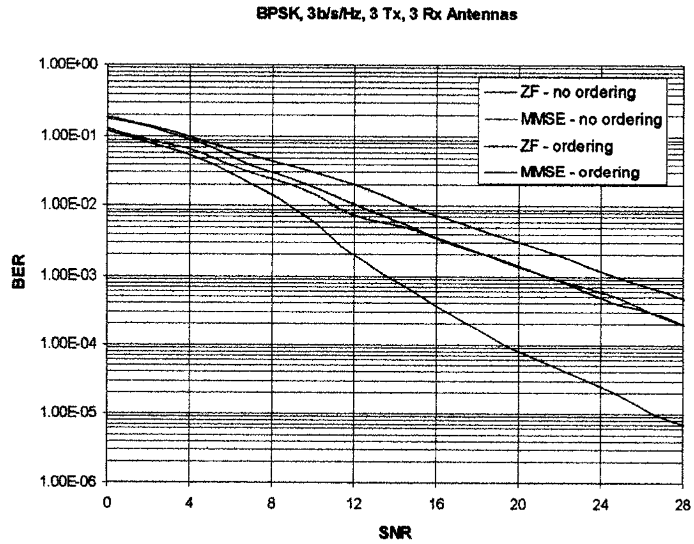


Figure 2.6. Comparison of ZF and MMSE detection, with and without ordering of the signals, n=m=4, QPSK modulation, V-BLAST architecture.

2.4 Simple Transmit Diversity Scheme

As previously mentioned, BLAST employs the capability of the MIMO channel by sending as much information in substreams as possible. This method does not provide for a proper use of the newly acquired spatial diversity. In [24], Alamouti expands on the notion of the classical maximal-ratio receive combining (MRRC) scheme. Figure 2.7 shows the baseband representation of a two branch diversity scheme. It has two transmit and two receive antennas.

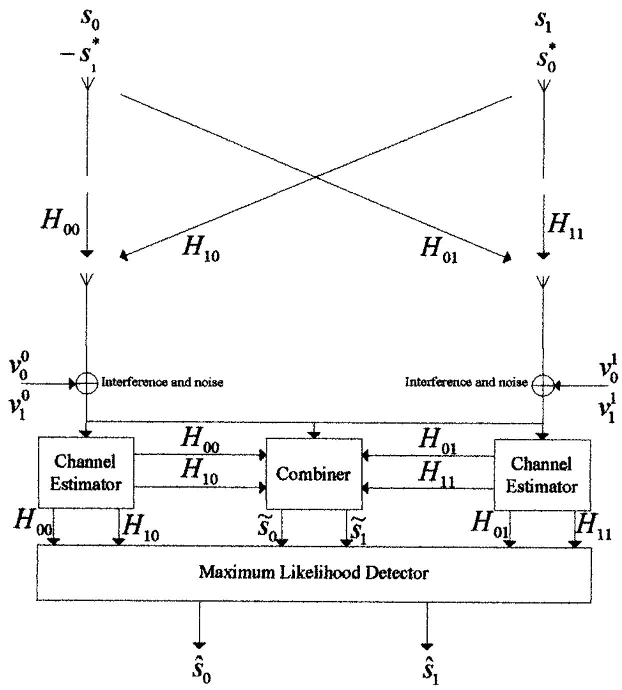


Figure 2.7. Two-branch transmit diversity scheme with two receivers.

The signal transmitted from antenna 0 is denoted by s_0 and from antenna 1 by s_1. During the following time interval, the signals $-s_1^*$ and s_0^* are transmitted from antennas 0 and 1 respectively. The encoding is thus done in space and time.

The received signals are expressed as

$$r_0^0 = H_{00}s_0 + H_{01}s_1 + v_0^0$$

$$r_1^0 = -H_{00}s_1^* + H_{01}s_o^* + v_1^0$$

$$r_0^1 = H_{10}s_0 + H_{11}s_1 + v_0^1$$

$$r_1^1 = -H_{10}s_1^* + H_{11}s_o^* + v_1^1 \tag{2.15}$$

where r_t^i is the receive symbol at antenna i and time t and H_{ij} is the channel matrix element for the subchannel from transmit antenna i to receive antenna j. The combiner in

Figure 2.7 builds the following signals to be used in a maximum likelihood decision rule [24],

$$\tilde{s}_0 = H_{00}^* r_0^0 + H_{01} r_1^{0*} + H_{10}^* r_0^1 + H_{11} r_1^{1*}$$

$$\tilde{s}_1 = H_{01}^* r_0^0 - H_{00} r_1^{0*} + H_{11}^* r_0^1 - H_{10} r_1^{1*}. \qquad (2.16)$$

Assuming PSK signals, the maximum likelihood decision rule for $\hat{s}_j$ is to pick s_i if and only if

$$d^2(\tilde{s}_j, s_i) \le d^2(\tilde{s}_j, s_k), \qquad \forall i \ne k. \qquad (2.17)$$

Alamouti [24] shows that using this scheme of two-transmit, two-receive antennas yields an improvement of about 15 dB at BER of 10^{-3} over a system with no spatial diversity. Also it is shown that with the increase in the number of antennas, comes a higher potential for spatial diversity and so the asymptotic error performance curve's slope becomes steeper.

2.5 Space-Time Code Design Criteria

Space-time codes were developed to take advantage of the diversity provided by increasing the number of transmit and receive antennas. If information is just separated into substreams without specific formation, there is no way to ensure that the spatial diversity will be fully exploited. Just like other forms of diversity, spatial diversity is where replicas of the transmitted signals are provided to the receiver in the form of redundancy; in this case, redundancy in the spatial domain.

In [11], Tarokh provides the designer with the goals and the tools to achieve those goals, in order to take full advantage of the possible spatial diversity offered by MIMO schemes. The method used to achieve these design criteria is by analyzing the probability that codeword $\mathbf{c} = c_1^1 c_1^2 ... c_1^n c_2^1 c_2^2 ... c_2^n ... c_l^1 c_l^2 ... c_l^n$ was sent and that instead codeword $\mathbf{e} = e_1^1 e_1^2 ... e_1^n e_2^1 e_2^2 ... e_2^n ... e_l^1 e_l^2 ... e_l^n$ was erroneously detected from the maximum likelihood detector. The system model is slightly modified from the previous section. In order to ensure the average energy of the constellation is 1, it is contracted by $\sqrt{E_s}$. The following equation expresses the signal observed by antenna j at time t,

$$r_t^j = \sum_{i=1}^{n} H_{ij} s_t^i \sqrt{E_s} + v_t^j \tag{2.18}$$

where $1<t<l$ and $1 < j \le m$.

Assuming that the channel characteristics are known, the probability of transmitting **c** and receiving **e** can be approximated by the standard approximation of the Gaussian tail function

$$P(\mathbf{c} \to \mathbf{e} \mid H_{ij}, i = 1,2,...,n, j = 1,2...,m) \le \exp(-d^2(\mathbf{c},\mathbf{e}) E_s / 4N_o) \tag{2.19}$$

where $N_o / 2$ is the noise variance per dimension and

$$d^2(\mathbf{c},\mathbf{e}) = \sum_{j=1}^{m} \sum_{t=1}^{l} \left| \sum_{i=1}^{n} H_{ij} (c_t^i - e_t^i) \right|^2 . \tag{2.20}$$

For the case of Rayleigh fading, the upper bound on the average probability of error is thus obtained as

$$P(\mathbf{c} \to \mathbf{e}) \le \left(\frac{1}{\prod_{i=1}^{n} (1 + \lambda_i \, {E_s}/{4N_o})} \right)^m \tag{2.21}$$

where λ_i's are the eigenvalues of

$$A_{pq} = \sum_{t=1}^{l} (c_t^p - e_t^p)\overline{(c_t^q - e_t^q)}.$$

It is interesting to note that a square root of the $A(\mathbf{c},\mathbf{e})$ matrix is the codeword difference matrix defined as

$$B(\mathbf{c},\mathbf{e}) = \begin{pmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \cdots & e_l^1 - c_l^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \cdots & e_l^2 - c_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_1^n - c_1^n & e_2^n - c_2^n & \cdots & e_l^n - c_l^n \end{pmatrix}.$$

The average probability of error can be further reduced to [11],

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \left(\prod_{i=1}^{d} \lambda_i\right)^{-m} \left(E_s/4N_o\right)^{-dm} \tag{2.22}$$

where d is the rank of the matrix $A(\mathbf{c},\mathbf{e})$, which is also the rank of $B(\mathbf{c},\mathbf{e})$. The diversity advantage of the system, which is the power of SNR in the denominator of the expression for pairwise error probability, is dm. While the coding advantage, defined as the measure of the gain above an uncoded system operating with the same diversity advantage, is given by $(\lambda_1\lambda_2...\lambda_d)^{1/d}$.

The above pairwise error performance analysis leads to the two following design criteria for quasistatic Rayleigh flat fading STCs as stated in [11]:

-*The Rank Criterion*: In order to achieve the maximum spatial diversity nm, the codeword difference matrix $B(\mathbf{c},\mathbf{e})$ has to be of full rank for any codewords **c** and **e**. If the minimum rank of all the realizations of the matrix $B(\mathbf{c},\mathbf{e})$ is d, then a diversity of dm is achieved.

-*The Determinant Criterion*: If a diversity of dm is targeted, the minimum dth roots of the sum of determinants of all $d \times d$ principal cofactors of $A(\mathbf{c},\mathbf{e})$ taken over the set of all pairs of dissimilar codewords **c** and **e** corresponds to the coding advantage. This quantity must be monitored for all codewords **e** and **c** when designing an STC system, and maximizing it is the objective. If a diversity of nm is targeted, then the minimum of the determinant of $A(\mathbf{c},\mathbf{e})$ taken over the set of all pairs of dissimilar codewords **e** and **c**, must be maximized.

A designer would choose to first optimize the diversity advantage before tackling the coding gain, since the diversity advantage determines the asymptotic slope of the performance curve, while the coding gain merely shifts it left.

The rank criterion remains the same when the fading is possibly dependent. However there is a penalty in the coding advantage. One other noteworthy case is where there is rapid fading, namely where the fading does not remain constant during one transmission block. In this case the system equation is given by

$$r_t^j = \sum_{i=1}^{n} H_{ij}(t) s_t^i \sqrt{E_s} + v_t^j \, . \tag{2.23}$$

This results in a loosening of the limitations of the design criteria. Namely the rank criterion, now named the distance criterion, only requires that codewords **c** and **e** be different for at least w values of $1 \le t \le l$ in order to achieve a diversity of wm. While the determinant criterion, now named the product criterion, requires that the minimum of the products

$$\prod_{t \in \upsilon(\mathbf{c},\mathbf{e})} \left| \mathbf{c}_t - \mathbf{e}_t \right|^2 \tag{2.24}$$

where $\upsilon(\mathbf{c},\mathbf{e})$ denotes the set of time instances where $\mathbf{c}_t \neq \mathbf{e}_t$,taken over dissimilar codewords **c** and **e** must be maximized.

2.6 Binary Rank Criterion

The diversity and coding advantage criteria developed by Tarokh in [11], create a fundamental difficulty, despite facilitating the task of a designer. As they are offered, the criteria apply to the complex domain of baseband modulated signals, rather than to the binary or discrete domain in which codes are traditionally designed. In [19] Hammons and El Gamal present a binary rank criterion which satisfies the more general rank criterion given by Tarokh, but also simplifies the task.

In the rank criterion derived by Tarokh, the sign of the differences between modulated codeword symbols is important. It is difficult to see how that information is retained in the binary domain. The following definition given in [19] helps to mitigate the problem.

Definition: In BPSK modulation, two complex matrices $\mathbf{r}_1$ and $\mathbf{r}_2$ are said to be (-1)-equivalent, if $\mathbf{r}_1$ can be transformed into $\mathbf{r}_2$ by multiplying any number of entries of $\mathbf{r}_1$ by powers of -1.

In BPSK modulation, the discrete alphabet is the field F={0,1} of integers modulo 2. The following theorem, given in [19], is presented without proof and logically leads to the binary rank criterion.

Theorem: The nxl ($l \geq n$) binary matrix $\mathbf{c} = [\bar{c}_1\, \bar{c}_2 \ldots \bar{c}_n]^T$ has full rank n over the binary field F, if and only if every real matrix $\mathbf{r} = [\bar{r}_1\, \bar{r}_2 \ldots \bar{r}_n]^T$ that is (-1)-equivalent to $\mathbf{c}$ has full rank n over the field $\Re$.

The binary design criterion for linear space-time codes, as proposed by Hammons, now follows directly:

Binary Rank Criterion: Let C be a linear nxl space-time code with $l \geq n$. If every nonzero binary codeword $\mathbf{c} \in C$ is a matrix of full rank over the binary field F, then, for BPSK transmission, the space-time code C achieves full spatial diversity.

It is necessary to note that codes that satisfy the binary rank criterion are a subset of codes that satisfy the complex domain baseband rank criterion. This criterion is of utmost importance to this thesis, as it is with it that all future codes are designed. Using the binary rank criterion, one can now easily develop algebraic code designs for which full spatial diversity is effortlessly verified. Linear codes are assumed in the previous theorems; however, one can easily extend to non-linear codes provided the results are not applied to the codewords, but to the modulo 2 differences between the codewords.

Hammons proceeds to develop simple code design rules for STCs to achieve full spatial diversity. The stacking construction is a simple technique that allows for the analysis of the spatial diversity of rigorously studied classical codes. Letting T_1, T_2, ..., T_n be linear vector-space transformations from F^k into F^n, and letting C be the nxl space-time code of dimension consisting of codeword matrices

$$\mathbf{c}(\overline{x}) = \begin{bmatrix} T_1(\overline{x}) \\ T_2(\overline{x}) \\ \vdots \\ T_n(\overline{x}) \end{bmatrix} \tag{2.25}$$

where $\bar{x}$ denotes a k-tuple of information bits and $l \geq n$. Codebook C satisfies the binary rank criterion, thus achieving full spatial diversity n, if and only if T_1, T_2, …, T_n have the property that

$$\forall a_1, a_2, \ldots, a_{n_T} \in \mathrm{F}:$$

$$T = a_1 T_1 \oplus a_2 T_2 \oplus \cdots \oplus a_{n_T} T_{n_T} \text{ is nonsingular, unless}$$

$$a_1 = a_2 = \cdots = a_{n_T} = 0 .$$

2.7 Space-Time Codes

With the rank criterion developed, the design of STCs is made much easier. There have been several types of codes that have been designed to achieve full diversity. Tarokh [11, 14] developed both trellis and block codes that ensure full diversity. The trellis codes developed are superior in that they always achieve full diversity at full rate and they are defines by a set of rules that allow for the design of other codes. On the other hand, the block codes designed don't all have full rank at full diversity. For example, in [11], Tarokh presents STCs with n=2 which achieve diversity advantage by following two design rules:

Design Rule 1: Transitions departing from the same state in the trellis, differ in the second symbol (namely the symbol transmitted via antenna 2).

Design Rule 2: Transitions arriving at the same state differ in the first symbol (namely the symbol transmitted via antenna 1).

These simple rules lead to a whole class of trellis STCs that achieve full diversity. The one thing that links both the trellis codes with the block codes is their decoding algorithm. Since these codes achieve full diversity, at best they will attain full rate; a fact that will be fully explained later. A code with full rate has as an interesting side-effect: it is simple enough to be decoded using a maxim likelihood decoder.

The trellis codes use the branch metric for transitions labeled $q_t^1 q_t^2 \cdots q_t^n$ (where q_t^i represents the symbol sent at time t by transmit antenna i) given by

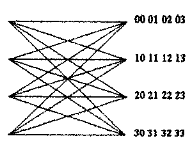
$$\sum_{j=i}^{m} \left| r_t^j - \sum_{i=1}^{n} H_{ij} q_t^i \right|^2 . \tag{2.26}$$

The Viterbi algorithm can then be used to calculate the path with the lowest sum of branch metrics.

On the other hand, the block codes' receiver computes the decision matrix

$$\sum_{t=1}^{l} \sum_{j=i}^{m} \left| r_t^j - \sum_{i=1}^{n} H_{ij} c_t^i \right|^2 \tag{2.27}$$

over all possible codewords $c_1^1 c_1^2 \cdots c_1^n c_2^1 c_2^2 \cdots c_2^n \cdots\cdots c_l^1 c_l^2 \cdots c_l^n$, and decides in favour of the codeword that minimizes it. Figure 2.8 shows a typical trellis representation for trellis STCs and a generator matrix for block STCs, as designed by Tarokh.



(a)

$$G_3 = \begin{pmatrix} x_1 & -x_2 & -x_3 & -x_4 & x_1^* & -x_2^* & -x_3^* & -x_4^* \\ x_2 & x_1 & x_4 & -x_3 & x_2^* & x_1^* & x_4^* & -x_3^* \\ x_3 & -x_4 & x_1 & x_2 & x_3^* & -x_4^* & x_1^* & x_2^* \end{pmatrix} \text{(b)}$$

Figure 2.8. (a) Space-time trellis code with n=2 and full rate, (b) Space-time block code with n=3 and rate ½.

Other codes have been designed using the binary rank criterion proposed by Hammons [19], one such code is the space-time turbo code presented by Su [25]. A simple rate=1/3 turbo code can be written as

$$\mathbf{c} = \begin{bmatrix} X(D) \\ X(D)G(D) \\ \pi^{-1}(\pi(X(D)))G(D) \end{bmatrix}. \tag{2.28}$$

where $X(D)$ is some information bit stream, $G(D)$ is some generator polynomial and $\pi(\cdot)$ is a bit permuter. The third row of the codeword ensures that no information bit stream can create a codeword that does not achieve full rank. The permuter is randomly chosen, but must also be made to satisfy the binary rank criterion.

Several other schemes combining turbo coding with space-time coding have been constructed. For a review of such codes, the reader is encouraged to see [26].

2.8 Adding Diversity to Layering

The early promise of MIMO systems showed the dramatic increase in capacity, and thus potential data rate, afforded in such a system. BLAST was the first method proposed to take advantage of such capacity. However, some shortcomings were present with BLAST, namely error propagation and the lack of employing spatial diversity to improve the error performance. Space-time codes, providing full spatial diversity, were then introduced. The problem with STCs was that while they improved the error performance of codes, they did so while not increasing the data rate of regular SISO codes. Researchers then began trying to combine the two, in order to obtain a code that would increas4e the transmission rate, while still using some of the newly available spatial diversity.

2.8.1 Combined Array Processing and Space-Time Coding

The main drawback in layering of data to be transmitted in MIMO systems is the diversity of the first layer decoded must be maximized in order to reduce the error propagation. In [15], Tarokh developed codes that combined the use of the layering approach with the use of space-time codes. Similar to H-BLAST, layers are not spread out among different antennas. However, each layer is now associated with one of the q groups of n_i antennas, where $\sum_{i=1}^{q} n_i = n$. The input bits are divided into q strings. Each of these strings is encoded using a space-time encoder C_i, the output of which goes through a serial-to-parallel converter providing n_i sequences of symbols simultaneously transferred from the n_i antennas in the j-th group (where $1 \leq j \leq q$). The set of q space-time encoders work in parallel on the same wireless communication channel, and each is received by the same m receive antennas.

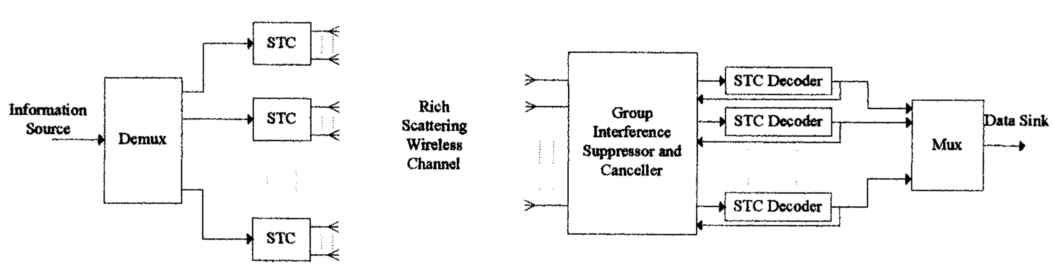


Figure 2.9. Combining layering with space-time codes.

The decoding in this scheme is similar to that used in BLAST. The receiver detects C_1 by using group interference suppression. It is then decoded by its constituent STC decoder. After decoding, the codeword is then cancelled from the over-all received signal, thus resulting in a system with $n - \sum_{j=1}^{i} n_j$ transmit antennas, but still m receive antennas. Assuming the codes are decoded correctly, the diversity gain afforded to each constituent code is $n_i \times (n_i + \sum_{j=1}^{i-1} n_j + m - n)$, which increases as the number of layers is cancelled out. As with regular BLAST, the problem of error propagation is very present in this scheme, and the assumption that the diversity increases with each detecting step is purely theoretical.

2.8.2 Threaded Space-Time Architecture

The problem with the Tarokh scheme is that it does not provide uniform performance from one space-time code decoder to the next. In [17], El Gamal proposes a system that under ideal interference cancellation assumption, can achieve the maximum possible spatial and temporal diversity. In this so-called threaded space-time architecture (TST), the encoding, interleaving and distribution of the symbols for each layer, among different antennas, are made to maximize the use of temporal and spatial diversity for a given rate, if it is assumed there is no interference from other layers. Each layer uses each of the n transmit antennas and each of the l transmission time periods, an equal amount of time. A simple example of such a system is given in Figure 2.10.

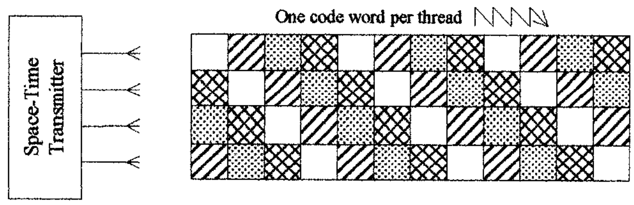


Figure 2.10. Simple TST architecture.

The construction of TST codes is one which allows several degrees of design freedom. One can choose to maximize either the diversity of the system or the throughput, or some combination of the two. In a system with a signaling constellation of size 2^b, L is considered a layer of the code that has a spatial span of n. The binary matrices $M_1, M_2, \ldots, M_n$ with dimension k x bl/n. The binary code C of dimension k is that which contains all the codewords of the form $g(\bar{x}) = \bar{x}M_1 \mid \bar{x}M_2 \mid \cdots \mid \bar{x}M_n$, where $\bar{x}$ is any k-tuple of information bits. The spatial modulator f_L has the property that $\bar{x}M_i$ is transmitted in the l/n symbol intervals of L that have been assigned to antenna i. This layer therefore can achieve a spatial diversity of dm, if and only if d is the largest integer such that $M_1, M_2, \ldots, M_n$ have the property that

$\forall a_1, a_2, \cdots, a_n \in F, a_1 + a_2 + \cdots + a_n = n - d + 1:$

M=[a_1M_1 $a_2M_2 \ldots$ a_nM_n] is of rank k over the binary field.

The maximum rate of transmission for TST systems with component codes achieving d-level transmit spatial diversity, is given as $b(n$-d-1) b/s/Hz.

The decoding of these codes is done by using an iterative multi-user detection type of receiver that employs the minimum-mean square error (MMSE) criterion. This detector will be further explained in the next section.

Chapter 3

Space-Time Layered Block Codes

In this chapter a new code using the advantage of MIMO channels is presented. This new form of coding, labeled Space-Time Layered Block Codes (STLBCs) aims to connect the two extremes that are full diversity space-time codes which maximize the diversity advantage, and BLAST codes, which maximize the throughput advantage. Other codes have been designed to employ both the diversity and multiplexing gains offered by MIMO, namely the so-called threaded space-time architecture presented by El Gamal [17]. The STLBCs presented in this paper provide more freedom of design and can thusly be better used to compare the performance of a wide variety of design objectives. In this chapter the inherent relationship between spatial diversity and rate is analyzed. Next the space-time layered block code design is presented. The sub-optimal non-linear iterative detector and decoders are then presented. Finally the performance of this iterative detector and decoder is analyzed.

3.1 The Trade-off Between Rate and Diversity

Up to this point the innate relationship between rate and diversity has been implied but not explicitly defined. The relationship between the rate of a code and its spatial diversity is inversely proportional. It is because of this that one cannot choose to optimize both diversity advantage and rate. Some codes have been designed to attempt to bridge the region where both diversity and rate are increased to a level superior to that offered by SISO systems. However these codes are either modified versions of pre-existing space-time codes or they impose design limitations. The combined array processing with space-time coding presented by Tarokh [15] requires a number of transmit antennas that is d times the total number of layers, given that each layer is sent as an independent space-time code. The threaded space-time architecture proposed by EL Gamal and Hammons [17] offers a better analysis of codes that are designed to operate between the full diversity and maximum rate. However, the design developed yield a code that is not very flexible in terms of the code parameters. For example, the ratio of bl/n must be a whole number. In other words, in a binary system where b=1, the length of the codeword must be a multiple of the number of transmit antennas used.

In this thesis, codes that have no parameter restrictions, other than those imposed by the rate and spatial diversity trade-off, are studied. Before presenting the codes, the relationship between rate and diversity is analyzed.

In [11], Tarokh a general equation showing the relationship between rate and diversity, namely

$$R \le \frac{\log\lfloor A_{2^{bl}}(n,d) \rfloor}{l} (4.1) \tag{3.1}$$

where $A_{2^{bl}}(n,d)$ is the maximum size of a code length n and minimum Hamming distance d defined over an alphabet of size 2^{bl}. While general and all-encompassing, this relationship does not provide an explicit relationship between the rate R and diversity d. In order to provide a linear equation representing their relationship, the simple case of binary coding, b=1 is used. By making this restriction, the binary rank criterion [19] can be used.

Firstly, the diversity d of codes that do not achieve full diversity must be defined. Binary codebooks are defined as a set of linear binary codeword matrices

$$\gamma = f(\underline{x}) = \begin{bmatrix} T_1(\underline{x}) \\ T_2(\underline{x}) \\ \vdots \\ T_n(\underline{x}) \end{bmatrix} \tag{3.2}$$

where T_j are linear vector-space transformations from F^k to F^n, and γ is an $n \times l$ codeword matrix from codebook Γ.

The binary codebook Γ achieves diversity d if and only if d is the largest integer such that

$$\forall a_1, a_2, \ldots, a_n \in \{0,1\}; a_1 + a_2 + \ldots + a_n \le d$$

$$T = a_1T_1 \oplus a_2T_2 \oplus \cdots \oplus a_nT_n \ne 0 \text{ unless } a_1 = a_2 = \cdots = a_n = 0 . \tag{3.3}$$

The proof of this assertion is given in two parts. First, it is assumed that the codebook Γ achieves rank d. By the binary rank criterion, it therefore also achieves diversity d. It is also assumed that $T = a_1T_1 \oplus a_2T_2 \oplus \cdots \oplus a_nT_n = 0$ for some realization of

$\forall a_1, a_2, ..., a_n \in \mathrm{F}; a_1 + a_2 + ... + a_n \le d$. In this case there exists a nonzero information sequence $\underline{x}_j$ such that

$T(\underline{x}_j) = a_1 T_1(\underline{x}_j) \oplus a_2 T_2(\underline{x}_j) \oplus \cdots \oplus a_n T_n(\underline{x}_j) = 0$.

In other words, there exists a linear combination of $e \le d$ rows of $\gamma(\underline{x}_j)$ that is dependent. By hypothesis, codebook Γ achieves rank *d*, therefore $a_1 = a_2 = \cdots = a_n = 0$. Conversely, if all the realizations of $T = a_1 T_1 \oplus a_2 T_2 \oplus \cdots \oplus a_n T_n \ne 0$, but there is a codeword $\gamma(\underline{x}_j) \in \Gamma$ which does not achieve a rank of *d*, then there exists $a_1, a_2, ..., a_n \in \mathrm{F}$ where $a_1 + a_2 + ... + a_n \ne 0$ for which there is a realization of

$T(\underline{x}_j) = a_1 T_1(\underline{x}_j) \oplus a_2 T_2(\underline{x}_j) \oplus \cdots \oplus a_n T_n(\underline{x}_j) = 0$.

Since *T* is defined as nonsingular, the only possible option is that $\underline{x}_j = \underline{0}$ and $\gamma(\underline{x}_j)$ is the all zero codeword matrix.

With the explicit definition of the diversity of a codeword given, it is now possible to derive the relationship between the rate and the diversity. In order to achieve a diversity advantage of *d*, the codeword matrix must have *d* independent rows. Assuming a direct layering of information bits and parity bits, a system that will be explored further in the remainder of this thesis, the codeword matrix can be visualized as in Figure 3.1.

Clearly rows that are filled completely or partially with parity bits can be made to ensure that no linear combination of other rows produces it. On the other hand, there is no control over rows filled with information bits. For example, in BLAST, the codeword matrix is filled with information bits *i*. It is possible to have codewords where all three

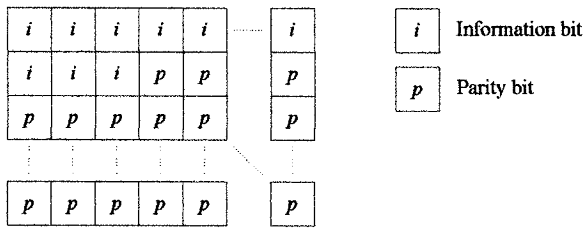


Figure 3.1. Layered codeword matrix.

rows of a codeword are the same, namely a string of bits that repeat every l bits. Therefore the diversity of this code is given as d=1. The rows including parity bits are therefore those that control the rank of the matrix. The rank of the matrix is defined as

$$rank = n + 1 - rows_i, \tag{3.4}$$

where $rows_i$ is the number of rows in the codeword matrix where information bits i are located. As defined previously, the rank of the matrix is also the diversity d. The rate of a codeword is defined as being

$$\text{rate}= R = \frac{k}{l} \tag{3.5}$$

where k is the number of information bits, n is the number of transmit antennas and l is the length of the codeword. The number of information bits lies within the range

$$(rows_i - 1)l < k \leq rows_i l\,. \tag{3.6}$$

Substituting for $rows_i$ in (3.6) produces

$$(n-d)l < k \leq (n-d+1)l\,. \tag{3.7}$$

The number of information bits is obtained from the rank as $k = Rl$, and substituting it in (3.7) yields,

$$n-d < R \leq n-d+1\,, \tag{3.8}$$

the relationship between the diversity and rate of an STLBC.

The trade-off is plotted for n=3 transmit antennas in Figure 3.2.

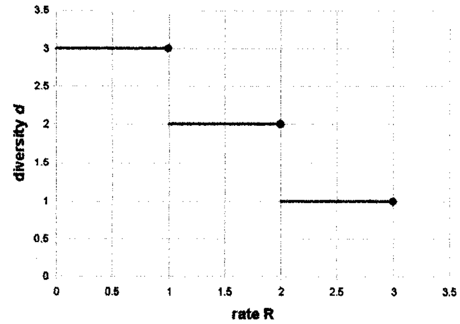


Figure 3.2. Diversity versus rate in a binary space-time layered block code with n=3.

3.2 Space-Time Layered Block Codes System Design

In this thesis a code that has no specific design parameter constraints is proposed. This STLBC also can be used to analyze any possible combination of diversity and rate that is achievable.

The codes presented herein will all be binary and use BPSK modulation. This allows for the use of binary rank criterion to be used to develop proper encoders. The STLBC is defined as a binary encoder, a mapper from the binary field to the {-1,1} field and a serial-to-parallel spatial modulator. The information bits, $\underline{x} = (x_1 x_2 \ldots x_k)$ where $x_j \in \{0,1\}$, are separated into k bits to be encoded by the STLBC encoder. The design of these encoders will be presented later. The output of the encoder is a $1 \times nl$ vector $\underline{\gamma}$ $\gamma_j \in \{0,1\}$, basically a SISO codeword. This codeword is then mapped from the binary field into the {-1,1} field, creating codeword $\underline{c}$ a vector of size $1 \times nl$. The final step in the generation of STLBCs is to convert the 1-dimensional codeword vector into a 2-dimensional codeword matrix. The serial-to-parallel spatial modulator thus converts the $1 \times nl$ codeword vector $\underline{c}$ into an $n \times l$ codeword matrix $\mathbf{s}$, where $s_t^i \in \{-1,1\}$. Note that the subscript j is not used to represent a time index, but rather a codeword index, the transmitted codeword matrix $\mathbf{s}$ uses the time index t as each column of the matrix are transmitted simultaneously. Figure 3.3 shows the block diagram of the STLBC transmitter. Figure 3.4 shows the layout of a simple and typical codeword as it is generated in the transmitter

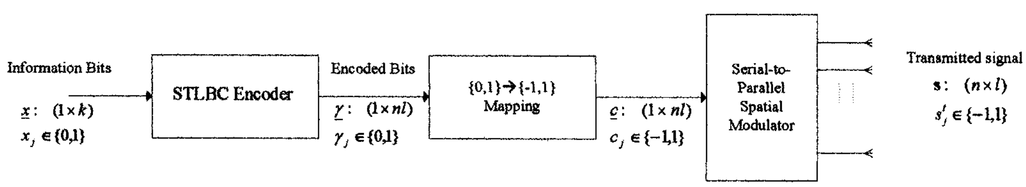


Figure 3.3. Space-time layered block code (STLBC) transmitter block diagram.

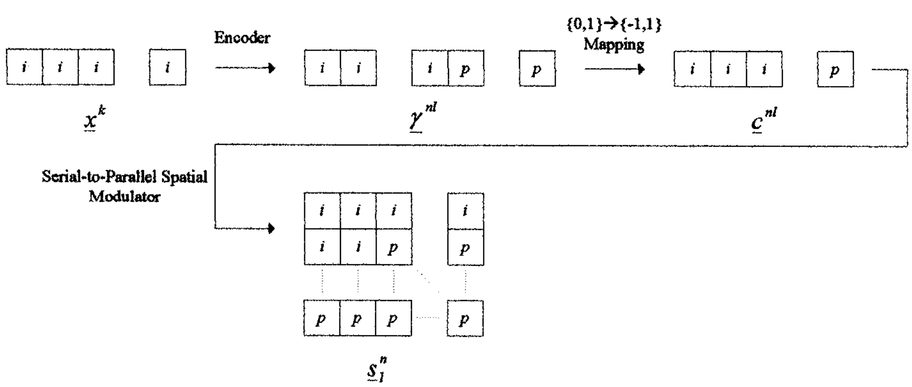


Figure 3.4. Space-Time Layered Block Code (STLBC) encoder structure.

3.3 Detection and Decoding of STLBC

Unlike Tarokh's space-time codes [11], the codes presented in the previous section can achieve greater rate than full rate. Therefore, much like BLAST [8], maximum likelihood detection is impossible as its complexity increases exponentially as the multiplexing advantage is increased. In BLAST, the detection is done using nulling and canceling. However, a much more powerful detector is one where space-time processing is formulated as a joint multiuser detection and decoding problem. Unlike BLAST and TST layering [17], the coding of the previously designed STLBC, is done both temporally and spatially. Therefore the information in each layer is not independent, and all layers must be fully detected before decoding can begin. In other words, it is akin to a multiuser detection problem where the users' data have some dependence. The turbo processing principle [27] can be used efficiently to allow trade-offs between complexity and performance. The block diagram of the iterative receiver is shown in Figure 3.5. In this block diagram, the SISO multiuser detector module provides soft-decision estimates of the n streams of data. The detected streams are multiplexed and decoded in one channel decoder. After each decoding iteration, the soft output from the channel decoder is used to refine the processing performed by the SISO multiuser detector. In this section, the multiuser detector and the channel decoder are examined.

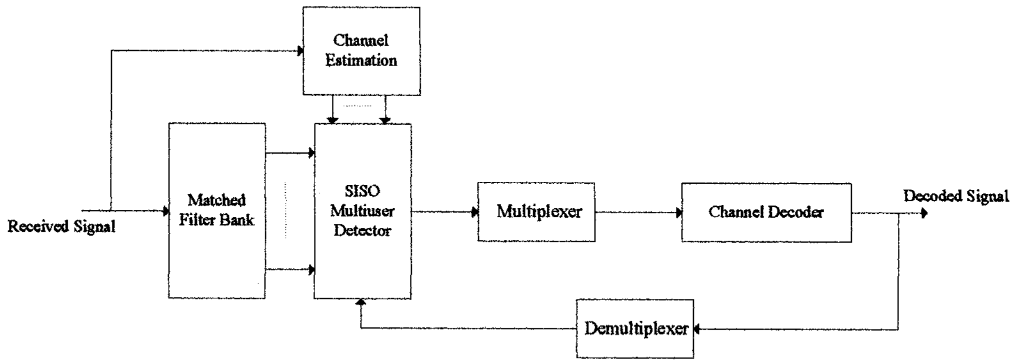


Figure 3.5. Iterative Multiuser detector for layered space-time signals.

3.3.1 The Iterative Minimum Mean Square Error (MMSE) Detector

There have been several SISO multiuser detection algorithms providing a tradeoff between complexity and performance, presented in literature. Some based on the maximum a posteriori (MAP) criterion [28, 29], while others on the MMSE criterion [18, 30]. In this paper, the iterative MMSE receiver, as presented for code-division multiple access (CDMA) systems by El Gamal [18], will be used and modified to take into account the dependence between the layers. In this scheme, the soft outputs of the decoder are used after each iteration to update the a priori probabilities of the transmitted symbols. These probabilities are used to calculate the conditional MMSE filter feed-forward and feedback weights. This detector is similar to that used in BLAST, in that the feedback represents the subtractive interference cancellation part of the receiver, while the feed-forward weights suppress the residual interference.

The received signal can be expressed in vector notation as

$$\underline{r_t} = H_t \underline{s_t} + \underline{v_t} \tag{3.9}$$

where $\underline{r}_t$ is the $m \times 1$ received vector at time t, H_t is the $m \times n$ complex channel matrix, $\underline{s}_t$ is the $n \times 1$ transmitted vector of encoded bits at time t and $\underline{v}_t$ is the white Gaussian noise vector. The estimate of the ith antenna symbol at time t, $y^{(i)}$, is given by (the subscript t is omitted for convenience)

$$y^{(i)} = \underline{w}_f^{(i)^T} \underline{r} + \underline{w}_b^{(i)^T} \underline{\hat{s}}^{(n/i)} \tag{3.10}$$

where $\underline{w}_f^{(i)}$ is the $m \times 1$ optimized feed-forward coefficients vector, $\underline{w}_b^{(i)}$ and $\underline{\hat{s}}^{(n/i)}$ are the $(n-1) \times 1$ vectors of optimized soft feedback weights and hard decision on the (n-1) bits from the other antennas, respectively. The second term of the addition in Eqn. (3.10) only appears through its sum; therefore it can be replaced without repercussions by a single coefficient that represents the sum of the coefficients,

$$w_b^{(i)} = \underline{w}_b^{(i)^T} \underline{\hat{s}}^{(n/i)}. \tag{3.11}$$

The values of the feedback and feed-forward coefficients are obtained through minimizing the mean square error e between the true value of the symbol and its estimate,

$$\begin{aligned} e &= E[(y^{(i)} - s^{(i)})^2] \\ &= E[(\underline{w}_f^{(i)} \underline{r} + \underline{w}_b^{(i)} - s^{(i)})^2] \\ &= E[(\underline{w}_f^{(i)} \{\underline{H}^{(i)} s^{(i)} + H^{(n/i)} \underline{s}^{(n/i)} + \underline{v}\} + w_b^{(i)} - s^{(i)})^2] \end{aligned} \quad , \tag{3.12}$$

where $\underline{H}^i$ is the $m \times 1$ complex channel vector of the ith transmit antenna, namely the ith column of the H matrix, $H^{(n/i)}$ is the $m \times (n-1)$ matrix made up of the complex channel vectors of the other n-1 transmit antennas; and $\underline{s}^{(n/i)}$ is the $(n-1) \times 1$ vector of transmitted data for the other n-1 transmit antennas. The MMSE solutions for the feed-forward and feedback have to satisfy the following conditions as obtained from standard minimization techniques:

$$E[\underline{s}^{(n/i)}]^T H^{(n/i)^T} \underline{w}_f^{(i)} + w_b^{(i)} = 0 \tag{3.13}$$

$$\{\underline{H}^{(i)}\underline{H}^{(i)^T} + H^{(n/i)}E[\underline{s}^{(n/i)}\underline{s}^{(n/i)^T}]H^{(n/i)^T} + E[\underline{v}\,\underline{v}^T]\}\underline{w}_f^{(k)} + H^{(n/i)}E[\underline{s}^{(n/i)}]w_b^{(i)} = \underline{H}^{(i)} \tag{3.14}$$

where

$$E[\underline{v}\,\underline{v}^T] = \sigma_v^2 I_{m\times m} \tag{3.15}$$

$$E[\underline{s}^{(n/i)}] = \underline{\tilde{s}}^{(n/i)} \tag{3.16}$$

and σ_v^2 is the white noise variance, $I_{m\times m}$ is an identity matrix of order m; $\underline{\tilde{s}}^{(n/i)}$ is the $(n-1)\times 1$ vector of the expected values of the transmitted symbols from the other (n-1) antennas.

The a priori probabilities used to evaluate the expectations are obtained from the previous iteration's channel decoder soft output

$$P(s_t^i = 1) = 1 - P(s_t^i = -1) = \frac{e^{\lambda_t^i}}{1+e^{\lambda_t^i}} \tag{3.17}$$

where λ_t^i is the extrinsic information corresponding to the symbol transmitted at time t through antenna i. The results of the detector are thus summarized by the following equations given in [18]

$$A = \underline{H}^{(i)}\underline{H}^{(i)^T} \tag{3.18}$$

$$B = H^{(n/i)}E[\underline{s}^{(n/i)}\underline{s}^{(n/i)^T}]H^{(n/i)^T} \tag{3.19}$$

$$F = H^{(n/i)}\underline{\tilde{s}}^{(n/i)} \tag{3.20}$$

$$R_v = \sigma_v^2 I_{m\times m} = N_0 I_{m\times m} \tag{3.21}$$

and

$$\underline{w}_f^{(i)^T} = \underline{H}^{(k)^H}(A + B + R_v + FF^T)^{-1} \tag{3.22}$$

$$\underline{w}_b^{(i)} = -\underline{w}_f^{(i)^T} F. \tag{3.23}$$

The Correlation Matrix of Expected Values

In a system without encoding over the layers, the assumption that the transmitted bits at time t from different layers are independent holds. In this case,

$$E[\underline{s}^{(n/i)}\underline{s}^{(n/i)^T}] = I_{(n-1)\times(n-1)} - \mathrm{Diag}(\underline{\tilde{s}}^{(n/i)}\underline{\tilde{s}}^{(n/i)^T}) + \underline{\tilde{s}}^{(n/i)}\underline{\tilde{s}}^{(n/i)^T}. \tag{3.24}$$

However, in the design presented in this thesis, coding over the layers is possible. Therefore, the independence assumption does not hold and the expectation of the multiplication of two transmitted bits does not equal the multiplication of the expectation. In this case, the $E[\underline{s}^{(n/i)}\underline{s}^{(n/i)^T}]$ matrix depends on the specific code used to generate the STLBC. Since coding is done over both space and time, the complexity of calculating $E[\underline{s}^{(n/i)}\underline{s}^{(n/i)^T}]$ is of order 2^k, where k represents the information size of the codeword. Using the properties of the expected value, a simple method arises to solve this complexity issue.

Suppose the following quantity is to be calculated $E[c_1c_2]$, where c_1 and c_2 are the {-1,1} mapped version of parity check bits x_1 and x_2, where $x_i \in \{0,1\}$. Assuming x_1 and x_2 are dependent, the expectation can therefore be rewritten as,

$$E[c_1c_2] = E[(2x_1 - 1)(2x_2 - 1)] = 4E[x_1x_2] - 2E[x_1] - 2E[x_2] + 1. \tag{3.25}$$

The parity check bits can be written from their information bits as

$$x_1 = A \oplus B \tag{3.26}$$

$$x_2 = A \oplus C \tag{3.27}$$

The A part denotes the modulo-2 sum of information bits that are common to both parity bits, while B and C denotes the modulo-2 sum of information bits uncommon to the parity bits. Replacing the x's with their component bits into Eqn. (3.25) yields the following,

$$\begin{aligned} 4E[x_1x_2] &= 4E[(A \oplus B)(A \oplus C)] = 4E[A'BC + AB'C'] \\ &= 4[P_{A'}P_BP_C + P_AP_{B'}P_{C'}] \end{aligned} \tag{3.28}$$

$$-2E[x_1] = -2E[A \oplus B] = -2E[AB' + A'B] = -2[P_AP_{B'} + P_{A'}P_B] \tag{3.29}$$

$$-2E[x_2] = -2E[A \oplus C] = -2E[AC' + A'C] = -2[P_AP_{C'} + P_{A'}P_C] \tag{3.30}$$

where A' is the inverse of A, $P_A = P(A = 1)$, $P_{A'} = P(A = 0)$ and $P_A + P_{A'} = 1$. The total sum to solve for the expectation in (3.25) is given by

$$E[c_1c_2] = 4P_BP_C - 2P_B - 2P_C + 1. \tag{3.31}$$

It is interesting to note that the common part A plays no role in determining the expected value. The expectation can be further reduced as follows,

$$E[c_1c_2] = 4P_BP_C - 2P_B - 2P_C + 1 = (1 - 2P_B)(1 - 2P_C). \tag{3.32}$$

In order to continue the analysis of the expected value, it is now assumed, without loss of generality, that C=0. In other words, all of the uncommon terms between the two parity bits occur in the parity check equation for x_1. In this case, $P_C = 0$ and (3.32) becomes

$$E[c_1c_2] = (1 - 2P_B). \tag{3.33}$$

The probability P_B is dependent on the terms that make up B. Letting $B = b_1 \oplus b_2 \oplus ... \oplus b_p$, the solution to the expected value can be solved for the two possible cases, where p is odd and where p is even.

Case 1 (p is even):

In this case, the information bits in B can be separated into two groups, yielding the following,

$$B = D \oplus E = DE' + D'E. \tag{3.34}$$

The value of the probability P_B can be determined from P_D and P_E as:

$$P_B = P_D P_{E'} + P_{D'} P_E = P_D + P_E - 2P_D P_E. \tag{3.35}$$

Therefore the value of $(1 - 2P_B)$ can be solved as a function of only P_D and P_E, namely

$$1 - 2P_B = 1 - 2(P_D + P_E - 2P_D P_E) = (1 - 2P_D)(1 - 2P_E). \tag{3.36}$$

If $p > 2$, then the values of P_D and P_E can be decomposed similarly to that of P_B, until the probabilities used are those of the information bits themselves. This leads to a final solution of

$$1 - 2P_B = (1 - 2P_{b_1})(1 - 2P_{b_2})...(1 - 2P_{b_p}). \tag{3.37}$$

Case 2 (p is odd)

This case is similar to *case* 1, save for the first step. In this step, the random variable B is decomposed as follows:

$$B = D \oplus E \oplus b_p = b_p DE + b_p D'E' + b_p'D'E + b_p'DE'. \tag{3.38}$$

Using the same method as in *case* 1, P_B is determined from P_D, P_E and P_{b_k}:

$$P_B = P_D + P_E + P_{b_p} - 2P_DP_E - 2P_DP_{b_p} - 2P_EP_{b_p} + 4P_DP_EP_{b_p} \tag{3.39}$$

and

$$\begin{aligned} 1-2P_B &= 1-2(P_D + P_E + P_{b_p} - 2P_DP_E - 2P_DP_{b_p} - 2P_EP_{b_p} + 4P_DP_EP_{b_p}) \\ &= (1-2P_D)(1-2P_E)(1-2P_{b_p}) \end{aligned} \tag{3.40}$$

As in *case* 1, the probabilities P_D and P_E can be further deconstructed, until arriving at the information bit probabilities, thus again yielding (3.37).

Lastly, the expected value can be reduced further by noting that

$$E[b_j'] = -E[b_j] = -(2P_{b_j} - 1) = (1-2P_{b_j}). \tag{3.41}$$

Therefore the complexity of the calculation for the expected value $E[c_1c_2]$ is reduced from exponential to linear and the final solution is given by

$$E[c_1c_2] = \prod_{j=1}^{p} E[b_j'] \tag{3.42}$$

where information bits b_j are the p bits not common to the parity check equations for x_1 and x_2, in other words they are not the bits that create the dependence. Knowing the generator matrix of the space-time code, and the expected values of the information bits, the correlation matrix of expected values can be easily filled with low complexity equations.

3.3.2 Decoding of STLBC

As detailed in the previous section, the detection and decoding of the STLBC is done in a turbo-like iterative manner. Therefore the decoder has to be capable of inputting soft-inputs and outputting soft-outputs. The decoder used in this design is based on the Chase algorithm [31] as modified by Pyndiah [32] to produce soft-outputs.

Using maximum likelihood decoding, the decision codeword q is given by

$$q = c^i \text{ if } \left|r - c^i\right|^2 \leq \left|r - c^j\right|^2 \quad \forall j \in [1, 2^k], \qquad j \neq i \tag{3.43}$$

where q, and c are vectors of length $1 \times nl$, c^i is the ith codeword of C and

$$\left|r - c^i\right|^2 = \sum_{j=1}^{nl} (r_j - c_j^i)^2 \tag{3.44}$$

is the squared Euclidean distance between r and c^i. Using an exhaustive search for the optimum codeword q is computationally demanding, so Chase [31] proposed a suboptimal algorithm of low complexity. The main idea behind this algorithm is that at high SNR, the optimal solution satisfying the ML criterion lies in a sphere of radius $d_{\min} - 1$ centered on $y=(y_1,\ldots,y_{nl})$, where $y_j=0.5(1+\text{sgn}(r_j))$. Note that the mapping from $\{0,1\} \rightarrow \{-1,1\}$ is not performed on y. The number of codewords used in (3.43) can be reduced to only those in the sphere of radius $d_{\min} - 1$ centered on y. The set of reviewed codewords can be further reduced by taking only the set of the most probable codewords within the sphere. There are several ways of constructing this set, however the most efficient is as follows. First the positions of the $p = \lfloor d_{\min} / 2 \rfloor$ least reliable bits of y are determined using r. These least reliable bits are obtained using the log-likelihood ratio

$$\text{LLR} = \Lambda(y_j) = \ln\left(\frac{\Pr(s_j = +1 \mid r_j)}{\Pr(s_j = -1 \mid r_j)}\right) = \left(\frac{2}{\sigma^2}\right) r_j, \tag{3.45}$$

namely, those producing the lowest magnitude of the received data, $\left|r_j\right|$. Next, a set of test patterns T^p are formed by constructing all the nl-dimensional binary vectors with a single "1" in the least reliable positions and "0" in all the other positions, two "1"s in the least reliable positions and "0" in all the other positions; the set includes all such codewords up to one having p "1"s in the least reliable positions. The test sequences z^p, where $z_j^p = y_j \oplus t_j^p$, are decoded using an algebraic decoder and the codewords and the codewords c^p are added to the subset. The decision codeword q is then obtained by mapping the set of c^p to {-1,1} and using the minimum Euclidean distance decoder in (3.43). The reliability of this decision codeword must be computed in order to generate the extrinsic value needed to update the a priori probabilities used in the MMSE detector.

The probability of the decision q_j is defined using the LLR of the transmitted symbol s_j,

$$\Lambda(q_j) = \ln\left(\frac{\Pr(s_j = +1 \mid r)}{\Pr(s_j = -1 \mid r)}\right). \tag{3.46}$$

Note that unlike in (3.45), (3.46) takes into account that q can only be one of the 2^k codewords of C. The numerator and denominator can therefore be rewritten as

$$\Pr(s_j = +1 \mid r) = \sum_{c^i \in T_j^{+1}} \Pr(s = c^i \mid r) \tag{3.47}$$

$$\Pr(s_j = -1 \mid r) = \sum_{c^i \in T_j^{-1}} \Pr(s = c^i \mid r) \tag{3.48}$$

where T_j^{+1} is the set of codewords $\{c^i\}$ such that c_j^i=+1 and T_j^{-1} is the set of codewords $\{c^i\}$ such that c_j^i=-1. Using Baye's Rule and denoting the codewords that achieve minimum Euclidean distance to r with either +1 or -1 at j as $c^{+1(j)}$ and $c^{-1(j)}$, the following is obtained,

$$\Lambda(q_j) = \frac{1}{2\sigma^2}\left(\mid r - c^{-1(j)} \mid^2 - \mid r - c^{+1(j)} \mid^2\right) + \ln\left(\frac{\sum_i A_i}{\sum_i B_i}\right), \tag{3.49}$$

where

$$A_i = \exp\left(\frac{|r-c^{+1(j)}|^2 - |r-c^i|^2}{2\sigma^2}\right) \le 1 \text{ with } c^i \in T_j^{+1} \tag{3.50}$$

and

$$B_i = \exp\left(\frac{|r-c^{-1(j)}|^2 - |r-c^i|^2}{2\sigma^2}\right) \le 1 \text{ with } c^i \in T_j^{-1}. \tag{3.51}$$

For high SNR, $\sigma \to 0$, therefore $\sum_i A_i \approx \sum_i B_i \to 1$, hence an approximation for the LLR of decision d_j is given by [32]

$$\Lambda'(q_j) = \frac{1}{2\sigma^2}(|r-c^{-1(j)}|^2 - |r-c^{+1(j)}|^2). \tag{3.52}$$

This equation can be rewritten as

$$\Lambda'(q_j) = \frac{2}{\sigma^2}\left(r_j + \sum_{u=1,u\neq j}^{nl} r_u c_u^{+1(j)} p_u\right) = \Lambda(y_j) + \lambda_j \tag{3.53}$$

where

$$p_u = \begin{cases} 0, & \text{if } c_u^{+1(j)} = c_u^{-1(j)} \\ 1, & \text{if } c_u^{+1(j)} \neq c_u^{-1(j)} \end{cases} \tag{3.54}$$

The LLR $\Lambda'(q_j)$ is thus the soft-output of the decoder while λ_j is the extrinsic value used to update the a priori probabilities of the MMSE detector.

In order to compute the reliability of the decision d_j at the output of the soft-input decoder, two codewords, $c^{+1(j)}$ and $c^{-1(j)}$ are needed to be included in Eqn. (3.52). The decision q obtained from the Chase algorithm is clearly one of the choices, while the second codeword can be labeled as c. The codeword c is basically a competing codeword to q with $c_j \neq q_j$. The soft output can be rewritten from (3.52) as

$$\Lambda'(q_j) = \left(\frac{|r-c|^2 - |r-q|^2}{2\sigma^2} \right) q_j . \tag{3.55}$$

The competing codeword c is found by increasing the value of the least reliable bits p, and hence the number of test patterns, used in the Chase algorithm. Since the decoder's complexity increases exponentially with an increase in p, a practical limit is placed on p. This limit creates the possibility that a competing codeword may not be found by the Chase algorithm. In this situation, Pyndiah proposes that in the case where a competing codeword could not be found, the following be used,

$$\Lambda'(q_j) = \beta \times q_j \text{ with } \beta \geq 0 . \tag{3.56}$$

The values of β are determined to increase as the decoding steps increase; in other words, the reliability on those bits increases with each iteration.

The block diagram for the combined iterative MMSE detector with Pyndiah decoder is given in Figure 3.6. A scaling of $\alpha(j)$, where j represents the iteration step, can be used on the extrinsic information. This is done to limit the effect of the extrinsic information on the detector, in the early iterations, when the extrinsic information has high variance.

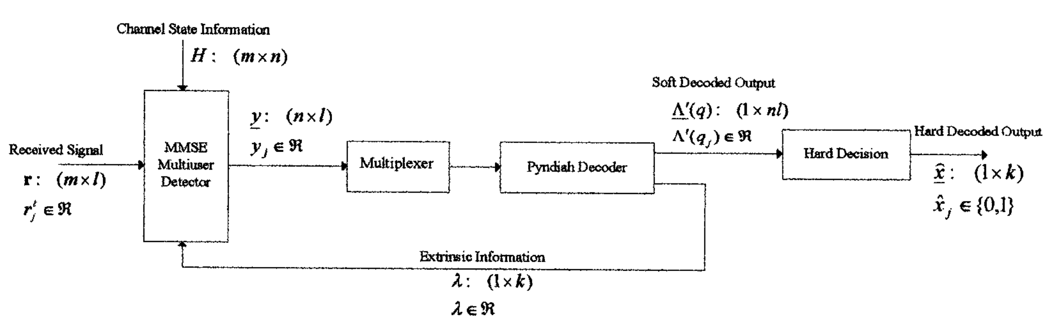


Figure 3.6. The combined iterative MMSE detector and Pyndiah Decoder.

3.4 Performance of Iterative MMSE Detector and Decoder

The design of the iterative MMSE detector provides a lot of leeway with respect to its use. Several questions must be answered in order to optimize its performance when used for STLBCs. For example the expected values used to update the feedback and fee-forward coefficients can be updated layer by layer, or the whole code can first be detected, then the expected values updated. In this section a performance analysis of the iterative MMSE detector coupled with a Pyndiah decoder are analyzed. These results lead to decoder design criteria used to analyze STLBC codes in the next chapter

A Note on Performance Analysis

As is true for all other performance analysis done within this thesis, the results are obtained via Monte Carlo simulation. Blocks of bits are detected and the number of errors tabulated. This mean number of errors is then calculated. In order to ensure that this statistical mean converges to the actual mean of the underlying probability distribution function, the following statistical stopping criteria is used.

The sample mean is given by

$$mean = \bar{e} = \frac{1}{b}\sum_{i=1}^{b} e_i \tag{3.57}$$

where e=1 if there is an error, 0 otherwise and b is the total number of information bits sent. It is known that the value of $\bar{e}$ converges to the underlying distribution's mean value as $b \to \infty$. The optimal simulation value of b is required, where the statistical mean achieves a value that is within a certain allowable estimation error, equivalent to the true mean. Since the sample mean is the sum of independent random variables, the central limit theorem is applicable and the sample mean can be treated as a Gaussian random variable. It is assumed that there are enough sample values, b, so that the sample variance can be taken to be the true variance. In such a case, the interval within which the sample mean lies with probability $1-\alpha$ is given as

$$P(-z_{\alpha/2} < Z \leq z_{\alpha/2} = \int_{-z_{\alpha/2}}^{z_{\alpha/2}} e^{-z^2/2} / \sqrt{s\pi} dz = 1-\alpha \tag{3.58}$$

where Z is a normally distributed random variable with mean 0, variance 1 and defined as

$$Z = \frac{\bar{e}-\mu}{\sigma/\sqrt{b}} \tag{3.59}$$

where μ and σ are the mean and variance of the underlying distribution. Therefore the statistical mean satisfies the following,

$$P(\bar{e} - z_{\alpha/2}\sigma/\sqrt{b} < \mu \leq \bar{e} + z_{\alpha/2}\sigma/\sqrt{b} = 1-\alpha \tag{3.60}$$

Hence, with a probability of $1-\alpha$ the error in the statistical mean with respect to the true mean is less than $z_{\alpha/2}\sigma/\sqrt{b}$. Some values used for $z_{\alpha/2}$ are given below:

$1-\alpha$	$z_{\alpha/2}$
0.95	1.96
0.99	2.576
0.999	3.291

An allowable tolerance of error on the mean is chosen, and a stopping criterion can then be found:

tolerance= $\varepsilon\bar{e}$, where $0 < \varepsilon \leq 1$,

tolerance= $z_{\alpha/2}\sigma/\sqrt{b}$.

Stop when

$$\frac{\bar{e}\sqrt{b}}{\sigma} \geq \frac{z_{\alpha/2}}{\varepsilon}. \tag{3.61}$$

As mentioned previously, enough samples are taken to make the assumption that the true variance is equivalent to the statistical variance, given by

$$\sigma_{stat}^2 = \sum_{i=1}^{b} \frac{(e_i - \bar{e})^2}{b-1}. \tag{3.62}$$

The values used in the simulations producing the results shown in this thesis are $1-\alpha$=0.95, hence $z_{\alpha/2}$=1.96 and ε=0.1. In other words, the value of probability of error is correct within ± 10% 19 times out of 20.

One problem encountered is that at high SNR, the probability of error becomes very small; therefore the running statistical variance is very unstable and never truly approaches the true variance. In this case, the stopping criterion under-estimates the necessary amount of samples and a stopping criterion of $\sum_{i=1}^{b} e_i = 150$ is used.

3.4.1 Results

In this section, an analysis of the iterative MMSE detector is used on layered codes without error control coding. This amounts to studying the iterative MMSE detector on H-BLAST codes. While the iterative detector is designed to operate with an error control code decoder and accept its extrinsic values to update the a priori probabilities, it is useful to study codes without parity in order to analyze its inherent possibilities. Unless stated otherwise, all the error performance results are obtained on a system using n=3 transmit antennas, m=3 receive antennas, with code length l=3. Binary signaling is used and BPSK modulation is assumed. For this chapter and the remainder of the thesis, the SNR is given in dB and is the ratio of the energy per information bit to the noise power (E_b/N_o).

One of the main features of iterative detecting is the updating of the a priori probabilities. In this case, there is no error control coding, therefore no extrinsic value obtained from a decoder. However, the a priori probabilities can still be updated using the output of the detector, to further refine the results. There are two approaches to updating the a priori probabilities: 1)detecting the whole block, then updating the a priori probabilities, and 2)detecting layer by layer and updating the a priori probabilities after each layer.

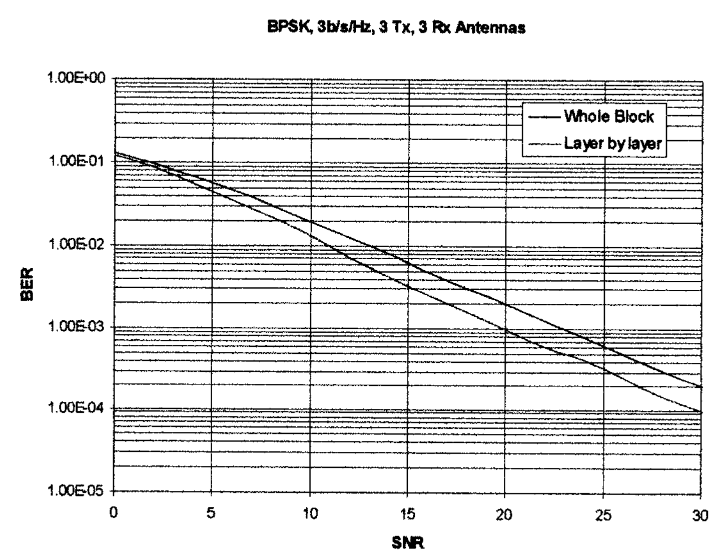


Figure 3.7. Performance updating a priori probabilities at the end of each block vs. at the end of each layer.

In Figure 3.7, an improvement is seen when the a priori probabilities are updated as each layer is detected. This is result stems from the fact that if the detector is to detect the whole block before an update of its a priori probabilities is obtained, there is error propagation. Updating layer by layer reduces the error propagation by ensuring that fewer false assumptions are made by the detector.

One of the main drawbacks of BLAST decoding is the error propagation. In order to improve the results, signal ordering is used. This ensures that the layer with the highest post-detection SNR is detected first. In doing so, the amount of errors to propagate is minimized and the performance of the detector does not degrade with each successive layer. In the iterative MMSE detector design provided by El Gamal [18], signal ordering is not taken into account. If the a priori probabilities were to be updated at the end of each detected block, order of detection would have no effect. However, from the previous discussion, updating the a priori probabilities after each layer is detected provides better results. Therefore the order of layers to decode is an important issue. In BLAST using the zero-forcing algorithm, signal ordering is achieved by detecting the layer, among the remaining undetected layers, which has associated row in the nulling matrix G_i with minimal norm. In the iterative detector, a nulling matrix is not employed, so another means of selecting the layer with strongest post-detection SNR must be developed. Since the nulling matrix G is actually the pseudo-inverse of the channel matrix H, the operation to solve for the signal ordering can be performed on the H matrix. The layer with the highest post-detection SNR is given as that associated with column i of H with highest norm

$$k_j = \arg\max_{i\in\{1,\ldots,m\}} \left\|(H_i)_j\right\|^2 \qquad (3.63)$$

Unlike BLAST, where the order of detected layers is calculated by selecting the required row of the updated G matrix, in this case, the order of all the layers is determined before detection begins on the whole block.

Figure 3.8 shows that error propagation occurs with the iterative MMSE detector when signal ordering is not used, by providing the performance of each layer. Without signal ordering, layer 1 is always detected first, since it may actually have the worst subchannel characteristics, the performance of the entire detector is diminished.

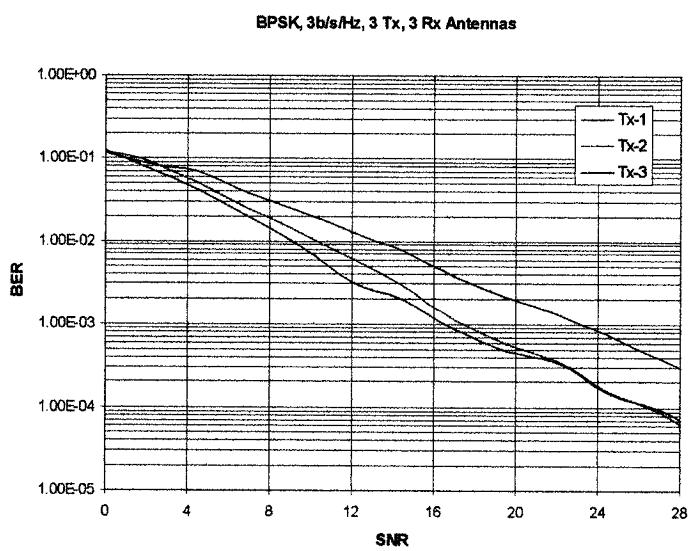


Figure 3.8. Performance without signal ordering.

On the other hand in Figure 3.9, signal ordering is used, and two conclusions can be drawn from it. The first is that by using signal ordering, the performance of all the layers is equalized. This is obvious, since the order of layers is a random variable with uniform distribution; therefore each layer is decoded first, second or third an equal average number of times. The increased diversity that is theoretically expected from being decoded last is thus averaged to all the antennas. The second conclusion reached from Figure 3.9 is that the average performance of the system using signal ordering, is better than the performance of the best layer when no ordering is used.

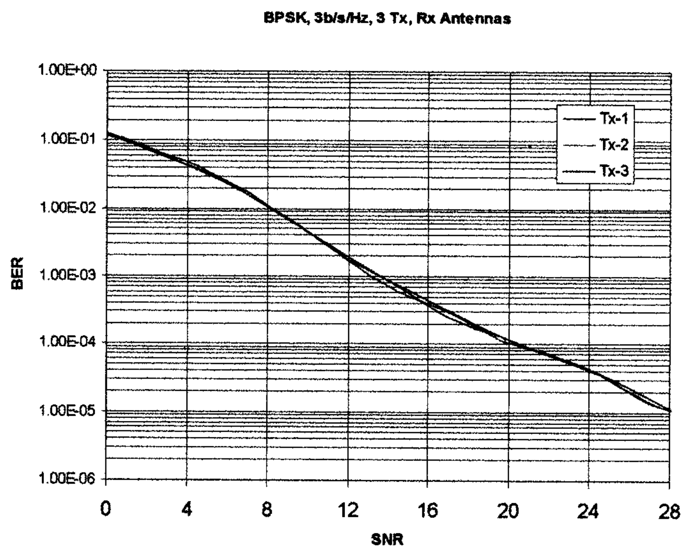


Figure 3.9. Performance using signal ordering.

Since the detector used is an iterative detector, it is necessary to study the improvement on the performance as the number of iterations increases. In Figure 3.10, the performance of the detector using signal ordering and updating the a priori probabilities as each layer is detected, is given at the end of each iteration. In this system, no spatial or temporal coding is used.

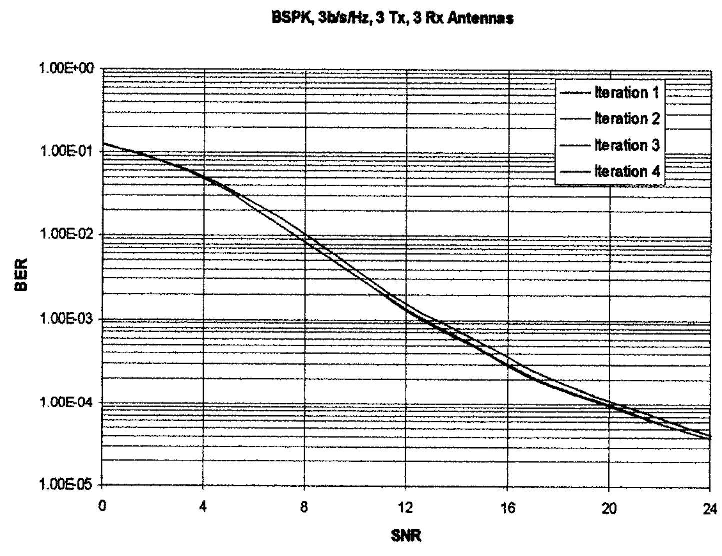


Figure 3.10. Performance after each iteration.

Soft detected data is used in the analysis of the performance for each iteration. After about three iterations it appears the performance converges to its final value.

The iterative MMSE detector can now be compared to the classical BLAST detector using the zero-forcing algorithm. The great probability of error performance advantage of the iterative MMSE detector over the ZF detector is evident from Figure 3.11. The gain is greatest in mid-range SNR, because as the SNR increases, the noise power decreases and the zero-forcing criterion approaches the minimum mean square error criterion.

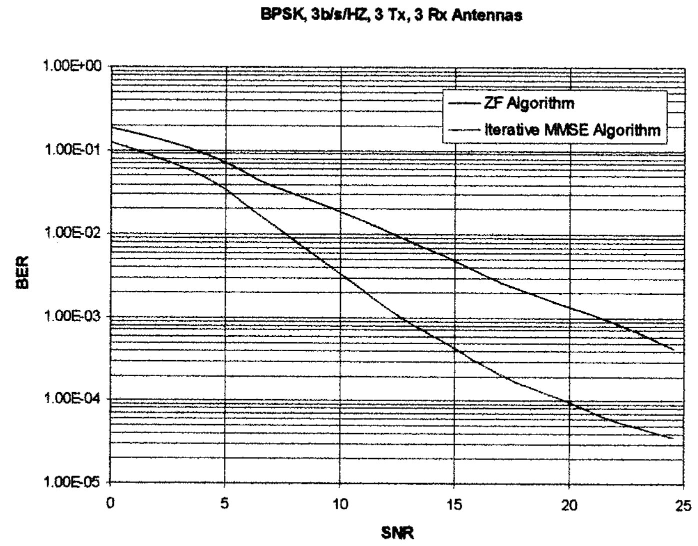


Figure 3.11. Performance for the iterative MMSE algorithm versus the zero-forcing algorithm.

In order to study function of the detector in conjunction with the decoder, a minimal coded scheme is devised. On each layer, a simple Hamming(7,4) code is used. This code, while providing error control coding, is not optimal as it does not explicitly address the issue of diversity. However in the next chapter, code construction guidelines are offered that can provide better results.

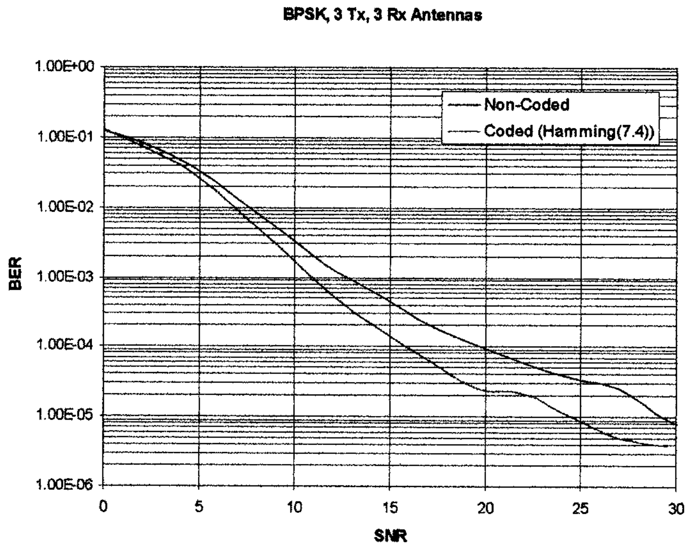


Figure 3.12. Performance of simple coded scheme with non-coded scheme.

By not introducing coding among elements of different layers, the dependence of each layer is negated; thus, the spatial diversity is not better used. However, despite this shortcoming, the results still show an improvement on the probability of error performance. Note that the increase in performance comes at a cost to the data rate. The code using Hamming (7,4) codes has a rate R=12/7=1.71 b/s, compared to a rate R=3 b/s for a non-coded system.

Chapter 4

Improvements on Space-Time Layered Block Codes

The previous chapter dealt with the detecting and decoding of codes that perform at points other than maximum rate or full diversity. These codebooks can be generated randomly for specific rates and achieving required diversity. However, two codebooks operating at the same rate and diversity were not differentiated. In this chapter a criterion for choosing a specific codebook over another is given. The comparison with codes designed using the classical criterion versus those designed using the new criterion is discussed. Next the issue of maximizing the use of spatial diversity on the information data in order to improve performance is discussed. Lastly, the probability of error performance of these codebooks is analyzed.

4.1 A New Criterion for Space-Time Codes

In [11] Tarokh derives the pairwise probability of error of space-time codes. The result from this derivation is a design criterion that states in part that if the minimum rank of all the possible codeword difference matrices is d, then a diversity of dm is achieved.

For a linear code, the minimum rank of the codeword matrices can be used. For the remainder of this thesis, this criterion as given by Tarokh will be termed the *minrank* criterion; that is to say, the criterion that ensures the minimum rank of all codewords in a codebook is a specific value.

The *minrank* criterion imposes restrictions that do not address the average performance of a codebook, but rather the worst case performance. A new criterion that takes into account the average performance of codebooks is obtained. The probability that a codeword matrix **e** is decoded when codeword matrix **c** was sent [11] is rewritten for clarity,

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \left(\prod_{i=1}^{rank} \lambda_i \right)^{-m} \left(E_s \big/ 4N_o \right)^{-rank \times m} . \quad (4.1)$$

The *minrank* criterion thus ensures that all pairwise probability of error achieve a certain require rank. The probability of error of this code can be written as

$$P_e = P(\mathbf{e} \neq \mathbf{c}) \leq \sum_{a=1}^{2^k} \sum_{b=1, b \neq a}^{2^k} \left(\prod_{i=1}^{rank} \lambda(a,b)_i \right)^{-m} \left(E_s \big/ 4N_o \right)^{-rank \times m} \quad (4.2)$$

where $\lambda(a,b)_i$ is the ith eigenvalues of the square of the codeword difference matrix produced from **e**=a and **c**=b, and a and b are the a-th and b-th codewords of the codebook, respectively. Assuming a linear codebook, this equation can be further reduced to

$$P_e \leq \sum_{a=1}^{2^k} \left(\prod_{i=1}^{rank(a)} \lambda(a)_i \right)^{-m} \left(E_s \big/ 4N_o \right)^{-rank(a) \times m} \quad (4.3)$$

where $\lambda(a)_i$ is the ith eigenvalues of the square of the a-th codeword in the codebook. The codewords can each only have finite values of rank $rank(a) \leq n$, therefore using the label *rank*(a) to denote the rank of the a-th codeword matrix in the codebook, the probability of error can be rewritten as

$$P_e \leq \sum_{j=1}^{n} A_j \left(\frac{E_s}{4N_o} \right)^{-jm} \tag{4.4}$$

where

$$A_j = \sum_{a=1; d(a)=j}^{2^k} \left(\prod_{i=1}^{j} \lambda(a)_i \right)^{-m} . \tag{4.5}$$

Note that the value for the probability of error is dominated by the terms for j=1, then by the terms for j=2, and so on. It is clear from this that ensuring no low rank codewords, such as is achieved by the *minrank* criterion, ensures a lower probability of error. However, selecting codebooks based on their *minrank* criterion leads to two problems. The first is that when studying codes with rates greater than that achieving diversity d but less than those capable of achieving d+1, there is no proper way of selecting a good codebook. Secondly, forcing the minimum rank to be a certain value may lower the typical rank levels of the codes in the codebook. This, while excluding the dominant terms from Eqn. (4.4), puts more weight on the terms with the accepted rank, or "second-level" dominant terms.

Therefore, a new criterion is proposed: the *avgrank* criterion. Choices of codebooks are made on the basis of selecting the codebook with the highest expected value of codeword rank. Thus ensuring that in a codebook while there may be codewords contributing to the dominant terms of Eqn. (4.4), the average codeword contributes to less dominant terms. Using the *avgrank* criterion, two codebooks achieving the same *minrank* criterion can hence be analytically compared.

4.2 Maximizing the Use of Spatial Diversity on the Information

In the elementary design of STLBCs presented in the previous chapter, simple 1-dimensional (1-D) systematic block codes are layered to produce codeword matrices of diversity d. These codes, while achieving the required diversity advantage, do not take into account the effect of the potential harm of the channel characteristics. An example codeword matrix generator is given in Figure 4.1, where the i_j's are the information bits.

Time →

Antenna ↓

i_1	i_2	i_3
i_4	$i_3 \oplus i_4$	$i_2 \oplus i_3$
$i_2 \oplus i_3 \oplus i_4$	$i_1 \oplus i_3$	$i_1 \oplus i_2 \oplus i_3$

Figure 4.1. Example of a systematic codeword generator matrix.

Since the 1-D systematic codeword is just layered, it is seen that the information data bits fill spatial layer 1 first, then go on to fill the next layers in order. In the detecting of layered codes it was seen that each layer has its own post-detection SNR, and to minimize error propagation, the layer exhibiting the highest such value was detected first. In 1-D codes using SISO, the effect of the channel can be assumed constant over one codeword and thus all bits are affected similarly. However, unlike in 1-D codewords, in 2-D codeword matrices, the channel characteristics affect bits in a manner dependent on the layer. It is hence possible that in a transmitted codeword matrix, the layers containing the information data have been distorted much more than other layers.

In order to ensure the best performance, it is imperative to ensure the information bits are as spread out onto the most time periods and the most spatial layers. A simple method is to modify the systematic 1-D codeword into one that will provide maximal use of the spatial dimension as well as the temporal one. The end result will be a 1-D codeword that is pseudo-systematic; where the location of the information bits is determined to maximize the use of the spatial diversity as well as the temporal diversity. Maximizing the use of both the temporal and spatial diversity implies that each spatial layer and each time period has close to an equal amount of information bits. Positioning the information bits in the 1-D codeword in a manner that will ensure their diagonal placement in the codeword matrix provides the best use of the spatial and temporal diversity. Figure 4.2 compares the basic layering of the systematic 1-D codeword with the codeword that maximizes the spatial diversity use.

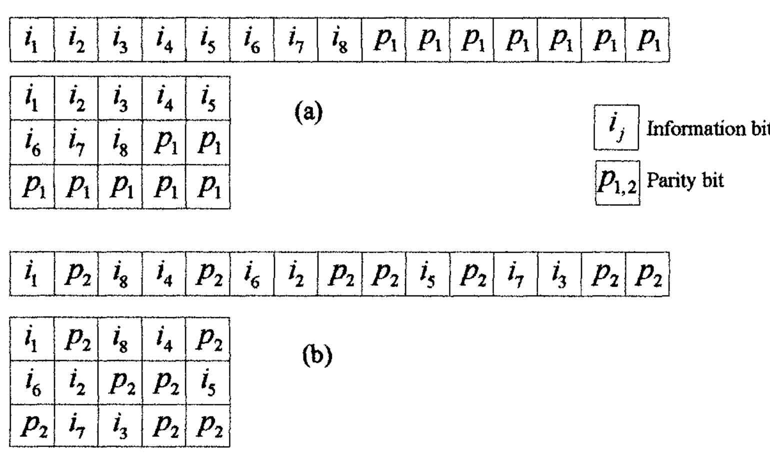


Figure 4.2. (a) Layering of systematic 1-D codeword into codeword matrix. (b) Layering of modified 1-D codeword into codeword matrix.

It must be noted that the parity bits in the modified 1-D codeword are not just the displaced parity bits from the systematic codeword. They satisfy new information bit combinations that ensure the required diversity of the code. In Figure 4.2 (a), the layering of the systematic codeword provides a codeword matrix with information bits using only two spatial layers unequally. In this case layer 1 has five information bits, layer 2 has three information bits and layer 3 has zero information bits. Figure 4.2 (b) shows the adjusted codeword matrix and the information bits now use the three spatial layers almost equally. Layer 1 has three information bits, layer 2 also has three information bits and finally layer 3 has two information bits. Notice that this spatial spreading did not come at the expense of temporal spreading.

The position of the information bits within the codeword matrix is given as follows,

$$\mathrm{Pos}_{2\text{-D}}(i_j) = \{[j - \lfloor \tfrac{j-1}{l} \rfloor - 1]_n + 1 + \lfloor \tfrac{j-1}{l} \rfloor, [j-1]_l + 1\} \tag{4.6}$$

where $0 < j \leq k$ and $[\cdot]_n$ is the modulo n function and $\lfloor \cdot \rfloor$ is the function producing the integer value of the equation.

The information bits are hence located in the 1-D codeword at positions

$$\mathrm{Pos}_{1\text{-D}}(i_j) = l([j - \lfloor \tfrac{j-1}{l} \rfloor - 1]_n + \lfloor \tfrac{j-1}{1} \rfloor) + [j-1]_l + 1, \tag{4.7}$$

where in this case the position function returns a 1-D location.

Looking back at the example codeword matrix given in Figure 4.1, it can be translated to a codeword satisfying the spreading requirements. If the new information bits are designated as i'_j, then the following codeword matrix is obtained:

i'_1	$f(i'_1, i'_2, i'_3, i'_4)$	$f(i'_1, i'_2, i'_3, i'_4)$
i'_4	i'_2	$f(i'_1, i'_2, i'_3, i'_4)$
$f(i'_1, i'_2, i'_3, i'_4)$	$f(i'_1, i'_2, i'_3, i'_4)$	i'_3

Figure 4.3. Pseudo-systematic codeword matrix.

This new codeword matrix can be translated from the one presented in Figure 4.1 by using the following relations,

$i'_1 = i_1 \Rightarrow i_1 = i'_1$

$i'_4 = i_4 \Rightarrow i_4 = i'_4$

$i'_2 = i_3 \oplus i_4 = i_3 \oplus i'_4 \Rightarrow i_3 = i'_2 \oplus i'_4$

$i'_3 = i_1 \oplus i_2 \oplus i_3 = i'_1 \oplus i_2 \oplus i'_2 \oplus i'_4 \Rightarrow i_2 = i'_1 \oplus i'_2 \oplus i'_3 \oplus i'_4$.

From these relationships, new parity bit equations can be calculated, and the translated codeword matrix is illustrated in Figure 4.4.

i'_1	$i'_1 \oplus i'_2 \oplus i'_3 \oplus i'_4$	$i'_2 \oplus i'_4$
i'_4	i'_2	$i'_1 \oplus i'_3$
$i'_1 \oplus i'_3 \oplus i'_4$	$i'_1 \oplus i'_2 \oplus i'_4$	i'_3

Figure 4.4. Translated codeword matrix.

It is interesting to note that the occurrences of each information bit per time period or per spatial layer is increased compared to the codeword matrix drawn from the systematic 1-D codeword. In Figure 4.3, it is seen that layer 1 is lacking i_4, layer 2 is lacking i_1, while time period 3 is lacking i_4. On the other hand, in Figure 4.5, it is seen that the only incomplete dimension is time period 1 lacking i'_2.

4.3 Performance of Pseudo-Systematic STLBC

In this section an analysis on the probability of error of STLBCs is given. The codebooks used for analysis are obtained by an exhaustive search on codebooks achieving the required design characteristics. All schemes analyzed include 3 transmit antennas and 3 receive antennas with BPSK modulation.. The code length is l=n=3, which implies the 1-D block code is of length nl=9. This is the minimal code length required to ensure the possibility of achieving the full diversity potential. Each codebook is defined by a generator matrix and includes $2^k = 2^{Rl}$ codeword matrices.

In order to show the relevance of the *avgrank* criterion, two codebooks of rate R=1.33 b/s/Hz with different *minrank* are compared. The codebooks are defined by their rank spreads, namely the number of codewords achieving each rank, without including the all-zero codeword. For a rate of 1.33 b/s/Hz, k=4 and there are 15 codewords not including the all-zero codeword. The codebooks along with their rank spreads are given below (where d_{min} is the minimum Hamming distance of the codebook):

Codebook 1, *minrank*=2, *avgrank*=2.4, d_{min}=4

Rank:	d=1	d=2	d=3
Number of codewords	0	9	6

Codebook 2, *minrank*=1, *avgrank*=2.533, d_{min}=4

Rank:	d=1	d=2	d=3
Number of codewords	2	3	10

Figure 4.5 provides the probability of error performance for these two codes. Despite the different *minrank*, thus different diversity level of the total code, their performance is very similar at low SNR. However, at higher SNR, codebook 2, with lower *minrank*, starts outperforming codebook 1. From Eqn. (4.4) it is seen that as the signal-to-noise ratio increases, so too does the contribution to the probability of error from low-rank codewords, as compared to the contribution of higher rank codewords. So while codebook 1 achieves a higher *minrank*, it has fewer codewords attaining full rank. The effect of the mid-rank codewords in codebook 1 greatens as the SNR increases, and eventually overshadows the effect of the low-level rank codewords in codebook 2.

Figure 4.6 compares three codebooks with the same d_{min}, but with differing *avgrank*. Again the same result is obtained; however, in this case, the effect of the lower rank codewords is better seen. The codebook with the lowest *avgrank* has several codewords with the lowest rank, and the degradation of performance is seen at lower SNR. As the *avgrank* is increased, the performance degradation at high SNR is reduced. The total effect at BER 10^{-5}, is up to 3dB between the worst case and the best case scenario.

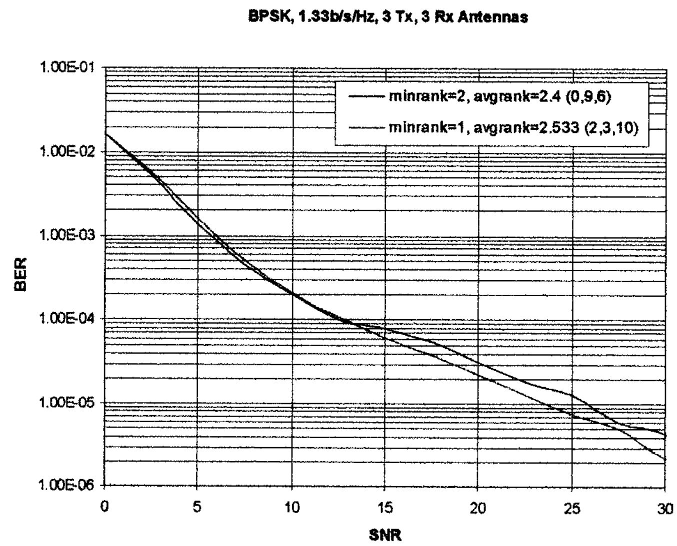


Figure 4.5. Comparison of error performance between codebooks achieving different *minrank*.

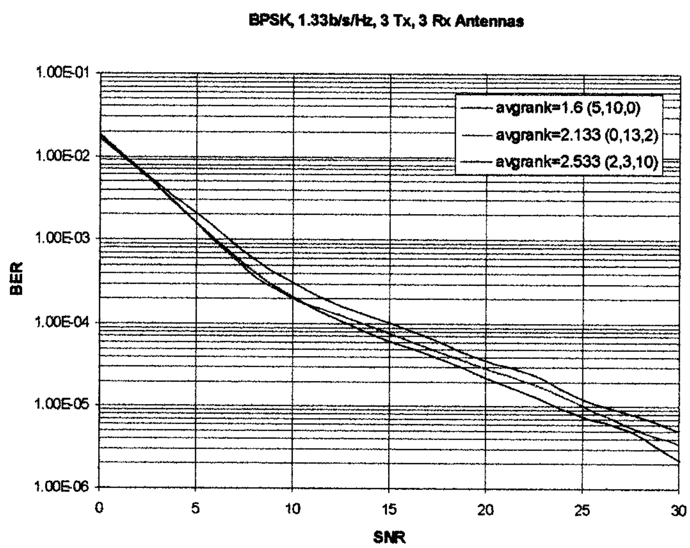


Figure 4.6. Comparison of error performance between codebooks achieving different *avgrank*.

The effect of established coding characteristics is examined next. In classical error control coding theory, one wishes to maximize the minimum Hamming distance

d_{min}. As the d_{min} of a code is increased, its maximum *avgrank* is decreased. This is intuitive, since a higher d_{min} increases the number of "1"'s in the codewords and thus constrains the construction of codewords with higher rank. The effect of d_{min} is studied for two STLBC codebooks with R=1.33 b/s/Hz, similar rank spreads and different d_{min}. The two codebooks are defined as follows:

Codebook 3, *avgrank*=2.6, d_{min}=3

Rank:	d=1	d=2	d=3
Number of codewords	1	4	10

Codebook 4, *avgrank*=2.533, d_{min}=4

Rank:	d=1	d=2	d=3
Number of codewords	2	3	10

Figure 4.7 provides an analysis of the error performance of these codes. It is seen that despite having a slightly higher *avgrank*, the code with the lower d_{min} is greatly outperformed by the code with higher d_{min}.

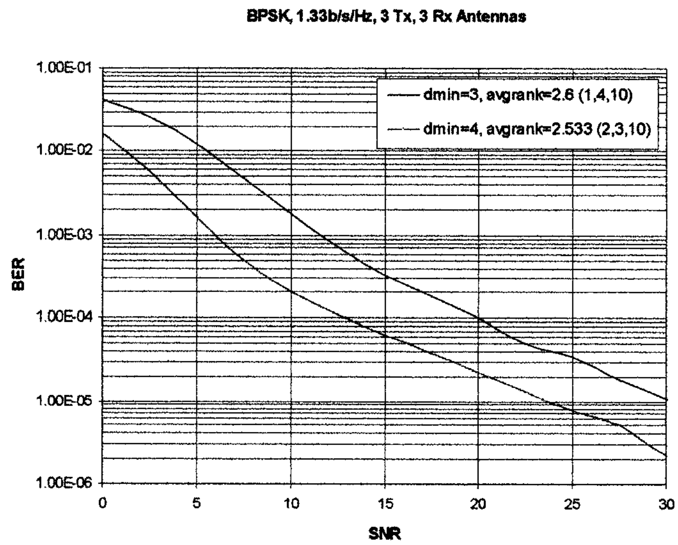


Figure 4.7. Comparison of error performance between two codes with different d_{min}.

This leads to the conclusion that it is imperative that the d_{min} be maximized firstly to ensure better error correcting capabilities. Next the *avgrank* of a codebook can be maximized, to ensure good performance at high SNR.

A STLBC achieving a combination of high rate and good probability of error performance is now given. In a system with $n=l=3$, having $k=7$ is the greatest rate achievable while still providing relevant error control coding capabilities. Any higher code rate results in a $d_{min}=1$. The code, thus achieving a rate of $R=2.33$ b/s/Hz, is compared to a non-coded BLAST code of full rate.

Codebook 5, *avgrank*=2.28, $d_{min}=2$

Rank:	$d=1$	$d=2$	$d=3$
Number of codewords	13	66	48

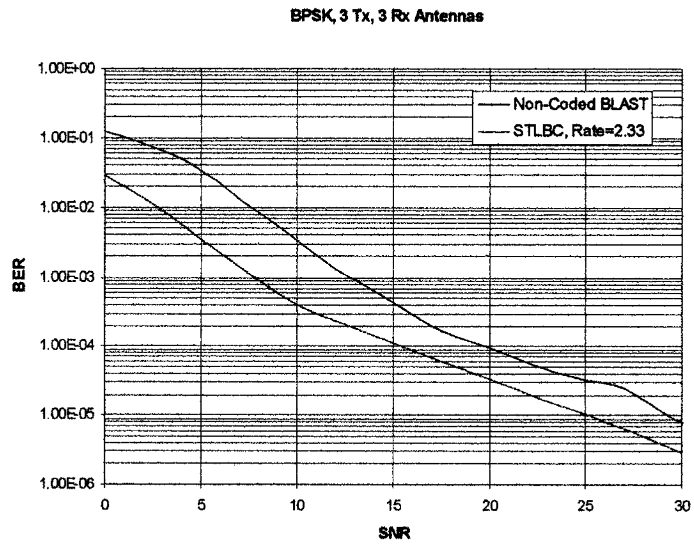


Figure 4.8. Performance of high rate STLBC compared to non-coded BLAST.

Despite being a simple code and only having two parity bits per codeword, the performance of the high rate STLBC is a vast improvement over uncoded BLAST, with low cost of information rate. In Figure 4.8, it is seen that the error control and spatial diversity coding of the STLBC code provides an improvement of about 4dB at BER of 10^{-5}.

Lastly, STLBC employing diversity parity is compared to the simple layered scheme provided in Figure 3.12. For comparative purposes, a codebook with similar rate is proposed. By employing k=5, the rate achieved by the n=l=3 STLBC is 1.66 b/s/Hz, comparable to the 1.71 b/s/Hz offered by using a Hamming (7,4) code on each layer. The STLBC used is presented below:

Codebook 6, *avgrank*=2.42, d_{min}=3

Rank:	d=1	d=2	d=3
Number of codewords	3	12	16

It is seen in Figure 4.9 that the STLBC outperforms the simple layered scheme at low SNR and is comparable at high SNR. The reason for this is because the simple layered scheme uses a larger value of l, namely l=7. The diversity of the code cannot improve, however, the *avgrank* can by having more possibilities of producing higher rank codewords. In fact the *avgrank* of the simple layered scheme is 2.59, while that of the STLBC is 2.42. Another difference between the two codes lies in the decoding. The STLBC code requires the decoding of one (9,5) codeword per iteration, while the layered scheme requires the decoding of three (7,4) codewords per iteration. The complexity of the STLBC is thus lower and offers comparable performance.

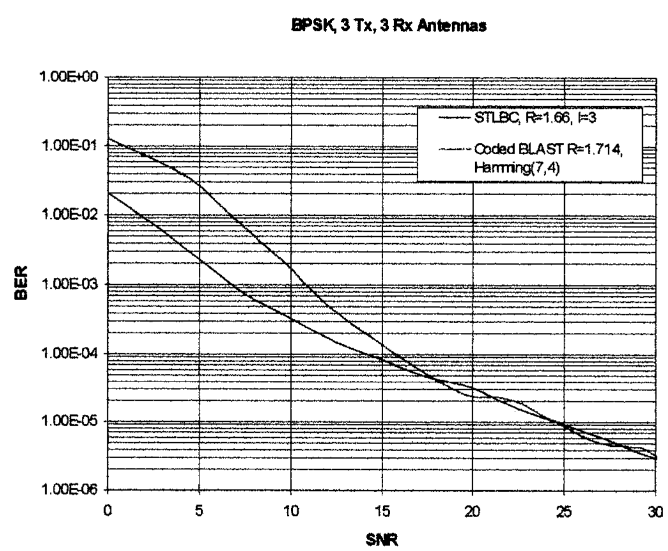


Figure 4.9. Comparison of STLBC with simple layered coded scheme.

Chapter 5

Conclusions and Further Work

Since the early 1990's, wireless telecommunications has advanced by leaps and bounds. Better channel coding, diversity combining and other techniques have improved the performance of wireless communications, which typically operates in harsh signal environment. However, none of these methods make use of the rich scattering channel provided by the wireless medium. Capacity results [4] show that multiple antennas at both the transmitter and the receiver can improve performance without the need to improve costly bandwidth. Recently proposed multiple-input multiple-output (MIMO) schemes can be separated into two groups: those who wish to maximize the data rate, such as BLAST [8], and those wishing to maximize the use of the newly available spatial diversity, such as space-time codes by Tarokh [11].

The first part of this thesis consisted of providing an overview of current results available in literature. First the capacity results for MIMO schemes were given; next methods to employ this increased diversity were provided. BLAST codes are a layered scheme that maximizes the data rate, and require non-linear detection to separate the layers at the receiver. Next space-time codes achieving full diversity were given. The rank criterion as proposed by Tarokh [11] was provided. in order to facilitate the work of code designers, the binary rank criterion [19] was presented. Using this criterion allows for typical code design techniques used in 1-D codes, to be used for space-time codes. lastly, codes that try to combine the increased rate potential with spatial diversity were detailed.

In the second part of the thesis, the trade-off between rate and diversity was derived for binary systems. In order to derive this trade-off, the diversity of codes not achieving full diversity, was defined using the binary rank criterion. At the opposite ends of the trade-off curve lie BLAST (full rate) and space-time codes (full diversity). The trade-off curve provides the inherent relationship between rate and diversity. Next, the design of simple space-time layered block codes (STLBC) was given. These codes are basically a superset of all available space-time technology. When care is taken to ensure no coding between layers is done, they can become simple BLAST codes or simple space-time block codes.

The detection and decoding of these codes using the iterative MMSE detector [18] and the Pyndiah decoder [32] is provided. Since coding between layers is possible in STLBC, the MMSE detector had to be modified. The correlation matrix of expected value was derived for codes which has dependence among layers. Results of the iterative MMSE detector and decoder were then given, showing the improvement over the traditional zero-forcing algorithm used in BLAST.

The last part of the thesis deals with improving the design of STLBC codes. The full probability of error for space-time codes was derived from the pairwise error probability given by Tarokh [11]. Using the total probability of error, a new criterion to ensure proper use of the spatial diversity was obtained. This criterion ensures the maximization of the average rank of codeword matrices within a codebook. Next the STLBC codes were further improved by tweaking the design to ensure the information bits would be spread out over as many spatial and temporal locations as possible. This ensures that the poor performance of the detector over one layer or for select time periods does not affect the over-all quality. The final design provides a code that can improve performance of the system at minimal cost of rate, and is less complex than a similar performing simple coded layered scheme.

There is some possible future work that involves:

- Studying codes with code length l greater than n. While the maximum spatial diversity potential does not change, the actual diversity of each codeword matrix can be increased, thus increasing the average diversity of the codebook.
- A set of code construction rules that ensure the required code parameters are met, without resulting in the need to use an exhaustive search on all the possible codes. This search becomes prohibitively large as the number of transmit antennas n and the code length l increases.
- A more detailed study of the performance of STLBC from an information theoretic point of view. While simulations provide good comparative analysis for codes, theoretical bounds on error performance would ensure optimal schemes are selected as practical schemes.
- Studying the effect of different channel models, such as the block fading channel, on STLBC codes.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

[2] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*, Kluwer Academic / Plenum Publishers, New York, NY, 1999.

[3] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and Decoding: Turbo-codes," *IEEE Transactions on Communications*, vol. 44, pp.1261-1271, Oct. 1996.

[4] G.J. Foschini and M.J. Gans, "On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas," *Wireless Personal Communications*, pp. 6:311-335, 1998.

[5] A. Shah and A.M. Haimovich, "A Performance Comparison of Optimum Combining and Maximal Ratio for Digital Cellular Mobile Radio Communication Systems With Co-channel Interference," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 1454-1463, Jul. 2000.

[6] S.M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451-1458, Oct. 1998.

[7] G.G. Raleigh and J.M. Cioffi, "Spatio-Temporal Coding for Wireless Communications," *IEEE Globecom '96*, vol. 3, pp. 1809-1814, 1996.

[8] G.J. Foschini, "Layered Space-Time Architecture for Wireless Communications in a Fading Channel Environment When Using Multi-Element Antennas," *AT&T, Bell-Labs Technical Journa*l, pp. 41-59, Autumn 1996.

[9] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection Algorithm and Initial Laboratory Results Using the V-BLAST Space-Time Communication Architecture," *Electronics Letters*, vol. 35, pp.14-15, Jan. 1999.

[10] G. J. Foschini, G.D. Golden, R.A. Valenzuela, and P.W.Wolniansky, "Simplified Processing for High Spectral Efficiency Wireless Communication Employing Multi-Element Arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1841-1852, Nov. 1999.

[11] V. Tarokh, N. Seshadri and A.R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communications: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory,* vol. 44, pp.744-765, Mar. 1998.

[12] A.F. Naguib, V. Tarokh, N. Seshadri, and A.R. Calderbank, "A Space-Time Coding Modem for High-Data-Rate Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1459-1478, Oct. 1998.

[13] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 451-460, Mar. 1999.

[14] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Theory of Generalized Orthogonal Designs," *IEEE Transactions on Information Theory,* vol. 45, pp. 1456-1467, July 1999.

[15] V. Tarokh, A. Naguib, N. Seshadri, and A.R. Calderbank, "Combined Array Processing and Space-Time Coding," *IEEE Transactions on Information Theory*, vol. 45, pp. 1121-1128, May. 1999.

[16] D. Bevan and R. Tanner, "Performance Comparison of Space-Time Coding Techniques," *Electronics Letters*, vol. 35, pp. 1707-1708, Sept. 1999.

[17] H. El Gamal and A.R. Hammons Jr., "A New Approach to Layered Space-Time Coding and Signal Processing," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2321-2334, September 2001.

[18] H. El Gamal and E. Geraniotis, "Iterative Multiuser Detection for Coded CDMA Signals in AWGN and Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 30–41, Jan. 2000.

[19] A.R. Hammons Jr. and H. El Gamal, "On the Theory of Space-Time Codes for PSK Modulation," *IEEE Transactions on Information Theory,* vol. 46, no. 2, pp. 524-542, March 2000.

[20] W.C. Jakes, Jr., *Microwave Mobile Communications*, John Wiley and Sons, New York, Chapters 1 and 5, 1974.

[21] Pinsker M.S., *Information and Information Stability of Random Processes*, Holden Bay, San Francisco, Chapter 10, 1964.

[22] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich Scattering Wireless Channel." In *Proc. Int. Symposium on Advanced Radio Technologies*, Boulder, CO, Sept. 10 1998.

[23] S. Baro, G. Bauch, A. Pavlic and A. Semmler, "Improving BLAST Performance Using Space-Time Block Codes and Turbo Decoding," *IEEE Global Telecommunications Conference, (GLOBECOM '00)*,Vol. 2, pp.1067-1071, Nov.27-Dec.1, 2000.

[24] S. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications", *IEEE Journal on Select Areas in Communications*, vol. 16, no. 8, pp.1451-1458, October 1998.

[25] Su, H.J. and Geraniotis, E., "Space-Time Turbo Codes with Full Antenna Diversity", *IEEE Transactions On Communications*, vol. 49, no. 1, pp.47-57, January 2001.

[26] D. Cui, "Turbo Space-Time Coded Modulation Principles and Performance Analysis," PhD dissertation, New Jersey Institute of Technology, pp. 13-16, May 2001.

[27] J. Hagenauer, "Iterative Decoding of Binary Block and Convolutional Codes," *IEEE Transactions on Information Theory*, vol. 47, no.2, pp,429-445, March 1996.

[28] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer, "Iterative Multiuser Detection for DS-CDMA With FEC," in *Proc. Int. Symposium. Turbo Codes and Related Topics*, Brest, France, Sept. 1997, pp.162–165.

[29] M. Moher, "An Iterative Multiuser Decoder for Near-Capacity Communications," *IEEE Transactions on Communications*, vol. 46, pp. 870–880, July 1998.

[30] X.Wang and H. V. Poor, "Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA," *IEEE Transactions on Communications*, vol. 47, pp.1046–1061, July 1999.

[31] D. Chase, "A Class of Algorithms for Decoding Block Codes with Channel Measurement Information," *IEEE Transactions on Information Theory*, vol. IT-18, pp.170–182, January 1972.

[32] R.M. Pyndiah, "Near-Optimum Decoding of Product Codes: Block Turbo Codes," *IEEE Transactions on Communications*, vol. 46. no. 8, pp.1003-1010, August 1998.