Hybrid ARQ Schemes for Non-Orthogonal Space-Time Block Codes

Rui Lin, B.E.(Hons)

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Engineering
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

March 2007

ABSTRACT

Automatic Repeat-reQuest (ARQ) schemes are extensively used in communication systems and computer networks to achieve reliable transmission. Using space-time codes (STCs) with multiple input multiple output (MIMO) or multiple input single output (MISO) systems is an effective way to combat multipath fading, which is the most severe impairment for wireless communication systems. STCs are designed to use the rich scattering multipath environment provided by using multiple transmit antennas. The work done in this thesis focuses on the use of ARQ schemes with non-orthogonal space-time block codes (NOSTBCs) based on Reed Solomon codes.

The truncated-selective ARQ (TS-ARQ) scheme is considered and three novel hybrid ARQ (HARQ) schemes are proposed. Simulation results reveal that, compared to using TS-ARQ with orthogonal space-time block codes (OSTBCs), using NOSTBCs with any of the three proposed HARQ schemes can provide significant gains in terms of dropped packet rate and spectral efficiency at the cost of increased decoding complexity. The performance can be further improved by using the water filling principle to adaptively allocate transmit power among transmit antennas.

ACKNOWLEDGEMENTS

To begin this thesis, I would like to express my thanks to my supervisor Dr.Philippa Martin. This project would not reach this end without her patient guidance, inspiration and encouragement.

Also, I would like to express gratitude to my parents for their support.

Finally, thanks must go to my friends and the members of the comms lab for their friendship and assistance.

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

Wireless network system design has been experiencing rapid development due to the high demand for wireless access to high speed communication networks. This thesis investigates new technologies which may be applicable to next generation wireless networks. These new technologies aim to provide reliable access with high data rates for mobile computer network terminals.

1.1 OVERVIEW

Traditionally, the design of a communications network can be divided into seven parts. Each part corresponds to a layer in the open systems interconnection (OSI) reference model developed by the International Organization for Standardization (ISO). The seven layers of the OSI reference model are the application layer, presentation layer, session layer, transport layer, network layer, data link layer and physical layer [1] as shown in Fig. 1.1. Each layer provides an independent function and can be designed individually without considering other layers. All the layers work together to provide a network service which allows computers connected on the same network to communicate with each other.

Only the design of the lowest two layers for a wireless communication system, the physical layer and the data link layer, is investigated in this thesis. The information bits are organized into frames and packets. A frame contains multiple packets and some control information. A packet is a variable-length block of bits up to some specified maximum size [1]. The packets and frames are often formed at the data link layer

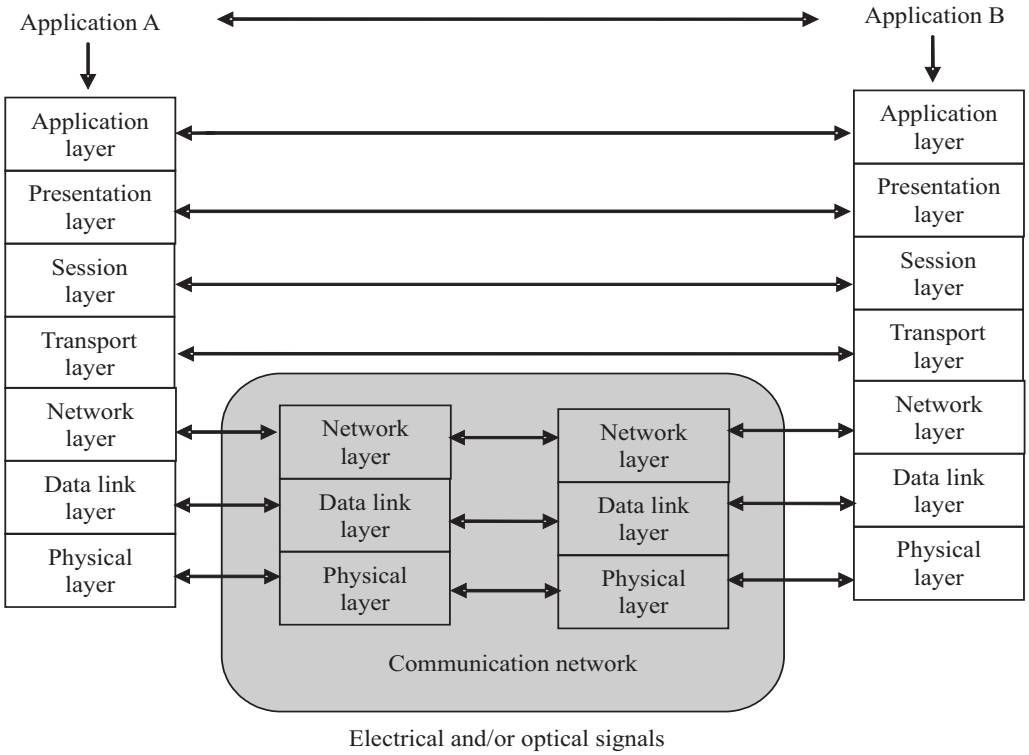

Figure 1.1 The seven-layer OSI reference model [1].

and network layer. The physical layer design consists of several components including coding, modulation and channel estimation. It deals with the transmission of these information frames and packets.

The primary difference between a wireless and a fixed network is the communication channel [6]. Fixed communication networks are connected by using communication cables or optical fibers of which the physical characteristics are almost constant or vary slowly over a relatively long period of time. In contrast, the physical characteristics of wireless radio channels can vary significantly during a transmission. How the radio waves propagate over a wireless channel depends on many factors including the movement of users, and the material and shape of the obstacles between the transmitter and the receiver. The energy level of the received radio waves depends on many other factors including the distance the radio waves have traveled, how much energy is absorbed by obstacles and the phases of the reflected waves when they arrive at the receive antenna. This phenomenon is called multipath fading. All these channel impairments contribute to the time variation of the wireless channel. Therefore, wireless channels

are often described using statistics rather than a deterministic channel model [6]. In order to overcome the channel impairments, a lot of research has been done to improve the design of the physical layer and the data link layer. Recent research has shown that if they are designed jointly rather than individually, then much better performance can be achieved.

In order to achieve nearly error free transmission, Automatic-Repeat-reQuest (ARQ) schemes are normally employed as a key component of the data link layer [4]. This is a practical way to ensure that the information will eventually be received correctly assuming perfect error detection and that an infinitive number of retransmissions is allowed. The ARQ scheme defines how to retransmit information received incorrectly and how to use the received information from the multiple transmissions in the receiver. From the communication system point of view, one of the major drawbacks of using ARQ is that it incurs a loss in terms of the spectral efficiency, which is defined in section 5.1.2, due to the retransmission of the same information. Therefore, minimizing the probability of future retransmission and using as little redundancy as possible are two of the main design goals of ARQ schemes.

Traditional ARQ schemes were designed for communication between a transmitter and receiver, where each has a single antenna. The resulting communication channel is called a single input single output (SISO) channel. A lot of research effort has been put into the joint design of ARQ schemes and the physical layer for a wireless communication system over a SISO channel. These joint designs achieve much better performance, but the amount of gain possible is limited by the channel fading, which is the most severe impairment for wireless communication systems.

Recent advances in physical layer research have shown that an effective way to combat multipath fading is to use multiple antennas at both the transmitter and the receiver. The resulting communication channel is called a multiple input multiple output (MIMO) channel. Theoretical analysis and computer simulations have shown that a MIMO system with specially designed codes can achieve much better performance than a SISO system. The improvement in performance is due to the increased diversity offered by a MIMO system. Multiple antennas provide space diversity if they are sep-

arated far enough. The codes designed for most MIMO systems introduce correlation into the transmitted signal in both space and time domains because, normally, they span multiple time slots and use multiple transmit/receive antennas. Therefore, this is called space-time coding [3]. Many coding schemes have been proposed for MIMO systems including space-time block codes and space-time trellis codes. Now new ARQ schemes need to be designed for these communication schemes.

The performance of an ARQ scheme is normally measured by several parameters including packet or frame error rate and data throughput to the end user. In this thesis, the packet error rate is shown using *dropped packet rate*. Since the practical systems considered only allow a finite number of retransmissions (a finite delay), the retransmission of a packet will be truncated after the number of retransmissions reaches a preset maximum allowed value. If errors still remain, the packet is declared a "Dropped Packet". The *dropped packet rate* can be defined as the ratio of the number of dropped packets to the total number of transmitted packets (including the dropped packets and the packets received correctly).

Throughput is defined as the ratio of the number of accepted information bits for the ARQ scheme to the number of information bits transmitted using a non-ARQ scheme with the same physical layer design during the same amount of time [7] [8]. In this thesis, we want to take into account the constellation size. Therefore, we measure the data throughput performance of the ARQ scheme using the spectral efficiency and throughput spectral efficiency, which are defined in section 5.1.2. They are measured in bits per second per Hertz. The difference in computing them is that throughput spectral efficiency only considers the packets accepted by the receiver, whereas spectral efficiency considers all packets. They are closely related to the average number of transmissions for each packet and the modulation used in the physical layer design. Together the dropped packet rate, spectral efficiency and throughput spectral efficiency quantify the performance of the ARQ schemes under investigation. They are defined in Chapter 5 and are obtained from computer simulations in this thesis.

1.2 BACKGROUND

In this section, modulation and channel fading are introduced. An overall wireless ARQ communication system diagram is presented in Fig. 1.2. The transmitter may contain an error control encoder, space-time encoder and modulator (Fig.1.3). Depending on the number of antennas used at the transmitter and receiver side, the message signals are sent through a SISO or MIMO wireless channel. The received faded signals are further corrupted by additive white Gaussian noise (AWGN). The receiver decodes the received information and the decoded results are checked by the ARQ generator. If errors are detected, a retransmission request will be sent by the ARQ generator to the ARQ controller through a feedback channel.

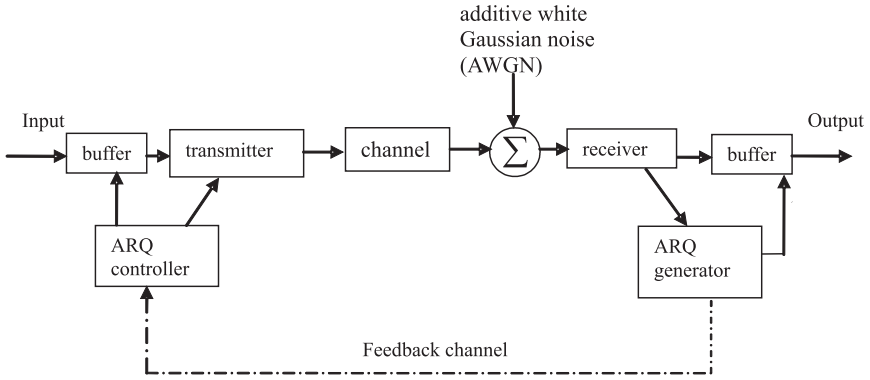

Figure 1.2 Overall system diagram.

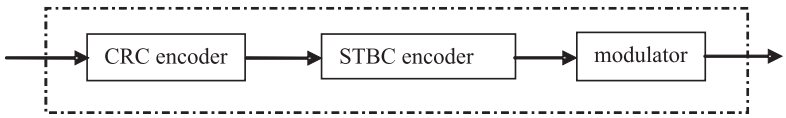

Figure 1.3 Transmitter block diagram.

1.2.1 Modulation and Signal Representation

In digital wireless communication systems, the information bearing signals are modulated using a carrier, which is normally a sinusoid with a fixed frequency [2]. A

sinusoidal wave can be specified by three parameters, namely amplitude, frequency and phase. By manipulating one of these parameters using the message, *amplitude-shift keying(ASK), frequency-shift keying(FSK)* and *phase-shift keying(PSK)* can be obtained, respectively.

A general mathematical representation for a sinusoidal wave is given by

$$s(t) = A\cos(2\pi f_c t + \theta) \tag{1.1}$$

where A is the amplitude and f_c and θ are the frequency and the phase of the carrier, respectively. The carrier signal $s(t)$ in (1.1) can also be described in terms of its in-phase and quadrature components as

$$\begin{aligned} s(t) &= A\cos(2\pi f_c t)\cos(\theta) - A\sin(2\pi f_c t)\sin(\theta) \\ &= s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t) \\ &= Re[\tilde{s}(t)exp(j2\pi f_c t)], \end{aligned} \tag{1.2}$$

where $Re[\cdot]$ is the real part of the expression contained inside the square brackets and $\tilde{s}(t) = s_I(t) + js_Q(t) = A\cos(\theta) + jA\sin(\theta)$ is the complex envelope of the carrier signal $s(t)$. Both $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are passband basis functions. The in-phase and quadrature components are represented as $s_I(t)$ and $s_Q(t)$, respectively. They are both low pass signals. Therefore, $\tilde{s}(t)$ is also a low pass signal [2]. It is a baseband representation of a passband modulated signal. This baseband representation is used throughout this thesis and in the computer simulations. In a real system, after being demodulated, matched filtered and sampled, the baseband representation of the received signal is obtained [9].

The k^{th} block of m bits are transmitted using the modulated signal $s_k(t)$. Its low pass representation is called a $M = (2^m) - ary$ symbol. All $M - ary$ symbols can be represented using a signal-space diagram. This signal-space is defined using two normalized passband basis functions which are orthogonal and have unit energy over

one period. The two normalized passband basis functions are given by [2]

$$\phi_I(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t), \qquad 0 \le t \le T \tag{1.3}$$

$$\phi_Q(t) = \sqrt{\frac{2}{T}}\sin(2\pi f_c t), \qquad 0 \le t \le T \tag{1.4}$$

They are normalized meaning

$$\int_T (\phi_i)^2 dt = 1, \qquad i = \text{I, Q}$$

and are orthogonal to each other meaning

$$\int_T \phi_I \phi_Q dt = 0,$$

where T is the symbol duration which is related to the bit duration T_b. Assuming one symbol represents m bits, then $T = mT_b$.

If the values of both $s_I(t)$ and $s_Q(t)$ are chosen from a set of discrete numbers according to a message, then M-ary *quadrature amplitude modulation (QAM)* can be obtained. In the QAM signal-space diagrams, the k^{th} message point s_k is denoted by $(s_I^k = a_k\sqrt{d_{min}}/2, s_Q^k = b_k\sqrt{d_{min}}/2)$, where d_{min} is the minimum squared Euclidean distance between two message points, a_k and b_k are chosen from an integer set, $\pm 1, \pm 3, \cdots$. It has the following general representation [2]

$$\begin{aligned} s_k(t) &= \sqrt{\tfrac{2E_0}{T}}a_k\cos(2\pi f_c t) - \sqrt{\tfrac{2E_0}{T}}b_k\sin(2\pi f_c t) \\ &= \sqrt{E_0}a_k\phi_I - \sqrt{E_0}b_k\phi_Q \qquad\qquad 0 \le t \le T, \\ &= s_I^k\phi_I - s_Q^k\phi_Q \qquad\qquad k = 0, \pm 1, \pm 2, \cdots \end{aligned} \tag{1.5}$$

where $E_0 = d_{min}/4$ is the energy of the projection of the signal with the lowest amplitude to basis functions. In this thesis, E_0 is assumed to be 1 for ease of analysis. 16-QAM is the constellation used in this thesis and is shown in Fig.1.4. It has 16 symbols and each symbol can be used to modulate up to $\log_2(16) = 4$ bits. Normally, Gray mapping is used to map bits to symbols in order to minimize the bit error rate. The

labels of adjacent symbols only differ in one bit when Gray mapping is used. Normally, for PSK or QAM signal space diagrams, several different Gray mapping schemes exist, among which the binary reflected Gray code [10] has the best performance.

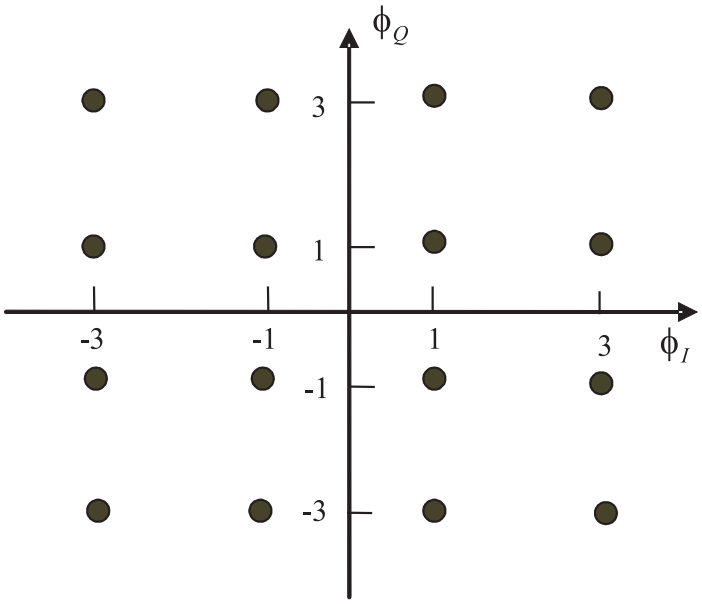

Figure 1.4 Signal space diagram of 16-QAM.

1.2.2 Wireless Channel

In a wireless communication environment, the received signals normally have different amplitude and phase than the transmitted signals. The amplitude fluctuation of the received signals is due to multipath fading and noise. As mentioned previously, multipath fading is caused by many factors, which can be classified into two classes, large-scale propagation effects and small-scale propagation effects [6].

Large-scale propagation effects, which occur over long distances, include path loss and shadowing. On the other hand, the small-scale propagation effects, which occur over very short distances, include the constructive and destructive addition of multipath signal components [6]. The path loss represents the dissipation of transmit power over distance. This loss results in received signals with much lower power. The shadowing is caused by obstacles located on the paths along which the radio waves travel. These obstacles absorb, reflect, scatter and diffract the transmitted radio wave. As a

consequence, the transmitted signal will be further attenuated when it arrives at the receiver. Radio waves are reflected by many reflectors before they arrive at the receive antennas. In an indoor environment, reflectors include furniture, walls, the floor, the ceiling, windows and doors. In an outdoor environment, reflectors include trees, mountains, buildings, moving vehicles and other objects. All the reflected radio waves will be collected by the receive antennas constructively or destructively depending on the phases of the different reflected radio waves [3]. The total effect of all these versions is normally described using statistical models.

The coherence bandwidth is the bandwidth over which the channel response has a constant gain and linear phase response [9]. The channel gain causes fluctuations in the received signal energy and the linear phase response introduces a time delay. If the coherence bandwidth is larger than the bandwidth of the transmitted signal, then the spectral characteristics of the received signal will not be distorted. This is called flat fading or frequency non-selective fading. Otherwise, the fading is frequency selective and the channel will introduce inter-symbol interference (ISI) in the received signals [9]. Only flat fading is considered in this thesis.

Another phenomenon, which is caused by the mobility of objects in a wireless communication environment, is the Doppler spread. The relative movement between the transmitter and receiver and the movement of the surrounding environment will cause frequency shift which is known as Doppler shift. This results in the received signals having a wider bandwidth than the bandwidth of the transmitted signals [6] [9]. This is called Doppler spread. It is defined as the bandwidth caused by the Doppler shift. If the transmitted signal has much wider bandwidth than the Doppler spread, then it is called slow fading. Otherwise, it is called fast fading. The performance of the proposed schemes under both slow and fast fading is studied and simulated in this thesis.

In this thesis, we consider the Rayleigh fading channel model which represents a typical wireless communication environment with no line of sight path between the transmitter and receiver [11]. When the receive antennas can collect a large number of reflected waves, the central limit theorem can be applied. It allows the two quadra-

ture components of the received signal to be treated as uncorrelated Gaussian random processes with zero mean and variance σ_s^2 [3]. The envelope of the received signal, which is defined as $\alpha = \sqrt{s_I(t)^2 + s_Q(t)^2}$ and related to the power of the signal, has a Rayleigh probability distribution and its phase has a uniform distribution between $-\pi$ and $+\pi$ [3]. The probability density function (pdf) is given by [3]

$$p(\alpha) = \begin{cases} \frac{\alpha}{\sigma_s^2} \cdot e^{\frac{-\alpha^2}{2\sigma_s}} & \alpha \geq 0 \\ 0 & \alpha < 0 \end{cases} \tag{1.6}$$

The received baseband signal for a SISO flat Rayleigh fading channel is given by

$$y = a \cdot x + n \tag{1.7}$$

where y is the baseband received signal, a is the channel coefficient generated, x is the transmitted baseband signal and n is the noise signal.

1.3 SCOPE AND CONTRIBUTION OF THIS THESIS

In this thesis, three novel hybrid ARQ (HARQ) schemes are designed and investigated for use with the space-time block codes proposed in [12]. Their performance is simulated over a Rayleigh fading channel. HARQ schemes designed for MIMO systems are designed for a given modulation, error control code and space-time code (STC), as a result it is very difficult to make a direct comparison between different HARQ schemes. The traditional type-I and II HARQ techniques cannot be directly applied to the HARQ schemes proposed in this thesis. Therefore, in this thesis, simulation results are compared to the conventional truncated selective ARQ (TS-ARQ) scheme applied to a STC system as has been done in the existing literature, e.g. [13]. TS-ARQ is considered as it offers the highest throughput out of all the conventional ARQ schemes [7], [8]. OSTBCs are also used for comparison purposes in this thesis. This is because OSTBCs can achieve full diversity gain and have a simple decoder. Therefore, they have been used in other proposed ARQ schemes [14] [15] [16].

ARQ schemes require a feedback channel to send an ACK or NACK signal back to

the transmitter. They are used to positively or negatively acknowledge a correctly or incorrectly received information packet, respectively. If channel information can also be fed back to the transmitter, then the power used for each transmit antenna can be adaptively chosen according to the "water filling" principle [3] to further improve performance. This adaptive modulation scheme is also investigated with the three proposed ARQ schemes and the TS-ARQ scheme.

This thesis is organized as follows. Chapter 2 introduces relevant theory on error control coding such as definitions of groups, fields and cyclic codes. Some traditional ARQ schemes including stop-wait, go-back-n and selective repeat ARQ, which were initially designed for SISO channels, are also introduced in Chapter 2.

Chapter 3 presents a detailed description of MIMO systems. It starts with a general introduction to MIMO systems including their capacity and the water filling principle. Then, the coding design criteria used for MIMO systems are presented. One of the most popular STC families, space-time block codes (STBCs), will be introduced and analyzed in detail. They can be divided into three main categories, orthogonal space-time block codes (OSTBCs), quasi-orthogonal space-time block codes (QOSTBCs) and non orthogonal space-time block codes (NOSTBCs). Other types of STC including space-time trellis codes (STTCs) are also briefly introduced in Chapter 3.

Chapter 4 proposes several hybrid ARQ (HARQ) schemes designed for the NOSTBCs in [12]. They are designed to achieve good performance with high rates. In addition, recent advances in ARQ schemes designed for MIMO systems are presented.

Chapter 5 introduces the simulation environment and presents simulation results for the proposed schemes for both a quasi-static fading channel and a time varying Rayleigh flat fading channel. Various fade rates are considered for the time varying channel. Conclusions and details of possible future work are given in Chapter 6.

The main contributions of this thesis are presented in Chapter 4 and Chapter 5. The main focus of this thesis is on the design of HARQ schemes for the NOSTBCs in [12]. In Chapters 4 and 5, three HARQ schemes, namely HARQ-1, HARQ-2 and HARQ-3, are proposed and their performance is presented and compared. They are designed for MIMO systems using the NOSTBCs of [12], but HARQ-1 and HARQ-2

can easily be used with other STCs including OSTBCs, QOSTBCs and STTCs. The HARQ-3 scheme is designed using the properties of the NOSTBCs in [12] and so cannot be used with other STCs. Simulation results show that a significant coding gain can be achieved using NOSTBCs with HARQ schemes compared to using OSTBCs with TS-ARQ. Adaptively distributing transmit power among all transmit antennas according to the water filling principle is also investigated and simulated for these HARQ schemes. This results in further gains at low signal to noise ratios.

Chapter 2

ERROR CONTROL CODING

This chapter introduces the basic theory of error control coding used in this thesis, particularly focusing on cyclic codes. Firstly, an overview of Galois field (GF) theory is presented. Then an introduction to cyclic codes and Automatic Repeat reQuest (ARQ) schemes is given. These techniques are used extensively in the following chapters. The contents of Sections 2.2, 2.3 and 2.4 are based on [7], [17] and [18].

2.1 INTRODUCTION

Digital communication systems are designed to deliver digitized information from the source to the end user through a channel with an acceptable reliability and at an acceptable data rate. This goal must be achieved with two major constraints: transmit power and bandwidth. For a given application, a certain level of signal to noise ratio (SNR) is required to achieve an acceptable quality and reliability. SNR is defined as the ratio of energy per data bit to noise spectral density (E_b/N_0). Reliability is often measured by bit error rate (BER). For most practical systems, there is a limit on the available transmit power. Since the noise variance at the receiver is fixed, the available signal power may not be high enough to achieve the required quality and reliability. For such cases, error control coding can be used to achieve the desired data quality at lower SNRs. Alternatively, if the signal power is high enough to obtain the desired reliability, then error control coding can be used to reduce the transmit power required to achieve the desired reliability. Error control coding adds redundant information to the information stream. Therefore, the rate of data transmission decreases.

Error control codes can be classified into two major classes[18]:

1. Forward error control (FEC) codes:
   Information only flows in one direction which is from the source to the user. FEC codes include (but are not limited to):

   - Cyclic redundancy check (CRC) codes
   - Hamming codes
   - Bose-Chaudhuri-Hocquenghem (BCH) codes
   - Reed Solomon (RS) codes
   - Convolutional codes

2. Error control for feedback channels:
   Information can be fed back to the transmitter. ARQ schemes are used for error control with feedback channels. Normally, ARQ schemes are used when the transmitted information needs protection to achieve nearly error-free transmission. The use of ARQ schemes, however, reduces the transmission rate due to retransmitting the same information.

2.2 GROUP, RING AND GALOIS FIELD

The mathematical basis of coding theory including groups, rings and Galois fields is introduced in this section.

2.2.1 Group

A set of elements can be called a ***group***, if a binary operation " + " can be defined over any pair of elements contained in the set. The " + " operation must satisfy the following requirements if the set of elements $\mathbf{G}$ is a group:

1. Closure: $a + b = c \in \mathbf{G}$ for all $a, b \in \mathbf{G}$

2. Associativity: $(a+b)+c = a+(b+c)$ for all $a, b, c \in \mathbf{G}$

3. Identity: There exists an element I $\in \mathbf{G}$ such that $a + \mathrm{I} = \mathrm{I} + a = a$ for all $a \in \mathbf{G}$

4. Inverse: for all elements a $\in \mathbf{G}$, a unique element b $\in \mathbf{G}$ can be found to satisfy $a + b = \mathrm{I}$

The group **G** is said to be *commutative*, if it also satisfies

- Commutativity: $a + b = b + a,$ for all a, b $\in \mathbf{G}$

One of the simplest ways to construct a group is to define the operation " + " as modulo m addition, which can be expressed as:

$$a + b \equiv c \text{ modulo m} \tag{2.1}$$

For example, a group constructed under modulo 4 addition consists of elements $\{0, 1, 2, 3\}$. The " + " operation over any pair of elements in this group is completely defined by

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

From the table above, it is clear that the set $\{0, 1, 2, 3\}$ is a group. The identity element is 0. The inverse elements for 0, 1, 2 and 3 are 0, 3, 2 and 1, respectively. It is a commutative group as it satisfies the commutativity constraint. Only *finite* groups are considered in this thesis. They contain a finite number of elements. A group can also be constructed by using modulo m multiplication for the operation " + ", but m must be a prime integer. This is because if m is not a prime integer, there exists pairs of elements in the set which result in " 0 ", which does not have an inverse element to satisfy $a + inv(a) = 1 = I$, where $inv()$ denotes the inverse.

2.2.2 Ring

The definition of a ***ring*** can be obtained by extending the definition of a group to two operations over a set of elements. It is defined as:
A ***ring*** is a set of elements over which two binary operations " $\cdot$ " and " $+$ " can be defined. These two operations must satisfy the following properties:

1. This set of elements forms a commutative group under the operation " $+$ ". The identity element under this operation is labeled " 0 ".
2. The operation " $\cdot$ " is associative for the elements contained in the set, meaning $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in \mathbf{G}$.
3. The operation " $\cdot$ " is distributive over the operation " $+$ " meaning $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

If the operation " $\cdot$ " is commutative, then the ring is called a *commutative ring*. If the ring has an identity element under the operation " $\cdot$ ", it is called a *ring with identity*. The identity element is labeled " 1 ".

2.2.3 Field

A ***field*** is an extension of a ***ring***. Let " $+$ " and " $\cdot$ " be two operations defined over the set of elements $\mathbf{G}$. A set of elements is defined as a ***field*** if and only if the following conditions are satisfied:

1. $\mathbf{G}$ is a commutative group under operation " $+$ ". The identity element under this operation is labeled " 0 " and is called the additive identity.
2. After removing the additive identity element " 0 " from the set, the rest of the elements form a commutative group under operation " $\cdot$ ". The identity element under this operation is labeled " 1 " and is called the multiplicative identity element.
3. Both of the operations are distributive: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

Finite fields were discovered by Evariste Galois, hence, they are called *Galois fields* (GF). They have a finite number of elements. GF(q) is used to denote a Galois field with q elements. Some definitions need to be presented before we describe how to construct GF(q).

1. The number of elements in a Galois field is defined as the ***order*** of the field.

2. The ***order of a nonzero element***, α, in GF(q) is defined as the smallest positive integer k which satisfies $\alpha^k = 1$.

3. The nonzero element whose order is $q-1$ in GF(q) is defined as the ***primitive element***.

A Galois field is defined by its order [7]. The simplest Galois field is GF(2), which can be used to describe the two states of a bit. The two operations are binary addition and multiplication. GF(2) can be represented as $\{0, 1\}$. It only contains the two identity elements, " 0 " and " 1 ", which are the identity elements under operations " + " and " · ", respectively. The two operations are defined by

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

From the above tables, it is clear that the element " 1 " is the primitive element and its order is 2. GF(2) can be constructed by using modulo 2 addition and multiplication. Using modulo m addition and multiplication is the simplest way of constructing a finite field with elements $\{0, 1, 2, 3, \cdots, m-1\}$, where m is a prime number. If m is not a prime number, this method can not be applied for the reason mentioned before. A more complex method is required to construct a finite field which has non prime order.

It can be proven that a finite field must have order p^m, where p is a prime and m is a positive integer [7]. Assuming α is a primitive element of GF(q), it can be shown that the set of all q-1 consecutive powers of α, $\{\alpha^0, \alpha^1, \cdots, \alpha^{q-2}\}$, where $\alpha^0 = \alpha^{q-1} = 1$, consists of all nonzero elements in GF(q). GF(q) can be formed by adding the identity

element under operation " + ", " 0 " , to the above set of elements. Therefore, it follows that GF(q) can be represented as $\{0, 1, \alpha^1, \cdots, \alpha^{q-2}\}$.

In order to complete the introduction to finite fields, we need to define the primitive element, α. First some relevant definitions are presented.

1. The notation GF(p)[x] is used to denote polynomials with coefficients which are elements from GF(p).

2. If a GF(p)[x] polynomial, p(x), cannot be represented as a product of two GF(p)[x] polynomials of lower degree, then p(x) is defined as an irreducible polynomial.

3. If an irreducible polynomial, p(x), is a factor of $x^n - 1$, where n is a positive integer and the smallest value of n is $p^m - 1$, then p(x) is called a ***primitive polynomial***.

The addition or multiplication of two GF(p)[x] polynomials follow the normal rules for polynomials. However, the addition or multiplication of the polynomial coefficients use the " + " and " $\cdot$ " operations of the Galois field GF(p), where the coefficients are defined.

From the definition of the primitive polynomial, it is trivial to find that the roots of a primitive polynomial have order $p^m - 1$. Therefore, any of these roots can be used as the primitive element of GF(p^m). Based on the above discussion, it can be concluded that GF($q = p^m$) can be constructed as $\{0, 1, \alpha, \alpha^2, \cdots, \alpha^{q-2}\}$, where α is a root of a primitive polynomial GF(p)[x] which divides $x^{p^m-1} - 1$.

GF(16) is extensively used in this thesis. We now describe the construction of GF(16) using a primitive polynomial GF(2)[x]. Let α to be a root of the degree 4 primitive polynomial given by

$$g(x) = x^4 + x + 1.$$

Because α is a root of the above polynomial, the element $x = \alpha$ must satisfy $x^4+x+1 = 0$. Since the coefficients are from GF(2), we can write $x^4 = x + 1$. Therefore,

$$
\begin{array}{ll}
0 = 0 & 1 = 1 \\
\alpha = \alpha & \alpha^2 = \alpha^2 \\
\alpha^3 = \alpha^3 & \alpha^4 = \alpha + 1 \\
\alpha^5 = \alpha^2 + \alpha & \alpha^6 = \alpha^3 + \alpha^2 \\
\alpha^7 = \alpha^3 + \alpha + 1 & \alpha^8 = \alpha^2 + 1 \\
\alpha^9 = \alpha^3 + \alpha & \alpha^{10} = \alpha^2 + \alpha + 1 \\
\alpha^{11} = \alpha^3 + \alpha^2 + \alpha & \alpha^{12} = \alpha^3 + \alpha^2 + \alpha + 1 \\
\alpha^{13} = \alpha^3 + \alpha^2 + 1 & \alpha^{14} = \alpha^3 + 1
\end{array}
$$

These symbols are labeled 0 to 15 for 0 to α^{14}, respectively. Representing the elements of GF(16) as consecutive powers of α means the " $\cdot$ " operation can be calculated using modulo 15 addition for the powers of α. Representing the elements as polynomials means the " $+$ " operation can be calculated using modulo 2 addition over the coefficients of the polynomials involved. Two tables which fully define the operations on GF(16) can be found in Appendix A.

2.3 CYCLIC CODES

Cyclic codes are an attractive family of error control codes. They are **linear block codes**. Their encoders and decoders can be implemented using high speed shift register based circuits.

We now introduce some general definitions for linear block codes. A block code consists of a set of codewords generated by mapping k symbols from a data stream of GF(q) symbols to a codeword of n ($n > k$) GF(q) symbols. A block code generated in this way is called a q-ary (n, k) code. Adding $n - k$ redundant symbols to the original k-symbol block means not all combinations of n GF(q) symbols are valid codewords. This enables the receiver to detect or correct errors in the received signals. The q-ary (n, k) code consists of $t = q^k$ codewords, $\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_t$. The **rate** of the block code

is k/n. Any codeword, $\mathbf{C}_m$, in the q-ary (n, k) linear block code can be represented by

$$\mathbf{C}_m = a_1\mathbf{C}_1 + a_2\mathbf{C}_2 + \cdots + a_t\mathbf{C}_t,$$

where $a_k, k = 1, 2, \cdots, t$, are symbols in GF(q).

2.3.1 Theory

Cyclic codes are linear block codes. If each codeword $\mathbf{C} = (c_0, c_1, c_2, \cdots, c_{n-1})$ and its right cyclic shifted version $\mathbf{C}^1 = (c_{n-1}, c_0, \cdots, c_{n-3}, c_{n-2})$ are both codewords in an (n,k) linear block code, then it is called a **cyclic code**.

In order to see the underlining mathematical structure of the cyclic code, it is convenient to represent each cyclic codeword $\mathbf{C} = (c_0, c_1, \cdots, c_{n-1})$ using a polynomial

$$C(x) = c_0 + c_1x + c_2x^2 + \cdots + c_{n-1}x^{n-1}.$$

All code polynomials of a q-ary (n,k) cyclic code are within a ring of polynomials generated by defining the " $\cdot$ " operation of the ring as modulo $(x^n - 1)$ polynomial multiplication which is denoted as $\mathrm{GF}(q)[x]/(x^n - 1)$.

If C(x) is an arbitrary code polynomial in this cyclic code, it can be represented as $\mathbf{C}(x) = c_0 + c_1x + \cdots + c_{n-1}x^{n-1}$, where c_t for $t = 0, 1, \cdots, n-1$ are GF(q) symbols. It can be shown that the code polynomial generated by m right cyclic shifts can be represented as $C^m(x) = x^mC(x)$. For example, by using modulo (x^n-1) multiplication, right cyclic shifting a codeword by 1 position can be written as

$$\begin{aligned} x \cdot C(x) &\equiv (c_0x + c_1x^2 + \cdots + c_{n-1}x^n) \bmod (x^n - 1) \\ &\equiv (c_{n-1} + c_0x + \cdots + c_{n-2}x^{n-1}) \bmod (x^n - 1) \\ &\equiv C^1(x) \bmod (x^n - 1) \end{aligned}$$

Therefore, the multiplication of any two code polynomials of a cyclic code can be viewed as a summation of many right cyclic shifted versions of this codeword multiplied by corresponding coefficients c_t. The result of this summation must be a valid codeword

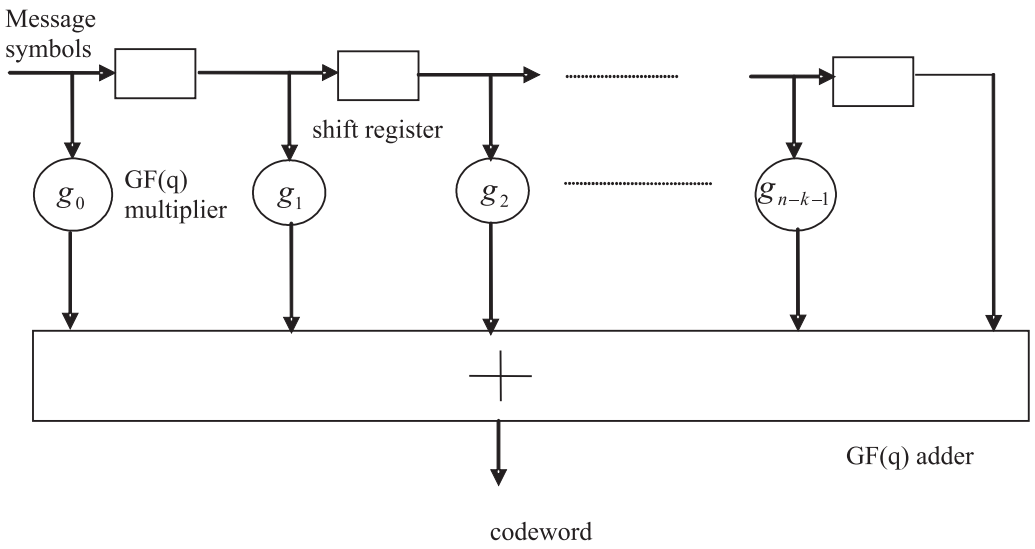

Figure 2.1 Encoder for non-systematic cyclic code.

polynomial because it is a linear code.

A q-ary (n,k) cyclic code also has the following properties [7]

1. The set of code polynomials contains a GF(q)[x] polynomial, $\mathbf{G}(x) = g_0 + g_1 x + \cdots + g_{n-k}x^{n-k}$, which divides $x^n - 1$ and has a degree of $n - k$. $\mathbf{G}(x)$ is called a generator polynomial.

2. Every code polynomial can be represented as $C(x) = M(x)G(x)$, where M(x) is a GF(q)[x] polynomial with highest degree less than k.

The set of all (n,k) cyclic codewords defined on GF(q) can be generated using $C(x) = M(x)G(x)$ for all q^k different message polynomials, $M(x)$. Therefore, a q-ary (n,k) cyclic code has q^k codewords. All these polynomials have GF(q) coefficients. A cyclic code encoded using this method is called a **non-systematic** code. The encoder can be implemented using a shift register based circuit similar to a finite impulse response (FIR) filter as shown in Fig. 2.1.

The message polynomial does not appear unaltered in the codeword generated by using the non-systematic encoder. So, it is not possible to recover the original message polynomial by simply looking at the encoded codeword. If the message polynomial appears unaltered in the encoded code polynomial, then the code is called **systematic**. Both encoding methods are used in this research.

A systematically encoded code polynomial can be obtained using the following method [7] [18]. A polynomial with n elements can be obtained by multiplying a message polynomial $M(x)$ by x^{n-k}. The resulting polynomial is a polynomial in the ring $GF(q)[x]/(x^n-1)$, but it is not a code polynomial unless the message is all zeros. This polynomial can be represented as

$$x^{n-k}M(x) = F(x)G(x) + R(x) \tag{2.2}$$

where $M(x) = m_0 + m_1x + \cdots + m_{k-1}x^{k-1}, F(x) = f_0 + f_1x + \cdots + f_{k-1}x^{k-1}, R(x) = r_0 + r_1x + \cdots + r_{n-k-1}x^{n-k-1}$ are polynomials in the ring $GF(q)[x]/(x^n-1)$ and $G(x) = g_0+g_1x+\cdots+g_{n-k}x^{n-k}$ is the generator polynomial. $R(x)$ is the remainder after polynomial division and has degree less than $n-k$. Equation (2.2) can be modified by moving the polynomial $R(x)$ to the left side of the equal sign giving the code polynomial

$$C(x) = x^{n-k}M(x) - R(x) = F(x)G(x). \tag{2.3}$$

Equation (2.3) shows that a systematic codeword, $\mathbf{C}(x) = c_0 + c_1x + \cdots + m_0x^{n-k} + \cdots + m_{k-1}x^{n-1}$, can be formed by filling the positions of a codeword indexed from 0 to $n-k-1$ by the coefficients of the polynomial $R(x)$ and the positions indexed from $n-k$ to $n-1$ by the coefficients of the message polynomial $M(x)$. This systematic encoding method is slightly more complicated than the non-systematic method.

The systematic encoder has a shift register implementation as shown in Fig. 2.2. In this circuit, the multiplication of x^{n-k} and $M(x)$ is done by feeding the message polynomial $M(x)$ into the right end of the encoder. This encoder uses a linear feedback shift register (LFSR) to perform the polynomial division. Initially, all the contents of the shift register are set to zero. At the first step of the encoding procedure, the switch S is positioned at point A. After k clock pulses, the message polynomial has been shifted into the encoder as well as straight to the transmitter. By doing this, the original message appears unaltered in the encoded code polynomial and occupies the last k coefficients. In step 2 of the encoding procedure, the switch S is positioned to point B. After n-k clock pulses, the n-k GF(q) symbols left in the shift registers are shifted out

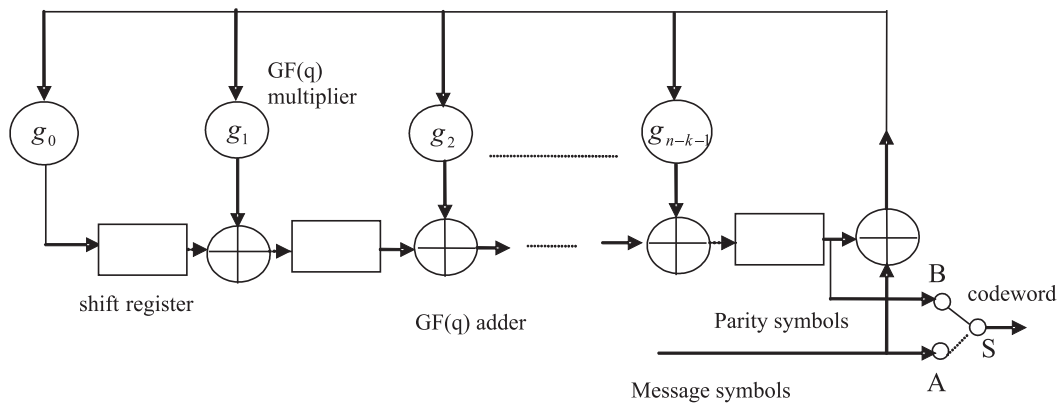

Figure 2.2 Encoder for a systematic cyclic code [2].

to become the coefficients $c_0, c_1, \cdots, c_{n-k-1}$ of the encoded code polynomial. All the operations in Fig. 2.1 and Fig. 2.2 are over GF(q). For GF(2) cases, the circuit can be further simplified by removing the GF(2) multiplier and connection if the corresponding coefficient is "0". The summation can be implemented simply by using an exclusive-OR gate. This attractive implementation makes binary cyclic codes very popular for practical communication systems.

The properties of cyclic codes allow the following decoding algorithm to be used. Property 1 of cyclic codes indicates that there exists a parity check polynomial $H(x)$ which satisfies $G(x)H(x) = x^n - 1$ for a given generator polynomial $G(x)$. Property 1 and 2 together suggest that a received vector $Y(x)$ is a code polynomial if and only if $Y(x)H(x) \bmod (x^n - 1) \equiv 0$. Therefore, a received vector can be decoded using this method. If the result of $Y(x)H(x) \bmod (x^n - 1)$, which is called the syndrome, is not 0, then errors are present in the received vector. Each syndrome value corresponds to a specific error pattern if the cyclic code is designed to have error correction capability. The receiver can correct the error pattern using the syndrome. The received signal is corrected by adding the error pattern to the received vector to obtain a valid codeword.

A different approach can be used to decode systematically encoded cyclic codes. This approach can be implemented using a slightly modified version of the encoder circuit shown in Fig. 2.2. This circuit is shown in Fig. 2.3 and is used in the simulations conducted for this research. In this case, the syndrome is the contents left in the shift register after the entire received vector is shifted into this circuit. The transmitted

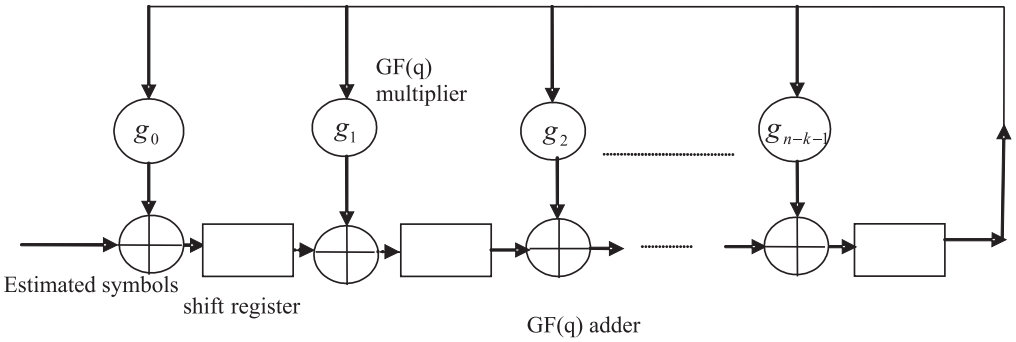

Figure 2.3 Syndrome calculator for a systematic cyclic code [2].

codeword is considered to be received correctly if the syndrome is all zeros. Otherwise, errors are presented in the received vector. These errors can be corrected by the receiver according to the syndrome.

2.3.2 Cyclic Redundancy Check (CRC) Codes

If only error detection is required, cyclic redundancy check (CRC) codes can be used. CRC codes are one of the most popular families of codes used in communication systems, especially for computer networks due to their simple en/decoder and the high code rate they can offer. They are basically systematically encoded cyclic codes with a shortened message. Therefore, they have the same en/decoder implementation as cyclic codes. Shortening the message means treating some of the coefficients in the message of a systematic codeword, $x^j, \cdots, x^{n-1}$, as zeros. Usually, n is a very large number.

A CRC code obtained from shortening a cyclic code is normally no longer a cyclic code. A CRC code can protect an arbitrary number of bits up to n, which is the length of a cyclic code defined by its generator polynomial. For example, the CRC-16 (USA) code has generator polynomial, $G(x) = x^{16} + x^{15} + x^2 + 1 = (x^{15} + x + 1)(x + 1)$, which is a divisor of $x^{32767} - 1$. This means, by using this generator polynomial, the CRC code can protect up to 32751 information bits by attaching only 16 parity bits. Obviously, it has a very high rate. This CRC code also offers very high error pattern coverage, which is defined as the ratio of the number of invalid n-tuples over GF(q) to the number of n-tuples over GF(q) and is the probability that a random word is not a

valid codeword. This can be calculated for the CRC-16 (USA) code as follows [7]

$$\lambda = 1 - q^{k-n} = 1 - 2^{-16} = 0.999985.$$

CRC codes are extensively used in ARQ systems which require error detection capability. They are used due to the simple implementation, shown in Fig. 2.2 and Fig. 2.3, and their high rate.

In this research, the 16 parity bits generated using the CRC-16 (USA) code are attached to every packet which consists of 1064 information bits. The decoder shown in Fig. 2.3 is used to check the received vector. If the syndrome is all zeros, that means the information has been received correctly. Otherwise, errors are contained in the received vector.

2.3.3 Bose-Chaudhuri-Hocquenghem (BCH) Codes and Reed Solomon (RS) codes

Bose-Chaudhuri-Hocquenghem (BCH) codes and Reed Solomon (RS) codes are cyclic codes. So, they share the same merits as the general cyclic codes already discussed. In this section, the encoding of BCH codes and RS codes is introduced. Decoding is not considered as these codes are used for modulation rather than for error control coding later in this thesis.

BCH codes are designed in such a way that a minimum error correction/detection capability can be guaranteed. This capability is normally measured by the minimum Hamming distance, which is defined as the minimum number of different coefficients between any two codeword polynomials in the code. Their generator polynomials are formed by using the following method. Assuming the designed BCH code is an (n,k) cyclic code and α is a primitive root which satisfies $x^n - 1 = 0$, the generator polynomial, $G(x)$, must be a minimal-degree polynomial in GF(q)[x] and has $\alpha^b, \alpha^{b+1}, \cdots, \alpha^{b+\xi-2}$ as its roots, where b ≥ 0 and ξ is the required minimum Hamming distance [7]. The BCH code defined by such a generator polynomial, $G(x)$, has minimum distance $\geq \xi$. In other words, the generator polynomial, $G(x)$, can be obtained by multiplying the irreducible polynomials containing $(x + \alpha^b), (x + \alpha^{b+1}), \cdots, (x + \alpha^{b+\xi-2})$. If $b = 1$, the

BCH code is called a narrow-sense BCH code.

RS codes are a type of BCH code, specifically they are defined on GF(q^m) and have length $q^m - 1$ [7]. RS codes have all the properties of BCH codes. They are maximum-distance separable (MDS) codes, which means the minimum Hamming distance is $(n - k + 1)$ for an (n, k) RS code.

RS and BCH codes can be extended or shortened by adding some overall parity symbols or puncturing some symbols according to a pattern, respectively. They are called extended RS (eRS), extended BCH (eBCH), shortened BCH (sBCH) or shortened RS (sRS), accordingly. Normally, extending a RS or BCH code slightly improves the error detection/correction capability and shortening a code improves the rate.

An extended (16,3) RS code defined on GF(16) is used in this research. This code is obtained by adding an overall GF(16) parity symbol to the (15,3) RS code. The generator polynomial of the (15,3) RS code is obtained as follows

$$\begin{aligned} G(x) = & \ (x+\alpha)\cdot(x+\alpha^2)\cdot(x+\alpha^3)\cdot\cdots\cdot(x+\alpha^{12}) \\ = & \ x^{12}+\alpha^8x^{11}+\alpha^{14}x^{10}+\alpha^8x^9+\alpha^3x^8+x^7+\alpha^2x^6 \\ & +\alpha^{13}x^5+\alpha^{14}x^4+\alpha^2x^3+\alpha^6x^2+\alpha^{13}x+\alpha \end{aligned} \tag{2.4}$$

where α is a root of the primitive polynomial $x^4 + x + 1$, which is a factor of $x^{15} - 1$. The overall GF(16) parity symbol is generated by adding the labels for all the GF(16) symbols in the generated codeword modulo 16 as in [12]. Alternatively, GF(16) addition could be used to create the overall parity symbol [12]. This extra symbol becomes the right most coefficient of the codeword.

2.4 CONVOLUTIONAL CODES

In this section, another class of forward error control codes, namely convolutional codes, is briefly introduced. Convolutional codes are only considered for comparison purposes in this thesis.

Convolutional codes are different from the previously introduced error control codes because the same code can be used on information sequences of any length. The encoder

introduces memory between adjacent symbols. A binary non-systematic convolutional encoder is shown in Fig. 2.4. A convolutional code has the following parameters [17]:

1. number of input streams: k
2. number of output streams: n
3. memory length: M
4. constraint length: v=M+1
5. code rate: r=k/n

The *memory length* is the length of the longest input shift register. The *constraint length* represents how many shifts are required to move a single input symbol from the input to the output of the shift register. It also indicates the maximum number of bits in the output stream affected by a single input bit. If the input information appears unaltered in the output, the encoder is systematic, otherwise it is non-systematic. The encoder in Fig. 2.4 can easily be converted to a systematic encoder by adding an output stream directly connected to the input.

The encoder in Fig. 2.4 does not contain a feedback path. So, it is called a feed forward encoder. Convolutional code encoders can also have feedback paths and these codes are called recursive convolutional codes. They do not have a specific constraint length due to the feedback path. The contents of the shift registers can be used to define the *state* of the encoder. The encoder has a finite number of states. Based on the current input and states, a trellis diagram with associated outputs can be obtained. The *Viterbi* algorithm [2] can be used to decode convolutional codes.

Using a convolutional code leads to bandwidth expansion [1]. This can be solved by combining modulation and convolutional coding. This combination results in trellis-coded modulation (TCM)[2] [2]. This scheme uses a constellation which has more points than required for uncoded transmission at the same data rate. This allows the redundancy introduced by the convolutional coding to be transmitted without increasing the

[1] Using block codes also leads to bandwidth expansion.
[2] In the case of block codes, this is achieved using multilevel codes.

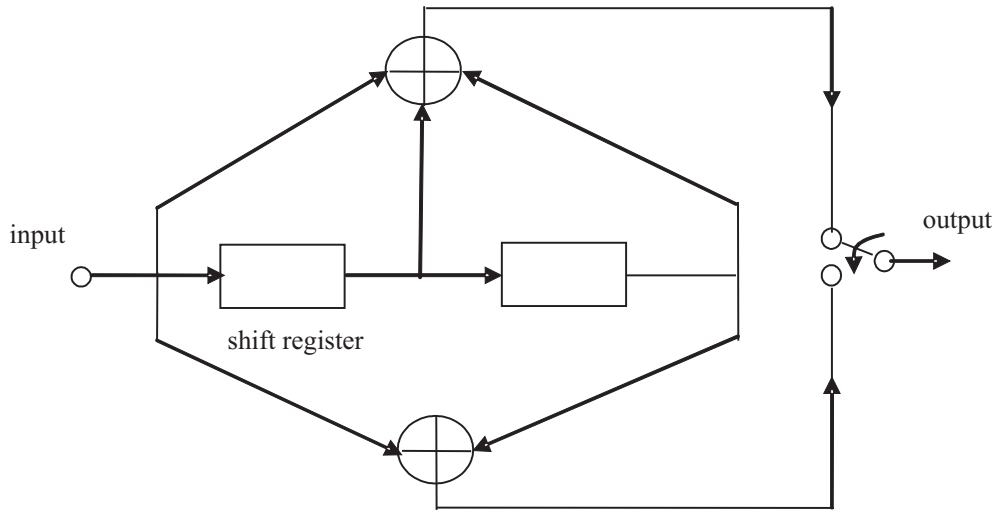

Figure 2.4 A convolutional code encoder [2].

bandwidth. Convolutional codes are used, so that the possible sequences of signals can be modeled as a trellis structure. The Viterbi algorithm with soft-decision decoding can be employed for the decoder. TCM is designed to maximize the Euclidean distance. The design procedure normally involves partitioning an M-ary constellation successively into subsets. The subsets resulting from each partitioning have a larger Euclidean distance between points than that of the previous subset[2].

Convolutional codes and TCM are strong codes with relatively simple decoding algorithms and can be used on an arbitrary number of bits. So, they are extensively used in ARQ scheme design to provide error correction capability across an entire packet, which normally consists of thousands of bits. Some of the ARQ schemes using convolutional codes or TCM are introduced in Section 2.5. However, the focus of this thesis is on block codes.

2.5 AUTOMATIC-REPEAT-REQUEST (ARQ) SYSTEM

In this section, the basic ARQ schemes considered in this thesis for comparison purposes are presented. ARQ scheme design is very different from the forward error control

[2]This assumes Ungerboeck set partitioning

coding design introduced previously because it allows information to be fed back to the transmitter using a feedback channel. The ARQ schemes introduced in this section are designed for single antenna (SISO) wireless communication systems. The extension of SISO ARQ schemes to multiple antenna (MIMO) ARQ schemes will be covered in Chapter 4.

Various communication systems such as packet-switching data networks and computer networks require some crucial data to be received without any error. Forward error control (FEC) codes alone can not achieve this goal because, after decoding, the estimated codeword is delivered to the user regardless of whether or not it is correct. Also, achieving high reliability by using FEC codes alone requires the use of long, powerful error correcting codes. This may make the decoder too complicated and expensive to be implemented. ARQ schemes are designed to solve this problem by sending back a positive or negative acknowledgment (ACK/NACK) from receiver to transmitter for correctly or incorrectly received information, respectively [7] [18]. However, when a feedback channel is not available, FEC codes are the only option.

2.5.1 Pure ARQ Schemes

As mentioned in Chapter 1, the ARQ scheme is part of the data link layer. Therefore, it normally deals with packets rather than individual bits. The earliest ARQ schemes only defined the retransmission strategy. They are called pure ARQ schemes [8]. In these pure ARQ schemes, a high rate error detection code, normally a high rate CRC code, is used to encode a block of bits to form a packet. The decoder checks the validity of the received and decoded packets by decoding the error detection code. Then it sends back an ACK/NACK accordingly. The same packet is resent from the transmitter if the corresponding NACK is received. There are three basic types of pure ARQ schemes, namely stop-and-wait, go-back-N and selective repeat. They differ in the retransmission strategy used when a NACK is received at the transmitter. The three pure ARQ schemes are described briefly as follows.

stop-and-wait: After transmitting one packet, the transmitter stops transmitting new packets until an ACK/NACK is received. Then, a new packet or the same packet

is transmitted accordingly.

go-back-N: The delay between when a codeword is transmitted and when the acknowledgment for this codeword arrives at the transmitter is defined as the time interval [8]. Let us assume that, during this interval, the transmitter sends N-1 distinct packets. If a NACK is received for a packet, this packet and the following N-1 packets are retransmitted again regardless of whether any of the following N-1 packets were received correctly. The efficiency of this ARQ scheme deteriorates when the time interval is large [8].

selective-repeat: The transmitter is continuously transmitting packets. When a NACK is received, only the packet corresponding to this NACK gets retransmitted.

Out of the three pure ARQ schemes above, the *stop-and-wait* scheme is the most inefficient due to the idle time used to wait for the arrival of an acknowledgment. The efficiency is improved by using the *go-back-N* scheme, but deteriorates for a communication system with high data rate and a long time interval [8]. The efficiency of the *selective-repeat* scheme is the highest. However, the efficiency is obtained at the cost of requiring extra memory.

For the *stop-and-wait* scheme, no memory is required at the transmitter or receiver. It can be used for communication systems which do not have duplex links, which means the information is only allowed to flow in one direction at a time. For the *go-back-N* scheme, a memory buffer, which is big enough to store N-1 codewords is required at the transmitter. The *selective-repeat* scheme has the highest memory requirements. It requires a memory buffer at both the transmitter and the receiver. A transmitted packet has to be stored in the transmitter until it has been acknowledged. The receiver memory is required because a correctly received packet has to be stored in the receiver until all packets which were transmitted before this packet are received correctly. The size of the required memory must be chosen carefully to prevent overflow [18].

Because most practical systems can only tolerate limited delays, the selective-repeat scheme with a specified maximum number of retransmissions is typically used

for practical systems. The transmission is truncated after the maximum number of retransmissions has been reached and this packet is declared a dropped packet. This is called *truncated selective-repeat* ARQ (TS-ARQ). Some modified selective-repeat schemes were proposed in [19] and [20]. In [19], a modified TS-ARQ scheme was proposed to avoid memory overflow. The scheme proposed in [20] deals with a very noisy channel.

2.5.2 Type-I and II Hybrid ARQ (HARQ) Schemes

Apart from the retransmission strategy, the performance of an ARQ scheme is significantly affected by the probability of generating a retransmission request. If the FEC code used by an ARQ scheme has both error detection and correction capability, then hybrid ARQ (HARQ) schemes can be used. The *selective-repeat* scheme is adopted by all HARQ schemes discussed in this thesis as part of the retransmission strategy. All HARQ schemes for SISO communication channels can be roughly classified into two classes [7] [18].

Type-I: A FEC code is used which has error correction and detection capability simultaneously. So, the receiver attempts error correction on receipt of a packet containing errors. If the decoding is not successful, then a retransmission request is generated. The same packet will be retransmitted until an ACK is received.

Type-II: In this HARQ scheme, a codeword C_1 is first transmitted with the parity bits/symbols required for error detection only as done for the pure ARQ schemes. If the receiver detects the presence of error bits/symbols in the received codeword, this erroneous codeword will be stored in memory and a retransmission will be requested. Before the second transmission, the same data is encoded using another error correction code $\mathbb{C}_3$, which is normally a powerful low rate code and can correct most common error patterns for the particular communication environment under consideration. The code $\mathbb{C}_3$ is also invertible, which means the original data can be recovered from the parity bits/symbols alone. A codeword C_2 is obtained by encoding the parity bits/symbols of $\mathbb{C}_3$ using the same error

detection code used for the first transmission. Then, C_2 is transmitted for the second transmission. If no errors are detected, then the original data is recovered from the received C_2 and delivered to the upper layers. Otherwise, these error correction parity bits/symbols are combined with the data part of the stored version of C_1 to perform error correction. If this error correction is not successful, a third transmission will be requested. For this retransmission, either the codeword from the first transmission or the parity bits from the second transmission will be retransmitted. The decoding procedure is the same as for the first or second transmission, accordingly. This scheme is also called incremental redundancy (IR) because it transmits additional redundant information during the retransmissions.

Compared with pure ARQ schemes, the type-I HARQ schemes decrease the probability of retransmissions due to the use of an error control code with error correction capability. But it is not efficient to use this scheme for wireless communication systems. This is because, for wireless communication systems, the channel condition are time varying. The instant SNR for the decoder varies depending on the current channel condition. This makes it very hard to choose a good code, which has minimum redundancy and sufficient error correction capability. When the SNR is very high, the error correction capability becomes redundant, but when the SNR is very low the error correction capability is inadequate.

Type-II HARQ schemes were proposed to overcome the problems of type-I HARQ. They act like an adaptive system. When the channel is quite good, they act like pure ARQ schemes. The error correction capability is added only when it is needed. In conventional type-II HARQ schemes, only data symbols or error correction parity symbols get transmitted with each transmission. Hence, the previous erroneously received packets need to be stored in memory for combining purpose instead of being discarded. The error correction capability offered by type-II HARQ schemes is provided by the low rate code $\mathbb{C}_3$. Type-II HARQ schemes normally use block or convolutional codes. Several type-II HARQ schemes were proposed in [21], [22], [23] and [24].

MDS codes, especially Reed Solomon (RS) codes, have attracted lots of attention for HARQ schemes. These HARQ schemes are designed based on the error correction capability of the MDS codes. MDS codes have several attractive properties which are suitable for both type-I and II HARQ schemes. These properties include [24]

1. RS codes are MDS codes. This means they can provide maximum Hamming distance between codewords for a given amount of redundancy. Therefore, they have excellent performance for error detection or error correction.

2. Punctured MDS codes are also MDS codes. This means shortened versions can still provide good performance.

3. Normally, MDS codes can be easily partitioned into data symbols and parity symbols.

4. MDS codes are strongly invertible. So, the data can be easily recovered from the parity symbols alone.

Several hybrid ARQ schemes based on the error correction capability of RS codes have been proposed in [22] [23] [24] [25] [26]. This is the traditional use of MDS codes in HARQ scheme design. They are designed for SISO channels. For MIMO systems, other design criteria should be used as detailed in Chapter 3. In Chapter 4, several novel HARQ schemes, which are also based on RS codes, are proposed. In these schemes, RS codes are used for modulation rather than error correction. Therefore, traditional type-I and II HARQ schemes cannot be directly applied to them. Hence, in this research, TS-ARQ scheme is used for comparison purposes as in [13]

Now, some existing type-I and II HARQ schemes for SISO channels are discussed. They are designed based on the error correction capability of error control codes including MDS codes and convolutional codes. In [25], a type-I HARQ scheme was proposed. It uses the error correction capability of RS codes. The receiver decodes the estimated RS codes using the Berlekamp-Massey algorithm. This type-I hybrid scheme was extended to allow erasure decoding in [26].

A type-II HARQ scheme is designed for MDS codes including RS codes in [24]. In this scheme, codeword C_3 is obtained by encoding the information packet using an (n, k) MDS code $\mathbb{C}_3$. Then, two punctured codewords with equal length, C_1 and C_2, are obtained by puncturing C_3. They are used for the first transmission and retransmission, respectively. C_3 can be reconstructed in the receiver by combining C_1 and C_2 according to their puncturing pattern to perform error correction. $\mathbb{C}_3$ should provide sufficient error correction capability to ensure reliable transmission under the considered channel condition.

Another type-II HARQ scheme was proposed in [22]. In this scheme, the receiver will have a punctured RS codeword with a rate of $k_R \quad (k_1 > k_2 > \cdots > k_{R-1} > k_R)$, obtained from puncturing a full length RS code, after the R^{th} transmission. The transmitter only transmits the difference between the R^{th} and $(R-1)^{th}$ transmission. Erasure decoding was also considered [22]. Punctured RS codes with type-II HARQ scheme are also studied in [23]. They showed that type-II HARQ schemes using a non-binary RS code can achieve better performance than type-II HARQ schemes using a binary error correction code.

Convolutional codes are also used in many type-II HARQ schemes. A type-II HARQ scheme using a rate half convolutional code as the error correction code is described in [8].

2.5.3 Other HARQ Schemes

Since all conventional ARQ schemes introduced so far simply continue transmitting a packet until a good one is received, these ARQ schemes are only effective for burst-error channels [27], but are ineffective when the duration of bad channel conditions is so long that all the received packets contain errors [28]. The error correction capability of type-I and II HARQ schemes is provided by the chosen error correction code, so it does not increase as the number of transmissions increases. This problem can be overcome by the HARQ scheme proposed in [28], which extracts useful information from all copies to increase the error correction capability as the number of retransmissions increases.

In the scheme of [28], packets are combined using a maximum-likelihood decod-

ing approach. In contrast, the combining scheme used in the type-II HARQ schemes introduced in Section 2.5.2 is based on the error correction capability of the chosen error correction code, which is closely related to the minimum Hamming distance of the code. Therefore, in these schemes, the receiver employs an algebraic decoder to perform the error correction. In the scheme of [28], a CRC encoded packet is encoded using a convolution code with rate R. This packet is repeatedly transmitted until an ACK is received. When a packet is first received, it is decoded using a rate R maximum-likelihood decoder. If there are detected errors in the current received packet, the same packet is retransmitted. Assuming packets $\mathrm{Y}^1, \mathrm{Y}^2, \cdots, \mathrm{Y}^n$ have been received and each packet has failed the CRC check, then all n received packets are combined as a code $\mathbb{C}_{low}$ of rate R/n. The receiver has to use a maximum-likelihood decoder rather than an algebraic decoder to decode the combined packet $\mathbb{C}_{low}$. This is because the minimum distance for a codeword repeated n times is n times the minimum distance of the chosen codeword set[3]. Therefore, this minimum Euclidean distance increases with the number of retransmissions. This combining scheme is referred to as Chase Combining (CC). In [28], each packet can also be weighted by its reliability to further improve performance. Chase Combining can be applied to block or convolutional codes.

Another combining scheme which selectively combines some packets was proposed in [29]. This scheme only considers punctured convolutional codes and is similar to [28]. The decoder selectively combines some of the received packets and a retransmission request is sent only when the decoding processes for all combinations fail. For example, if Y^1 and Y^2 have been received and the CRC checks for Y^1, Y^2 and Y^1Y^2 have failed, then Y^3 is requested. Once Y^3 is received, the decoder for this HARQ scheme tries combinations, $Y^3, Y^3Y^1, Y^3Y^2, Y^3Y^2Y^1$, until the decoding result of one of them passes the error detection code. Otherwise, a NACK is initiated.

Apart from linear block codes and convolutional codes, TCM is also extensively used in HARQ schemes. The HARQ scheme proposed in [30] uses different TCM schemes for each transmission. This combining scheme is also designed to maximize the Euclidean distance. Different TCM schemes should share the same trellis structure

[3]This assumes the same codeword is transmitted every time.

and the same input code assignment for each state transition in order to use the same Viterbi decoder at the receiver. The decoded bits are accepted if they pass the error detection code. In order to find good codes for each transmission, a code search was conducted in [30].

Another novel HARQ scheme and decoder was proposed in [31] [32] and [33]. In this scheme, the bit-symbol mapping varies for each retransmission and all copies of the packet are combined and decoded using the decoding algorithm proposed in [33] [34]. This mapping is chosen to increase the combined squared Euclidean distance to minimize the bit error rate. The focus of this work is to introduce a new type of diversity if an ARQ system is employed. Therefore, no FECs, except an error detection code, are used in [31] [32] [33] and [34].

Other codes including Turbo codes, which are based on systematic recursive convolutional codes, and low-density parity-check (LDPC) codes can approach the channel capacity due to their concatenated coding and iterative decoding. To date, their application with ARQ schemes has been focused on BPSK modulation. Bit-interleaved coded modulation (BICM) [35] can be used to achieve higher spectral efficiency while still maintain good bit error rate performance. But this is outside the scope of this thesis.

2.5.4 Summary for ARQ Schemes

All the ARQ schemes introduced so far are designed for SISO systems. The above discussion shows that using a powerful HARQ scheme can significantly improve the performance compared with pure ARQ schemes. Recent MIMO systems research show that MIMO systems can provide a more reliable wireless link than SISO systems [3] [6]. Therefore, these HARQ schemes designed for SISO sytems should be modified or new HARQ schemes need to be designed for MIMO systems. Some of the introduced HARQ schemes are only suitable for SISO systems. Others have been extended to MIMO systems during recent years [30] [36]. These extensions will be introduced in Chapter 4. Also, in Chapter 4, several novel HARQ schemes are proposed for MIMO systems, which use RS codes. The RS codes will be used in in a different manner than in the HARQ

schemes using RS codes introduced in this chapter.

2.6 SUMMARY

In this chapter, the fundamentals of *group*, *ring* and *Galois field* theory as related to error control coding were summarized. A powerful and popular family of error control codes, *cyclic codes*, was described. Its en/decoder structure has a very simple and efficient hardware implementation, especially for cyclic codes defined on GF(2). Therefore, cyclic codes are extensively used in many practical communication systems as well as this research. The encoding and/or decoding algorithms for some specific types of cyclic codes, namely CRC, BCH and RS codes, were explained. Convolutional codes, which are used for many ARQ schemes, were introduced very briefly. In Section 2.5, various ARQ schemes were introduced. The approach for HARQ design in these schemes is to use the error correction capability of the chosen error control codes including MDS codes, convolutional codes and TCM. Some of these schemes were extended to MIMO systems by other researchers. It also becomes clear that a powerful combining scheme is one of the keys to designing powerful ARQ systems, which is the main goal of this thesis.

Chapter 3

MIMO SYSTEMS

In this chapter, multiple input multiple output (MIMO) wireless communication systems are described. They use multiple transmit and receive antennas to obtain spatial diversity. The channel associated with a MIMO system is called a MIMO channel. Code design criteria for MIMO systems are introduced in Section 3.3 followed by the introduction of MIMO code families including space-time block codes and space-time trellis codes in Section 3.4 and 3.5.

3.1 INTRODUCTION

As mentioned in Chapter 1, multipath fading is the most severe impairment for wireless communication systems. One of the best techniques to combat channel fading is diversity combining of independently faded signal paths [6]. The possibility that the multiple faded paths experience deep fades simultaneously is much smaller than that of a single path. By sending the same information over several independent signal paths, the received signals from the multiple paths can be combined in such a way that the fading effect is reduced. Diversity techniques can be classified into several categories including (but not restricted to) [3] [37]

1. Temporal Diversity: Diversity is provided in the temporal domain. The same message is sent during multiple time intervals. The separation time between adjacent time intervals must be longer than the coherence time of the channel in order to achieve independent channel states. An ARQ system is a typical temporal diversity system.

2. Frequency Diversity: Diversity is provided in the frequency domain. The same message is sent using different carrier frequencies which are separated by at least the coherence bandwidth of the channel.

3. Spatial (antenna) Diversity: Diversity is provided in the spacial domain, normally by employing multiple antennas at the transmitter and/or receiver. In order to achieve independently faded paths between different transmit to receive antenna links, the multiple antennas at the transmitter and/or receiver need to be separated by a sufficient distance (typically a few wavelengths) [3].

Antennas can also be used to provide other types of diversity, for example, angular diversity and polarization diversity.

If spatial diversity is achieved using one transmit antenna and multiple receive antennas, it is called receive diversity. The multiple received versions of the transmitted signals can be combined to improve performance. Common combining schemes include selection combining and maximum ratio combining [3] [6]. If the spatial diversity is provided by using multiple transmit antennas, it is called transmit diversity.

A wireless communication system should be designed to use as many types of diversity as possible to provide a fast, reliable communication link. But, normally, not all types of diversity are available at the same time. The communication system considered in this thesis uses both spatial diversity and temporal diversity. Temporal diversity is obtained by the use of ARQ protocols and codes designed for a MIMO system. Spatial diversity, which contains both transmit and receive diversity, is realized by using multiple transmit and receive antennas and codes designed for a MIMO system. The codes designed for a MIMO system, generally, exploit the spatial and temporal diversity available from using multiple antennas and multiple time slots. So, they are called space-time codes (STCs). A typical MIMO system is shown in Fig. 3.1.

A MIMO system is equipped with n_t transmit antennas and n_r receive antennas. Assuming the fading is flat, the received signal at each of the n_r receive antennas is a linear superposition of faded versions of the signals transmitted from all transmit

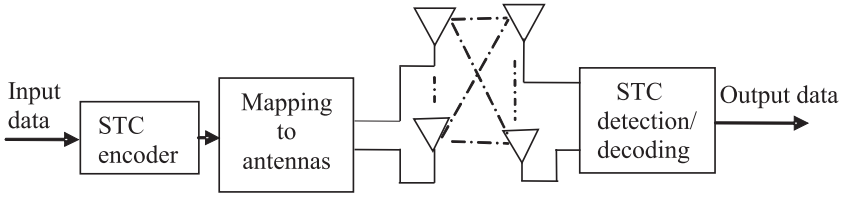

Figure 3.1 A MIMO system.

antennas plus noise. This received signal can be represented as

$$y_i(t) = \sum_{j=1}^{n_t} a_j h_{i,j} \cos(2\pi f_c t + \phi_j + \triangle\phi_{i,j}) + n(t), \quad i = 1, 2, 3 \cdots, n_r \tag{3.1}$$

where $a_j \cos(2\pi f_c t + \phi_j)$ is the signal transmitted from the j^{th} transmit antenna, $h_{i,j}$ and $\triangle\phi_{i,j}$ are the amplitude coefficient and phase shift caused by the fading between the j^{th} transmit antenna and the i^{th} receive antenna and $n(t)$ is the additive Gaussian noise at time t. For ease of analysis, the baseband representation of (3.1) can be written in matrix form as [3] [6] [9]

$$\begin{aligned}
\mathbf{Y} &= \begin{pmatrix} Y_1^1 & Y_1^2 & \cdots & Y_1^T \\ Y_2^1 & Y_2^2 & \cdots & Y_2^T \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n_r}^1 & Y_{n_r}^2 & \cdots & Y_{n_r}^T \end{pmatrix} = \mathbf{HX} + \mathbf{N} \\
&= \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n_t} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r,1} & h_{n_r,2} & \cdots & h_{n_r,n_t} \end{pmatrix} \times \begin{pmatrix} X_1^1 & Y_1^2 & \cdots & X_1^T \\ X_2^1 & X_2^2 & \cdots & X_2^T \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_t}^1 & X_{n_t}^2 & \cdots & X_{n_t}^T \end{pmatrix} \\
&\quad + \begin{pmatrix} N_1^1 & N_1^2 & \cdots & N_1^T \\ N_2^1 & N_2^2 & \cdots & N_2^T \\ \vdots & \vdots & \ddots & \vdots \\ N_{n_r}^1 & N_{n_r}^2 & \cdots & N_{n_r}^T \end{pmatrix}
\end{aligned} \tag{3.2}$$

where $\mathbf{X}$ is the $n_t \times T$ transmitted codeword matrix, $\mathbf{Y}$ is the $n_r \times T$ received signal matrix, $\mathbf{H}$ is the $n_r \times n_t$ channel matrix, where the element $h_{i,j}$ represents the coefficient of the path connecting the j^{th} transmit antenna and the i^{th} receive antenna, and $\mathbf{N}$ is the $n_r \times T$ noise matrix generated at the $t^{th}, t = 1, 2, \cdots, T$, sampling instant. This assumes the STC spans T time slots.

The matrix $\mathbf{H}$ is assumed to be constant for these T time slots. Under the assumption that the fading is flat and has a Rayleigh envelope, the "in phase" and "quadrature" terms of all fading coefficients are independent Gaussian random variables with zero-mean and $1/2$ variance. In other words, the variance of each channel coefficient is normalized to be 1. We assume that all the fading coefficients in $\mathbf{H}$ are uncorrelated. Theoretical analysis shows that uncorrelated fading paths can be achieved by separating the antennas by a few wavelengths [3] [9], but how fast each coefficient varies depends on the environment considered. The channel matrix $\mathbf{H}$ can be estimated by the receiver by sending and using training information.

Elements in the same column of matrix $\mathbf{X}$ represent the modulated signals transmitted from each transmit antenna at the same signaling interval and the j^{th} row of matrix $\mathbf{X}$ describes the modulated signal transmitted from the j^{th} antenna during T consecutive signaling intervals within one block. All elements of the noise matrix $\mathbf{N}$ are iid Gaussian random variables with zero-mean and variance $N_0/2$ per dimension.

If channel state information (CSI) is available at the receiver, then a maximum likelihood decoding algorithm can be used to estimate the transmitted information. The decision metric is calculated based on the squared Euclidean distance between the actual received signal matrix and all possible hypothesized received symbol matrices. The decoder picks the codeword which gives the smallest distance as the decoded result, meaning

$$\tilde{\mathbf{X}} = \arg\min_{\tilde{\mathbf{X}} \in \psi} \sum_{t=1}^{T} \sum_{j=1}^{n_r} \left| Y_j^t - \sum_{i=1}^{n_t} h_{j,i} X_i^t \right|^2 \tag{3.3}$$

where ψ denotes all possible codewords.

3.2 MIMO CHANNEL CAPACITY AND WATER FILLING PRINCIPLE

Theoretical analysis in [3] [6] [9] has shown that a MIMO channel has a much larger capacity than its SISO counterpart due to the multiple paths between the transmitter and the receiver. We now describe how the MIMO channel capacity can be calculated based on the description in [3] [6] [9].

The mutual information between the inputs and outputs of the MIMO channel is given by

$$\mathbf{I}(\mathbf{x};\mathbf{y}|\mathbf{H}=H) = \log_2(\det\left[I_{n_r} + (1/N_0)H^H \cdot \mathbf{R}_x \cdot H\right]) \tag{3.4}$$

where det[A] denotes the determinant of the matrix A, I_{n_r} is an $n_r \times n_r$ identity matrix, $(H)^H$ denotes the Hermitian of H, N_0 is the total noise variance and $\mathbf{R}_x$ is the covariance matrix of the inputs to the MIMO channel. Then, the normalized capacity is the amount of information obtained by maximizing (3.4) over all inputs and is given by

$$C/W = \max_{\mathbf{R}_x|\mathbf{Tr}(\mathbf{R}_x)\leq\mathbf{P}_{total}} \log_2(\det\left[I_{n_r} + (1/N_0)H^H \cdot \mathbf{R}_x \cdot H\right]) \quad \text{bits}/s/\text{Hz} \tag{3.5}$$

where C is the channel capacity and W is the given bandwidth. $\mathbf{Tr}(\mathbf{R}_x)$ denotes the trace of the matrix $\mathbf{R}_x$, which is defined as the sum of all elements along the main diagonal. It is the total transmit power which should be less than or equal to the total power constraint, $\mathbf{P}_{total}$.

The capacity of a time-invariant MIMO channel $\mathbf{H}$ is now considered. Equation (3.5) can be maximized by optimizing $\mathbf{R}_x$. But it is normally assumed that the transmit power is equally distributed among all transmit antennas, because the CSI is not known at the transmitter. Under such assumption, the power for each transmit antenna is $\mathbf{P}_{total}/n_t$. This means the covariance matrix of the inputs, $\mathbf{R}_x$, equals $(\mathbf{P}_{total}/n_t) \cdot I_{n_t}$. Therefore (3.5) can be written as

$$(C/W|\mathbf{H}=H) = \log_2(\det\left[I_{n_r} + (\mathbf{P}_{total}/(N_0 \cdot n_t))H^H \cdot H\right]) \quad \text{bits}/s/\text{Hz} \tag{3.6}$$

Singular value decomposition (SVD) is used to decompose the $n_r \times n_t$ matrix $\mathbf{H}$

with rank r_H as

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H \tag{3.7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $n_r \times n_r$ and $n_t \times n_t$ unitary matrices, which means $\mathbf{U}\mathbf{U}^H = \mathbf{I}_{n_r}$ and $\mathbf{V}\mathbf{V}^H = \mathbf{I}_{n_t}$. $\mathbf{D}$ is an $n_r \times n_t$ non-negative and diagonal matrix. The diagonal entries of $\mathbf{D}$ equal the non-negative square root of the eigenvalues, $\lambda_i, i = 1, 2, \cdots, r_H$, of matrix $\mathbf{H}\mathbf{H}^H$. Now (3.6) can be rewritten as

$$(C/W|\mathbf{H} = H) = \sum_{i=1}^{r_H} \log_2 \left[1 + (\mathbf{P}_{total}/(N_0 \cdot n_t)) \cdot \lambda_i\right] \tag{3.8}$$

This shows that the MIMO channel can be decomposed into r_H parallel subchannels and that the total channel capacity is the sum of the capacities of all r_H subchannels for a given realization of the channel $\mathbf{H}$. This means a MIMO channel has larger capacity than a SISO channel.

If all entries of the channel matrix $\mathbf{H}$ are random variables, the capacity of this random channel realization should be a random variable [6]. The related capacity definitions in this case are ergodic capacity and outage capacity [9]. The ergodic capacity defines the maximum capacity averaged over all possible channel realizations, which is based only on the distribution of $\mathbf{H}$, for a given transmission strategy. The ergodic capacity of the MIMO channel, under the assumption that the power is evenly distributed between all transmit antennas, can be represented as

$$C/W = \mathrm{E}_{\mathbf{H}} \left[\log_2(\det[I_{n_r} + (\mathbf{P}_{total}/(N_0 \cdot n_t))H^H \cdot H])\right] \quad \text{bits}/s \tag{3.9}$$

where $\mathrm{E}_{\mathbf{H}}[\cdot]$ is the expectation with respect to the distribution of the channel matrix $\mathbf{H}$.

Alternatively, we can use outage capacity, which is defined as follows. Assuming that the transmitter fixes a minimum signal to noise ratio (SNR_{min}) and encodes the information using a data rate $C_{rate} = B \log_2(1 + SNR_{min})$, an outage is declared if the received SNR is below the SNR_{min}, because the probability of a correct decoding approaches zero [6]. This data rate, C_{rate}, is called the outage capacity C_{out}. The

probability of such an outage is defined as the outage probability $\mathcal{P}_{out}$. For a fixed SNR_{min}, C_{out} is a function of $\mathcal{P}_{out}$. Therefore, given a $\mathcal{P}_{out}$, a corresponding C_{out} can be found and vice versa.

The capacity of a MIMO channel can be further increased, if there exists a feedback channel that can feed CSI back to the transmitter. Under this assumption, it is found that the transmit power can be distributed unevenly according to the water filling principle to maximize the channel capacity. This principle for MIMO case is derived from (3.5). For this case, the normalized capacity can be shown to be [3]

$$(\mathbf{C}/\mathbf{W}|\mathbf{H} = H) = \sum_{i=1}^{n_t} \log_2 \left[1 + \frac{\mathbf{P}_i \lambda_i}{N_0}\right] \quad i = 1, 2, \cdots, n_t \tag{3.10}$$

where $\mathbf{P}_i$ denotes the power allocated to the i^{th} transmit antenna. By applying Lagrange multipliers, the following function can be obtained [3]

$$\mathbf{Z} = \sum_{i=1}^{n_t} \log_2 \left[1 + \frac{\mathbf{P}_i \lambda_i}{N_0}\right] + \mathbf{L}\left(\mathbf{P}_{total} - \sum_{i=1}^{n_t} \mathbf{P}_i\right) \tag{3.11}$$

where $\mathbf{L}$ is the Lagrange multiplier. The unknown transmit power for each antenna, $\mathbf{P}_i$, are the roots of the functions, which are obtained by taking partial derivatives of $\mathbf{Z}$ with respect to $\mathbf{P}_i$ in (3.11). $\mathbf{P}_i$ can be represented as

$$\mathbf{P}_i = \mu - \frac{N_0}{\lambda_i}, \quad i = 1, 2, \cdots, n_t. \tag{3.12}$$

where μ is a constant which is used to satisfy the total power constraint. Equation (3.12) shows that, according to the water filling principle, more power will be allocated to transmit antennas with higher channel gains, indicated by λ_i. The use of water filling with the proposed HARQ space-time scheme will be discussed in Chapter 4. This water filling principle is only optimal for independent Gaussian inputs for each transmit antenna. A modified water filling principle called mercury/water filling was proposed in [38]. It is designed for arbitrary inputs for each transmit antenna, but this is beyond the scope of this thesis.

3.3 SPACE-TIME CODE DESIGN CRITERIA

As mentioned in Section 3.1, in order to achieve good performance, MIMO systems use specially designed codes called STCs. In this section, some guidelines for the design of STCs are given. The following descriptions are based on the descriptions in [3] [6] [9] [37] [39] [40].

Assuming that the STC codeword

$$\begin{pmatrix} X_1^1 & X_1^2 & \cdots & X_1^T \\ X_2^1 & X_2^2 & \cdots & X_2^T \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_t}^1 & X_{n_t}^2 & \cdots & X_{n_t}^T \end{pmatrix}$$

is sent and that the CSI is perfectly known at the receiver, a maximum likelihood decoder will make an erroneous decision in favor of another codeword

$$\begin{pmatrix} \hat{X}_1^1 & \hat{X}_1^2 & \cdots & \hat{X}_1^T \\ \hat{X}_2^1 & \hat{X}_2^2 & \cdots & \hat{X}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ \hat{X}_{n_t}^1 & \hat{X}_{n_t}^2 & \cdots & \hat{X}_{n_t}^T \end{pmatrix}$$

if and only if

$$\sum_{t=1}^{T}\sum_{j=1}^{n_r}\left|Y_j^t - \sum_{i=1}^{n_t} h_{j,i}X_i^t\right|^2 > \sum_{t=1}^{T}\sum_{j=1}^{n_r}\left|Y_j^t - \sum_{i=1}^{n_t} h_{j,i}\hat{X}_i^t\right|^2 . \tag{3.13}$$

Therefore, the pairwise error probability is a function of the squared Euclidean distance between the faded versions of two codeword matrices $\mathbf{X}$ and $\hat{\mathbf{X}}$. It is given by

$$\mathcal{P}(\mathbf{X}, \hat{\mathbf{X}}|\mathbf{H}) = \mathcal{Q}\left(\sqrt{\frac{E_s}{2N_0} d_h^2(\mathbf{X}, \hat{\mathbf{X}})}\right), \tag{3.14}$$

where E_s is the energy per symbol at each transmit antenna, $d_h^2(\mathbf{X}, \hat{\mathbf{X}})$ is the Euclidean

distance between the faded versions of $\mathbf{X}$ and $\hat{\mathbf{X}}$, which is defined by

$$\begin{aligned} d_h^2(\mathbf{X}, \hat{\mathbf{X}}) &= ||\mathbf{H} \cdot (\hat{\mathbf{X}} - \mathbf{X})||^2 \\ &= \sum_{t=1}^{T} \sum_{j=1}^{n_r} \left| \sum_{i=1}^{n_t} h_{j,i} (\hat{X}_i^t - X_i^t) \right|^2 \end{aligned} \tag{3.15}$$

and $\mathcal{Q}(\cdot)$ is the complementary error function defined by

$$\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt \leq \frac{1}{2} e^{-x^2/2}. \tag{3.16}$$

Therefore, (3.14) can be upper bounded by

$$\mathcal{P}(\mathbf{X}, \hat{\mathbf{X}} | \mathbf{H}) \leq \frac{1}{2} \exp\left(-d_h^2(\mathbf{X}, \hat{\mathbf{X}}) \frac{E_s}{4N_0} \right) \tag{3.17}$$

In order to further simplify (3.17), a codeword difference matrix, $\mathbf{B}(\mathbf{X}, \hat{\mathbf{X}})$, is defined by

$$\begin{aligned} \mathbf{B}(\mathbf{X}, \hat{\mathbf{X}}) &= \mathbf{X} - \hat{\mathbf{X}} \\ &= \begin{pmatrix} X_1^1 - \hat{X}_1^1 & X_1^2 - \hat{X}_1^2 & \cdots & X_1^T - \hat{X}_1^T \\ X_2^1 - \hat{X}_2^1 & X_2^2 - \hat{X}_2^2 & \cdots & X_2^T - \hat{X}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_t}^1 - \hat{X}_{n_t}^1 & X_{n_t}^2 - \hat{X}_{n_t}^2 & \cdots & X_{n_t}^T - \hat{X}_{n_t}^T \end{pmatrix} \end{aligned} \tag{3.18}$$

Now, (3.15) can be rewritten as

$$d_h^2(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{j=1}^{n_r} \mathbf{h}_j \mathbf{A}(\mathbf{X}, \hat{\mathbf{X}}) \mathbf{h}_j^H \tag{3.19}$$

where $\mathbf{h}_j$ denotes the vector $(h_{j,1}, h_{j,2}, \cdots, h_{j,n_t})$ and $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}}) = \mathbf{B}(\mathbf{X}, \hat{\mathbf{X}}) \cdot \mathbf{B}^H(\mathbf{X}, \hat{\mathbf{X}})$. Because matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ is a non-negative Hermitian matrix, it can be represented as

$$\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}}) = \mathbf{V}^H \mathbf{\Lambda} \mathbf{V} \tag{3.20}$$

where $\mathbf{V}$ is a unitary matrix. The row vectors of $\mathbf{V}$ are the eigenvectors of the matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ and the diagonal elements of the diagonal matrix $\mathbf{\Lambda}$ are the eigenvalues of the matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$. Substituting (3.20) and (3.19) into (3.17), the pairwise error

probability can be upper bounded by

$$\mathcal{P}(\mathbf{X}, \hat{\mathbf{X}}|\mathbf{H}) \leq \frac{1}{2} \exp\left(-\sum_{j=1}^{n_r} \sum_{i=1}^{n_t} \lambda_i |\beta_{j,i}|^2 \frac{E_s}{4N_0} \right) \tag{3.21}$$

where the complex number $\beta_{j,i}$ denotes the inner product between the row vectors of matrix $\mathbf{H}$, $\mathbf{h}_j, j = 1, 2, \cdots, n_r$, and the row vectors of matrix $\mathbf{V}$, $\mathbf{v}_i, i = 1, 2, \cdots n_t$. Inequality (3.21) shows the upper bound of the conditional pairwise error probability is a function of $|\beta_{j,i}|$, which depends on the distribution of matrix $\mathbf{H}$.

The design criteria for STCs are defined for two cases based on the value of rn_r, where r is the rank of matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$. Note that we only consider the Rayleigh fading channel. The rank r equals the number of non zero eigenvalues, $\lambda_i > 0, i = 1, 2, \cdots, r$, of matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$. If the value of rn_r is large enough ($rn_r \geq 4$), then the random variable $\sum_{j=1}^{n_r} \sum_{i=1}^{n_t} \lambda_i |\beta_{j,i}|^2$ can be considered to be Gaussian due to the central limit theorem. As a result, this Gaussian random variable, denoted as D, can be described using its mean, μ_D, and variance, σ_D^2. Then, the given channel condition can be removed and the pairwise error probability can be upper bounded by

$$\mathcal{P}(\mathbf{X}, \hat{\mathbf{X}}) \leq \int_{D=0}^{+\infty} \frac{1}{2} \exp\left(-\frac{E_s}{4N_0} D \right) p(D) d(D), \tag{3.22}$$

where $p(D)$ is the pdf of the Gaussian random variable D. Using the equation [3]

$$\int_{D=0}^{+\infty} \exp(-\gamma D) p(D) d(D) = \exp\Big(\frac{1}{2} \gamma^2 \sigma_D^2 - \gamma \mu_D \Big) \mathcal{Q}\Big(\frac{\gamma \sigma_D^2 - \mu_D}{\sigma_D} \Big), \tag{3.23}$$

the upper bound in (3.22) can be further expressed as

$$\mathcal{P}(\mathbf{X}, \hat{\mathbf{X}}) \leq \frac{1}{2} \exp\left(\frac{1}{2} \Big(\frac{E_s}{4N_0} D \Big)^2 \sigma_D^2 - \frac{E_s}{N_0} \mu_D \right) \mathcal{Q}\Big(\frac{E_s}{4N_0} \sigma_D - \frac{\mu_D}{\sigma_D} \Big). \tag{3.24}$$

For a Rayleigh fading channel, each element in $\mathbf{H}$ is a zero mean Gaussian random variable with variance 0.5 in each dimension. Hence, both the mean and the variance of the random variable $|\beta_{j,i}|^2$ are 1. As a result, the mean and variance of the random

variable $\sum_{j=1}^{n_r}\sum_{i=1}^{n_t}\lambda_i|\beta_{j,i}|^2$ can be represented as

$$\mu_D = n_r\sum_{i=1}^{r}\lambda_i \tag{3.25}$$

and

$$\sigma_D^2 = n_r\sum_{i=1}^{r}\lambda_i^2, \tag{3.26}$$

respectively. Therefore, (3.24) can be represented as

$$\begin{aligned}\mathcal{P}(\mathbf{X},\hat{\mathbf{X}}) \leq\ & \tfrac{1}{2}\exp\left(\tfrac{1}{2}\left(\tfrac{E_s}{4N_0}\right)^2 n_r\textstyle\sum_{i=1}^{r}\lambda_i^2 - \tfrac{E_s}{4N_0}n_r\sum_{i=1}^{r}\lambda_i\right)\\ & \cdot\mathcal{Q}\left(\tfrac{E_s}{4N_0}\sqrt{n_r\textstyle\sum_{i=1}^{r}\lambda_i^2} - \tfrac{\sqrt{n_r}\sum_{i=1}^{r}\lambda_i}{\sqrt{\sum_{i=1}^{r}\lambda_i^2}}\right)\end{aligned} \tag{3.27}$$

By using (3.16) and assuming that the SNR is high enough to ensure that $\frac{E_s}{4N_0} \geq \frac{\sqrt{n_r}\sum_{i=1}^{r}\lambda_i}{\sqrt{\sum_{i=1}^{r}\lambda_i^2}}$, (3.27) can be further upper bounded as

$$\begin{aligned}\mathcal{P}(\mathbf{X},\hat{\mathbf{X}}) \leq\ & \tfrac{1}{2}\exp\left(\tfrac{1}{2}\left(\tfrac{E_s}{4N_0}\right)^2 n_r\textstyle\sum_{i=1}^{r}\lambda_i^2 - \tfrac{E_s}{4N_0}n_r\sum_{i=1}^{r}\lambda_i\right)\\ & \cdot\tfrac{1}{2}\exp\left(-\tfrac{1}{2}\left(\tfrac{E_s}{4N_0}\right)^2 n_r\textstyle\sum_{i=1}^{r}\lambda_i^2\right)\\ =\ & \tfrac{1}{4}\exp\left(-\tfrac{E_s}{4N_0}n_r\textstyle\sum_{i=1}^{r}\lambda_i\right)\end{aligned} \tag{3.28}$$

This shows that the sum of all eigenvalues of $\mathbf{A}(\mathbf{X},\hat{\mathbf{X}})$, which is a square matrix, needs to be maximized over all codeword pairs in order to minimize the error probability. For a square matrix, the sum of all eigenvalues equals the trace of this matrix. From the definition of $\mathbf{A}(\mathbf{X},\hat{\mathbf{X}})$, the trace of $\mathbf{A}(\mathbf{X},\hat{\mathbf{X}})$ is also equal to the squared Euclidean distance between the pair of codewords $\mathbf{X}$ and $\hat{\mathbf{X}}$. Therefore, in order to minimize the error probability for a MIMO system with $rn_r \geq 4$, the Euclidean distance between codeword pairs or, equivalently, the trace of $\mathbf{A}(\mathbf{X},\hat{\mathbf{X}})$ should be maximized. This design criterion is referred to as the *trace criterion*.

For the case when $rn_r < 4$, it can be shown that, at high SNR, the error probability can be upper bound as

$$\mathcal{P}(\mathbf{X},\hat{\mathbf{X}}) \leq \left(\prod_{i=1}^{r}\lambda_i\right)^{-n_r}\left(\frac{E_s}{4N_0}\right)^{-rn_r} \tag{3.29}$$

This shows that, in order to minimize the error probability for a MIMO system with $rn_r < 4$, the minimum rank, r, and the minimum product, $\prod_{i=1}^{r} \lambda_i$, of matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ over all codeword pairs are required to be maximized. Note that $\prod_{i=1}^{r} \lambda_i$ is the absolute value of the sum of determinants of all the principle $r \times r$ cofactors of matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$. Therefore, these design criteria are referred to as the *rank and determinant criterion*. rn_r is called the *diversity gain*. The diversity gain is indicated by the slope of the performance curves [4]. This diversity gain can be traded for high transmission rate.

3.4 SPACE-TIME BLOCK CODES

In this section, we introduce a popular family of STCs, namely space-time block codes. They can be classified into three categories, namely orthogonal space-time block codes (OSTBCs), quasi-orthogonal space-time block codes (QOSTBCs) and non-orthogonal space-time block codes (NOSTBCs).

3.4.1 Orthogonal Space-Time Block Codes

OSTBCs were first introduced by Alamouti in [41] for a two transmit antenna MIMO system and, then extended to MIMO systems with more than two transmit antennas in [42]. The initial goal of this type of space-time coding was to achieve full diversity gain by using transmit diversity (multiple transmit antennas) instead of receive diversity (multiple receive antennas). Although using receive diversity and an associated combining scheme can achieve full diversity, it requires multiple radio frequency (RF) links at the receiver, which are not easily accommodated in relatively small mobile handsets [9]. It is also desirable that the decoding algorithm has a low computational complexity because the handset normally has limited data processing capability and power capacity. The Alamouti code is the simplest OSTBC and successfully achieves these goals. Of course, performance can be further enhanced if both transmit and receive diversity are used.

[4] These curves include bit error rate (BER) and symbol error rate (SER) and they are plotted against signal noise ratio (SNR).

An Alamouti encoder is designed to deal with symbols rather than bits. Therefore, the arriving bit stream is first modulated according to the chosen constellation and then pairs of modulated symbols are encoded to form a transmission matrix given by

$$\mathbf{X} = \begin{pmatrix} S_1 & -S_2^* \\ S_2 & S_1^* \end{pmatrix}, \tag{3.30}$$

where A^* denotes the complex conjugate of A. Matrix (3.30) shows that each code block consists of 2 symbols, S_1 and S_2, and spans 2 time slots. The symbols, S_1 and S_2, are transmitted from antenna 1 and 2 during the first time slot. The negative complex conjugate of S_2 and the complex conjugate of S_1 are transmitted from antenna 1 and 2 during the second time slot.

The *rate* of a STC is defined as the number of information symbols transmitted within one block divided by the number of time slots one block spans [3]. For example, the Alamouti code has rate $2/2 = 1$. Therefore, it is a full rate (rate=1) OSTBC. It is called orthogonal because the code design is based on orthogonality, which means the inner product (dot product) of two column vectors or row vectors is zero,

$$\mathbf{X}^1 \cdot \mathbf{X}^2 = -S_1 S_2 + S_2 S_1 = 0, \tag{3.31}$$

where $\mathbf{X}^i$ is the i^{th} column of (3.30). The distance matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ for the Alamouti code can be constructed as

$$\begin{aligned} \mathbf{A}(\mathbf{X}, \hat{\mathbf{X}}) = &\ \mathbf{B}(\mathbf{X}, \hat{\mathbf{X}})\mathbf{B}^H(\mathbf{X}, \hat{\mathbf{X}}) \\ = &\ \begin{pmatrix} |S_1 - \hat{S}_1|^2 + |S_2 - \hat{S}_2|^2 & 0 \\ 0 & |S_1 - \hat{S}_1|^2 + |S_2 - \hat{S}_2|^2 \end{pmatrix} \end{aligned} \tag{3.32}$$

Clearly, matrix (3.32) has full rank. Hence, it is clear that the Alamouti code can achieve full diversity gain for two transmit antennas. It can also be observed that the squared Euclidean distance, which is the trace of matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$, between any pair of codewords for a given constellation remains the same as an uncoded system using the same constellation. This means the Alamouti code cannot provide any coding gain.

The simulation results in [41] and [43] show that a MIMO system using the Alamouti code outperforms an uncoded MIMO system. This improvement is solely due to the increased diversity gain.

The orthogonality can be used not only to achieve a distance matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ with full rank, but also to enable the decoder to have a simple decoding algorithm. A joint maximum likelihood decoder based on (3.3) has to test M^2 codewords, where M is the number of constellation points, to find the estimated codeword for a MIMO system which transmits two information symbols within one block. This is because all the signals transmitted at the same time interfere with each other. The number of tests can be significantly reduced, especially, when a high order constellation is used, if the Alamouti code is used.

The decoding process for the Alamouti code is as follows. Assuming only one receive antenna is used, the received signals during the two time slots corresponding to one block are the superposition of faded versions of the transmitted signals corrupted by noise. They can be represented as

$$\begin{aligned} Y_1 &= h_{1,1}S_1 + h_{1,2}S_2 + N_1^1 \\ Y_2 &= -h_{1,1}S_2^* + h_{1,2}S_1^* + N_1^2 \end{aligned} \tag{3.33}$$

A maximum likelihood decoder uses the decision metric given by

$$\begin{aligned} & d^2(Y_1, h_1\tilde{S}_1 + h_2\tilde{S}_2) + d^2(Y_2, -h_1\tilde{S}_2^* + h_2\tilde{S}_1^*) \\ = \; & |Y_1 - h_1\tilde{S}_1 - h_2\tilde{S}_2|^2 + |Y_2 + h_1\tilde{S}_2^* - h_2\tilde{S}_1^*|^2 \end{aligned} \tag{3.34}$$

where $\tilde{S}$ is a test symbol from the set of all possible symbols. Substituting (3.33) into (3.34), the decision metric (3.34) can be represented as

$$(\tilde{S}_1, \tilde{S}_2) = \arg \min_{(\tilde{S}_1, \tilde{S}_2) \in \phi} (|h_1|^2 + |h_2|^2 - 1)(|\tilde{S}_1|^2 + |\tilde{S}_2|^2) + d^2(\bar{S}_1, \tilde{S}_1) + d^2(\bar{S}_2, \tilde{S}_2) \tag{3.35}$$

where ϕ is the set of all possible codewords, and $\bar{S}_i (i = 1, 2)$ are two decision statistics

given by

$$\begin{aligned} \bar{S}_1 &= h_1^* Y_1 + h_2 Y_2^* \\ &= (|h_1|^2 + |h_2|^2) S_1 + h_1^* N_1^1 + h_2 N_1^{2^*} \\ \bar{S}_2 &= h_2^* Y_1 - h_1 Y_2^* \\ &= (|h_1|^2 + |h_2|^2) S_2 - h_1 N_1^{2^*} + h_2^* N_1^1 \end{aligned} \tag{3.36}$$

The decision metric (3.35) can be further simplified as

$$\begin{aligned} \tilde{S}_1 &= \arg\min_{\tilde{S}_1 \in \phi} (|h_1|^2 + |h_2|^2 - 1)|\tilde{S}_1|^2 + d^2(\bar{S}_1, \tilde{S}_1) \\ \tilde{S}_2 &= \arg\min_{\tilde{S}_2 \in \phi} (|h_1|^2 + |h_2|^2 - 1)|\tilde{S}_2|^2 + d^2(\bar{S}_2, \tilde{S}_2) \end{aligned} \tag{3.37}$$

This shows that, by using the Alamouti code, the number of tests decreases to 2M. This indicates a significant reduction in the required processing capability of the receiver. The reason for this reduction is that the matrix (3.30) is an orthogonal matrix which has the following property

$$\begin{aligned} \mathbf{X} \cdot \mathbf{X}^H &= \begin{pmatrix} |S_1|^2 + |S_2|^2 & 0 \\ 0 & |S_1|^2 + |S_2|^2 \end{pmatrix} \\ &= (|S_1|^2 + |S_2|^2)^2 I \end{aligned} \tag{3.38}$$

This property enables the decoder to decouple the received signals by manipulating them according to (3.36).

The design of MIMO systems based on orthogonality was extended to more than two transmit antennas in [42]. Full rate (rate=1) OSTBCs using real signal constellations exist for MIMO systems with any number of transmit antennas, but $n \times n$ square transmission matrices can only be found for MIMO systems with 2, 4 or 8 transmit antennas. For MIMO systems using a complex signal constellation, OSTBCs with rate 1/2 exist for any number of transmit antennas. The Alamouti code is the only OSTBC which can achieve full rate with a complex signal constellation. There are two rate 3/4

OSTBCs given by

$$\begin{pmatrix} S_1 & -S_2^* & \frac{S_3^*}{\sqrt{2}} & \frac{S_3^*}{\sqrt{2}} \\ S_2 & -S_1^* & \frac{S_3^*}{\sqrt{2}} & -\frac{S_3^*}{\sqrt{2}} \\ \frac{S_3}{\sqrt{2}} & \frac{S_3}{\sqrt{2}} & \frac{(-S_1-S_1^*+S_2-S_2^*)}{2} & \frac{(S_2+S_2^*+S_1-S_1^*)}{2} \end{pmatrix} \tag{3.39}$$

and

$$\begin{pmatrix} S_1 & -S_2^* & \frac{S_3^*}{\sqrt{2}} & \frac{S_3^*}{\sqrt{2}} \\ S_2 & -S_1^* & \frac{S_3^*}{\sqrt{2}} & -\frac{S_3^*}{\sqrt{2}} \\ \frac{S_3}{\sqrt{2}} & \frac{S_3}{\sqrt{2}} & \frac{(-S_1-S_1^*+S_2-S_2^*)}{2} & \frac{(S_2+S_2^*+S_1-S_1^*)}{2} \\ \frac{S_3}{\sqrt{2}} & -\frac{S_3}{\sqrt{2}} & \frac{(-S_2-S_2^*+S_1-S_1^*)}{2} & \frac{(S_1+S_1^*+S_2-S_2^*)}{2} \end{pmatrix} \tag{3.40}$$

for MIMO systems with 3 or 4 transmit antennas, respectively.

The OSTBC given by (3.40) is considered in this thesis. Its performance is simulated and presented in Chapter 5 for comparison purposes. In order to avoid the estimation of CSI, there are some differential en/decoded STCs based on the orthogonal design proposed in [3] [44]. This is outside the scope of the present research.

In conclusion, OSTBCs can achieve full diversity gain and have a simple decoding algorithm. But they have the same minimum Euclidean distance as the uncoded system. Therefore, OSTBCs cannot provide coding gain. OSTBCs for systems with more than two transmit antennas and using complex constellations can not achieve full rate, which results in bandwidth expansion.

3.4.2 Quasi-Orthogonal Space-Time Block Codes

One of the drawbacks of OSTBCs is that their rate is less than 1 for MIMO systems with more than two transmit antennas and complex constellations. This is because OSTBCs maintain orthogonality in order to use a simple decoding algorithm. If the orthogonality constraint is relaxed to some extent, STBCs with rate equal to 1 or even greater than 1 can be obtained. This leads to another subfamily of STBCs, namely QOSTBCs. So far, many QOSTBCs have been proposed to achieve full rate(=1) or even rates greater than one with/without full diversity gain [45] [46] [47] [48]. They still keep some of the orthogonality properties to maintain the decoding complexity at an acceptable level.

In this section, some of these QOSTBCs are described.

The QOSTBC proposed in [45] achieves full rate for four transmit antennas. The columns of the transmission matrix are divided into groups. The columns in each group are not orthogonal to each other, but different groups are orthogonal to each other. In this scheme, the Alamouti code defined in (3.30) is used multiple times to form the transmission matrix. The QOSTBC of [45] is represented as

$$\mathbf{X} = \begin{pmatrix} A_{12} & A_{34} \\ -A_{34}^* & A_{12}^* \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & S_3 & S_4 \\ -S_2^* & S_1^* & -S_4^* & S_3^* \\ -S_3^* & -S_4^* & S_1^* & S_2^* \\ S_4 & -S_3 & -S_2 & S_1 \end{pmatrix} \tag{3.41}$$

where $A_{m,n}$ denotes an Alamouti code for the m^{th} and n^{th} symbol. The first column of (3.41) is orthogonal to the second and third column, but not the fourth column. Similarly, each of the other three columns is only orthogonal to two other columns. So, the first and fourth columns are in one group and the other two columns are in another group. Since this matrix is not fully orthogonal, a maximum likelihood decoder can no longer decode the four symbols separately. By manipulating the decision metrics

$$\sum_{t=1}^{4} \sum_{j=1}^{n_r} |Y_{j,t} - \sum_{i=1}^{n_t} h_{j,i} X_i^t|^2,$$

two decision formula can be obtained. Each of them is used to decode a pair of symbols. When a constellation with M points is used, the required number of tests is $2M^2$ rather than the $4M$ tests required by an OSTBC with rate 1 for 4 transmit antennas. Hence, the decoding complexity is increased. But this decoding complexity is still lower than the M^4 tests required by a joint maximum likelihood decoder. The diversity gain offered by this scheme can be examined by constructing the distance matrix. It can be shown that the minimum rank of the distance matrix is two, while full rank would be 4 [45].

There is another QOSTBC which can achieve full rate but not full diversity pro-

posed in [46]. It is described as

$$\mathbf{X} = \begin{pmatrix} A_{12} & A_{34} \\ A_{34} & A_{12} \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & S_3 & S_4 \\ -S_2^* & S_1^* & -S_4^* & S_3^* \\ S_3 & S_4 & S_1 & S_2 \\ -S_4^* & S_3^* & -S_2^* & S_1^* \end{pmatrix} \tag{3.42}$$

They both perform similarly. These QOSTBCs achieve full rate at the cost of having less diversity gain and more computational complexity. The simulation results in [45] show that the performance of the QOSTBC is better than the OSTBC over low SNRs, but worse at high SNRs. This is due to the fact that the diversity gain determines the slope of the BER-SNR curve.

A QOSTBC which can achieve full diversity and full rate for four transmit antennas was proposed in [47]. It achieves full diversity by using symbols from different constellations. These different constellations are obtained by rotating the original constellation by some angle, which is chosen to maximize the Euclidean distance between different codewords. It is shown in [47] that this QOSTBC achieves full rate and full diversity gain, but its decoder has higher computational complexity. The simulation results in [47] show that the BER-SNR curve of this QOSTBC has the same slope as the OSTBC. So, it can perform much better than the QOSTBC in [45] at high SNRs.

A STBC with rate greater than 1 is also possible by using the quasi-orthogonal concept. In [48], a QOSTBC which consists of two Alamouti codes transmitted from 4 transmit antennas is described. It achieves a rate of 2 (4 symbols in 2 time slots). Another QOSTBC with full rate and full diversity gain for 4 transmit antenna and quaternary phase-shift keying (QPSK) was proposed in [49]. Again, it trades increased computational complexity for higher code rate. Some other QOSTBCs are proposed in [50] [51] [52]. Generally, the QOSTBC family offers a trade off between the rate and decoding complexity.

3.4.3 Non-Orthogonal Space-Time Block Codes

A STBC with full rate or even higher rate can be easily obtained, if orthogonality is not required. This type of STBC is called a non-orthogonal STBC (NOSTBC) in this thesis.

In [53], a NOSTBC for 3 transmit antennas was proposed. It has rate 1 and diversity gain 2 and is defined as

$$\mathbf{X} = \begin{pmatrix} S_1 & S_2 & S_3 \\ S_2^* & -S_1^* & S_1 \\ -S_3^* & -S_3^* & S_2^* \end{pmatrix} \tag{3.43}$$

It is still based on the Alamouti code.

The NOSTBCs proposed in [12] are able to achieve much higher rates because they are designed using a different principle. These NOSTBCs are designed based on extended RS (eRS), extended BCH (eBCH) or shortened BCH (sBCH) codes, which were introduced in Chapter 2. The HARQ schemes proposed in this thesis are designed for these NOSTBCs. They are encoded using (n, k) eRS codes, eBCH codes or sBCH codes, where n is the length of the codeword and k is the number of information symbols. The code symbols are defined over GF(2), GF(4) or GF(16). In the GF(16) case, the binary information bits are mapped to $GF(16)$ symbols and then every k $GF(16)$ symbols are encoded to generate a codeword of the chosen eBCH, eRS or sBCH code. The encoded GF(16) symbols are then mapped to 16-QAM as [12]:

$$\begin{array}{cc|cc} 0 & 8 & 5 & 15 \\ 14 & 1 & 9 & 4 \\ \hline 7 & 13 & 2 & 10 \\ 11 & 6 & 12 & 3 \end{array} \tag{3.44}$$

After this mapping, the constellation points chosen by each codeword are written into the STBC matrix row by row. Each row of this matrix is transmitted from a different transmit antenna.

There are n_t symbols transmitted from the n_t transmit antennas simultaneously, but the rate of the NOSTBC must take the rate of the eRS, eBCH or sBCH code into account in order to make a fair comparison. Therefore, the rate of these NOSTBCs is calculated as

$$\mathcal{R}_{NOSTBC} = \mathcal{R}_{ecc} \cdot n_t, \tag{3.45}$$

where $\mathcal{R}_{ecc} = k/n$ denotes the rate of the error control code on which this NOSTBC is defined.

These NOSTBCs are designed using the trace criteria because signals coded in such a way can provide good Euclidean distance [12]. Therefore, these NOSTBCs suit a MIMO system with diversity gain of 4 or more ($rn_r \geq 4$). In order to ensure sufficient diversity gain, four receive antennas are used in [12] and this thesis. This is because these STBCs are not orthogonal. So, the distance matrix $\mathbf{A}(\mathbf{X}, \hat{\mathbf{X}})$ of these NOSTBCs does not have full rank. Using 4 receive antenna guarantees sufficient diversity gain, $rn_r \geq 4$.

NOSTBCs based on various eRS, eBCH and sBCH codes are considered in [12] to provide various rates ranging from 3/4 to 9/4. The optimal decoding algorithm for these codes is the joint maximum likelihood decoding algorithm given by (3.3). This algorithm has very high decoding complexity for these NOSTBCs, especially for high rate NOSTBCs. This is because high rate and using a constellation with relatively large numbers of points leads to a large set of possible codewords to test. Therefore, joint maximum likelihood decoding algorithm is not feasible for the high rate NOSTBCs of [12]. A list-based decoding algorithm, which has decreased decoding complexity, was proposed in [12]. This algorithm is suboptimal compared to the joint maximum likelihood decoding algorithm. The simulation results in [12] show these NOSTBCs offer significant performance gains compared to OSTBCs, but this is achieved at the cost of having higher decoding complexity.

Only the NOSTBC defined using the (16,3) eRS code is considered in this thesis, because we focus on the performance of the proposed HARQ schemes. This NOSTBC is now defined as follows. Three GF(16) symbols are obtained by mapping every 12 input

bits to GF(16) symbols. These symbols are encoded using the (15,3) RS code. One extra overall parity symbol is added to this codeword to form an eRS(16,3) codeword. This codeword is then mapped to 16-QAM according to (3.44). These 16 modulated symbols are written into a 4×4 transmission matrix row by row for a MIMO system with 4 transmit antennas.

This NOSTBC has relative low code rate (3/4). At such a low rate, the decoding complexity of the maximum likelihood decoding algorithm is acceptable. Therefore, a brute force maximum likelihood decoder is used to investigate the performance of the proposed HARQ schemes. The HARQ schemes proposed in Chapter 4 for the (16,3) eRS NOSTBC can also be used for all the other NOSTBCs proposed in [12].

3.5 OTHER SPACE-TIME CODES

We now quickly describe space-time trellis codes (STTCs) and layered space-time codes (LSTs) because many recent HARQ schemes are designed for them.

STTCs are another important subfamily of STCs. They were first introduced in [37]. They are an extension of the traditional trellis codes, introduced in Chapter 2, to MIMO systems [54]. The performance improvement is due to both increased diversity and Euclidean distance compared with the uncoded systems. Unfortunately, the decoding complexity of STTCs increases exponentially with the diversity and transmission rate [55].

A STTC encoder is now briefly introduced. Assuming a constellation with M points is used, the encoder generates n_t symbols for every $m_b = \log_2(M)$ information bits for each time slot. These n_t symbols are transmitted from n_t transmit antennas simultaneously. Each of these n_t symbols, $X_i \quad i = 1, \cdots, n_t$, at time T is given by [3]

$$X_i = \sum_{p=1}^{m_b} \sum_{l=0}^{v^p} g_{l,i}^p b_{T-l}^p \mod \mathrm{M}, \quad i = 1, 2, \cdots, n_t \tag{3.46}$$

where b_{T-l}^p is the p^{th} bit in a group of m_b bits at time $T-l$ and is stored in the l^{th} shift register of the path for the p^{th} bit stream, and $g_{l,i}^p$ and v^p are the coefficient of the l^{th}

shift register and the memory order of the path for the p^{th} bit stream, respectively. This encoder is illustrated in Fig 3.2 [3]. Because memory units are used in the encoder, a trellis diagram can be drawn according to the memory states associated with the inputs. Hence, STTCs can be decoded using the Viterbi decoding algorithm.

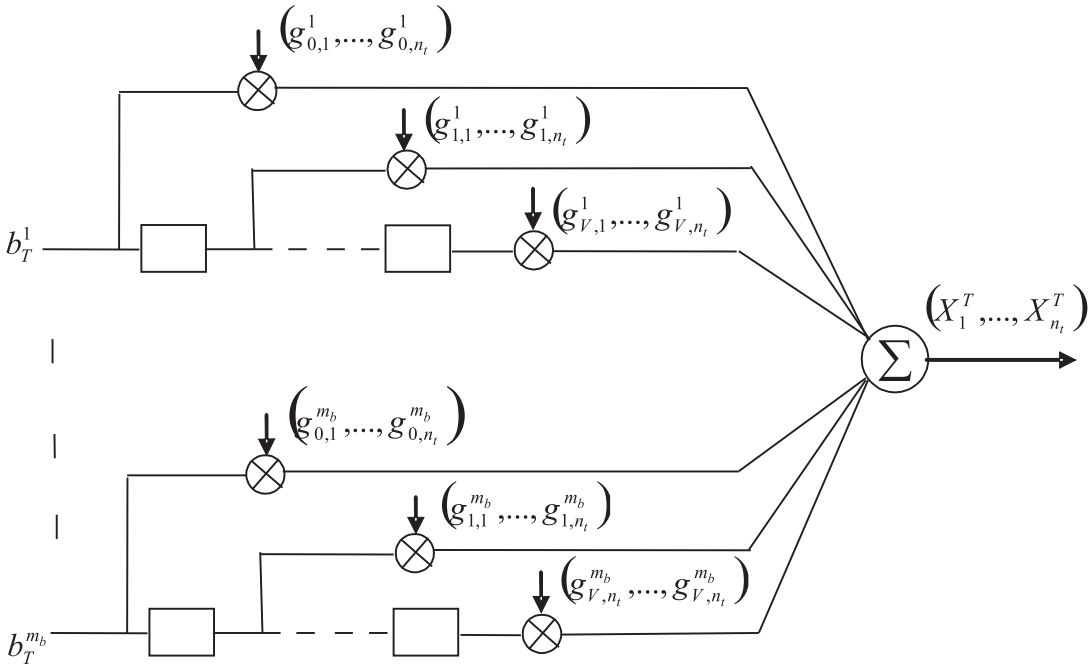

Figure 3.2 A STTC encoder [3].

Another type of MIMO system, which is quite different from all previously introduced MIMO systems, is the layered space-time (LST) architecture. This type of MIMO system requires the number of receive antennas n_r to be equal to or greater than the number of transmit antennas n_t. It can provide very high rate, but requires powerful signal processing technique such as successive interference cancellation employed by the receiver [3].

Here, the vertical layered space-time (VLST) scheme, is briefly introduced. It is used in the HARQ scheme of [56] which will be briefly considered in Section 4.1.1. In the VLST scheme, the information bit stream is demultiplexed into n_t substreams. These substreams are coded before they are mapped to a constellation. Alternatively, some

substreams may be left uncoded. Then they are transmitted from the corresponding transmit antennas as illustrated in Fig 3.3 [6]. The VLST scheme can achieve at most a diversity gain of n_r [6].

The decoding complexity can be significantly reduced by using *successive interference cancellation* instead of using joint detection. The decoding procedure is normally carried out by using interference suppression combined with interference cancellation. The interference suppression can be done by using QR decomposition interference suppression or interference minimum mean square error suppression. After the interference suppression is done, the n_t transmitted symbols are ordered in terms of their received SNR and the transmitted symbol with the highest received SNR is decoupled from the other $n_t - 1$ symbols. This symbol is estimated and a hard decision is made. Then, the interference caused by this symbol is subtracted out and the transmitted symbol with the second highest receive SNR is decoupled and estimated. A hard decision is made based on the estimation. This process is repeated many times until the decoding procedure is finished. Because each symbol is decoded individually, the decoding complexity is linear with the number of transmit antennas [6]. However, this decoding procedure indicates that, if an error is made for one symbol, all the decisions for the following symbols are likely to be erroneous.

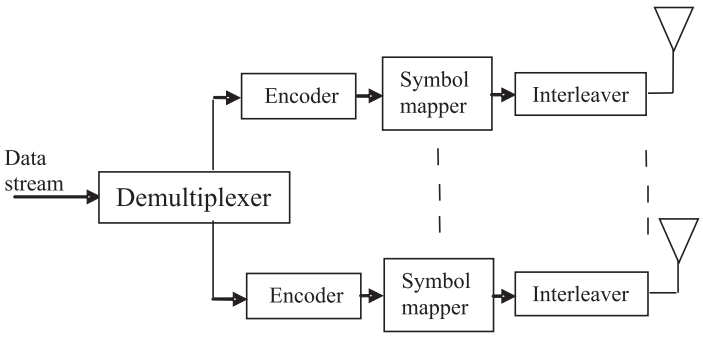

Figure 3.3 The VLST system of [3].

3.6 SUMMARY

In this chapter, MIMO wireless communication systems were introduced. They use multiple transmit antennas and multiple receive antennas to achieve larger channel capacity than SISO systems. They normally use specially designed STCs to achieve good performance. A good STC can be obtained if it is designed according to special design criteria for STCs [3] [6]. Two design criteria were discussed. The first one is called the trace criterion, which is used for MIMO systems with diversity gain ≥ 4 and the second one is called the rank and determinant criteria, which is used for diversity gain < 4.

STBCs including OSTBCs, QOSTBCs and NOSTBCs were described in detail in this chapter. OSTBCs have simple decoding algorithms and can achieve full diversity gains. But they cannot provide coding gains and full rate OSTBCs for MIMO systems with more than two transmit antennas do not exist. QOSTBCs can achieve higher rates than OSTBCs, but their decoding algorithms are more complicated. NOSTBCs can achieve even higher rates, but normally their decoding algorithms are even more complicated.

The good performance offered by MIMO systems has lead to their recent use with HARQ schemes. These existing HARQ schemes for MIMO systems will be introduced in Section 4.1. Three novel HARQ schemes will be proposed in Section 4.2. They use the NOSTBCs of [12] due to the good performance and high rate these NOSTBCs can offer.

Chapter 4

HARQ SCHEMES AND ADAPTIVE MODULATION FOR MIMO SYSTEMS

The design of hybrid automatic-repeat-request (HARQ) schemes for multiple input multiple output (MIMO) systems has recently attracted lots of research attention due to the good performance MIMO systems can achieve. HARQ schemes for MIMO communication systems can be designed by combining techniques introduced in previous chapters. Several recently proposed HARQ schemes for MIMO systems are summarized in this chapter. Some of them are motivated by the ARQ schemes for single input single output (SISO) systems introduced in Chapter 2 [13] [30] [36] [57]. In addition, some novel HARQ schemes for MIMO system are proposed in this chapter. We also discuss adaptive modulation and coding (AMC) in this chapter. It can be used to achieve better performance by reconfiguring the transmitter according to the current CSI. The current channel state information (CSI) is estimated by the receiver and fed back through a feedback channel.

This chapter is organized as follows. In section 4.1, a general summary of some recently proposed HARQ schemes and AMC schemes is presented. In Section 4.2, the novel HARQ schemes proposed in this thesis are described. In the last section, we introduce how to apply adaptive modulation based on the water filling principle to the proposed HARQ schemes.

4.1 ARQ PROTOCOLS AND ADAPTIVE MODULATION FOR MIMO SYSTEMS

ARQ schemes designed for SISO systems were introduced in Chapter 2. The ideas used in ARQ schemes designed for SISO systems can be applied to MIMO systems.

The fundamental concept behind the ARQ schemes designed for MIMO systems is, under the assumption of independent fading channel realizations for each transmission, to explore the time diversity provided by the ARQ schemes to enhance the overall diversity of the MIMO system at the cost of spectral efficiency (multiplexing gain) [58].

All the basic ARQ schemes, stop-wait, go-back-N, selective-repeat and truncated selective-repeat ARQ (TS-ARQ) [18] can be used for MIMO systems directly without any modification. The HARQ schemes designed for SISO systems using hard algebraic decoders can be easily modified for MIMO systems by being concatenated with a space-time code (STC). But the research conducted in the existing literature [13] [36] [57] and in this thesis shows that if the ARQ schemes are designed jointly with the MIMO systems and the STCs, then the performance can be improved significantly. An ARQ MIMO system is shown in Fig.4.1.

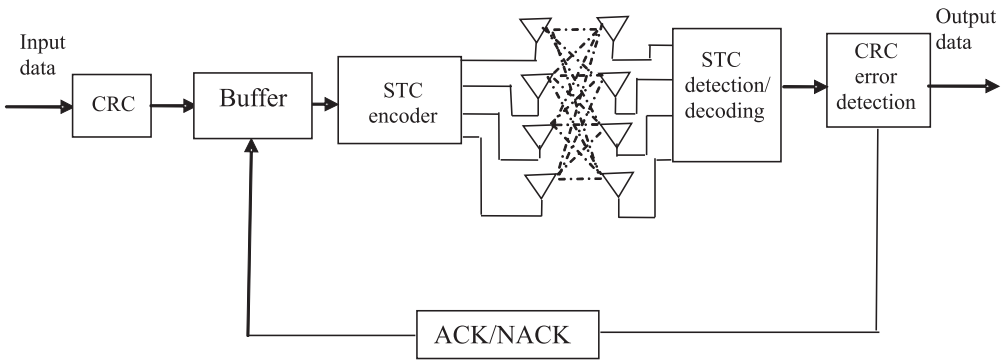

Figure 4.1 A MIMO ARQ system model.

In this section, some recently proposed HARQ schemes designed for MIMO systems are described. Simulation results show that these jointly designed HARQ-space-time coding schemes can offer significant gains compared to using space-time codes (STCs) with SISO ARQ schemes directly. So far, most of the proposed HARQ schemes are designed for MIMO systems with orthogonal space-time block codes (OSTBCs) [14], space-time trellis codes (STTCs) [36], turbo codes [15] [16] and layered space-time (LST) systems [56].

Normally, for an ARQ scheme, it is assumed that CSI is not available at the transmitter. If the current CSI is allowed to be fed back to the transmitter via a feedback channel, the transmitter can be reconfigured in such way that certain goals, for example, having more channel capacity or higher transmission rate, can be achieved. Adaptive modulation is one technique used to achieve this. Because an ARQ system must have a feedback channel to feed back the ACK/NACK to the transmitter, it is natural to consider ARQ and adaptive modulation together to further improve performance. So, adaptive modulation is also very briefly discussed in this section.

4.1.1 ARQ Protocols for MIMO Systems

Generally, in an ARQ scheme, if the error detection and correction capabilities are provided by different error control codes, the information bits should be encoded using an error detection code first, for instance a CRC code. Then the coded bit stream from the error detection encoder is encoded using an error correction code. If the CRC check fails, a retransmission is requested. Other error detection codes exist, which have better error detection capability than that of CRC codes, including extended Hamming codes and extended binary Golay codes [7]. But they are not convenient to be used with ARQ schemes for the following reasons. Firstly, these code place constraints on packet length. Secondly, these codes have much lower rate than a CRC code. Thirdly, a well chosen CRC code can offer an error detection capability close to these optimal ones. Therefore, in most of the existing literature and this thesis, CRC code is used for error detection purpose. It is also assumed that this CRC check is perfect in most of the existing literature (not in this thesis).

The HARQ schemes in the existing literature using space-time block codes (STBCs) focus on the use of OSTBCs. A HARQ scheme designed for OSTBCs was proposed in [14]. In this scheme, information packets, which are already CRC encoded, are encoded using a rate $1/R_c$ convolutional code. Therefore, a k bit CRC encoded information packet becomes an encoded packet with $R_c k$ bits. These $R_c k$ bits are then modulated using binary phase shift keying (BPSK) and encoded using an OSTBC encoder. Then, this STC encoded packet is transmitted using n_t antennas. In [14], the

Alamouti code was used.

The combining scheme used in [14] is now described. Assuming a packet is retransmitted N-1 times and the CRC check for each transmission has failed, the N received copies will be combined to form a single packet in the following way. Each bit of the combined packet is the hard decision from a quantizer which calculates the average value of N received symbols for each bit and makes a decision according to a threshold. Assuming that the a priori probabilities for 0 and 1 are equal, this threshold is 0 due to the use of BPSK modulation. If the average value for a bit is very close to the threshold, that bit is regarded as unreliable and will be punctured from the combined packet. Then, the combined packet with only reliable bits is decoded using the Viterbi algorithm. The Hamming distance is calculated as the metric for each branch. It is shown in [14] that, although puncturing the unreliable bits decreases the Hamming distance between different codewords, the performance can still be enhanced for the AWGN channel or fading channel due to removing the unreliable bits. This scheme has low decoding complexity, but it can only offer very low spectral efficiency (< 1 bit/s/Hz). Applying larger modulations to this scheme is not a trivial exercise.

The advantage of using OSTBCs is that they have very low decoding complexity. But most of the HARQ schemes designed for OSTBCs use them in conjunction with other codes in order to obtain coding gain. Recall that the OSTBC does not have coding gain. The use of other codes decreases the rate, which is already limited by the use of OSTBCs.

Convolutional codes, trellis coded modulation (TCM) and STTCs have recently attracted a lot of attention for MIMO ARQ systems [13] [36] [57]. A HARQ scheme designed for a recursive convolutional code was proposed in [13]. It uses the Turbo code concept. In this scheme, two transmit antennas and BPSK are used. The encoder uses a rate 0.5 recursive convolutional code. The information packets, which are already CRC encoded, are encoded using this encoder. Therefore, during one time slot, two coded bits will be generated for one bit of the packet. Then, these two bits are modulated using BPSK and transmitted from two transmit antennas. If a NACK is received, the same packet is interleaved and then encoded using the same recursive convolutional code

before being transmitted. The non-interleaved and interleaved versions are transmitted alternately, if retransmission is requested. All non-interleaved and interleaved copies are combined with non-interleaved and interleaved copies, respectively. Then they are decoded using the Turbo principle and iterative decoding. Simulation results show that this HARQ scheme can offer a significant improvement compared with the traditional TS-ARQ scheme using the same recursive convolutional code. But the drawback of this scheme is that the good performance is achieved using BPSK. This results in very low spectral efficiency.

Turbo codes are used in the HARQ schemes proposed in [15] and [16]. In [15], CRC coded packets are systematically encoded using a rate 1/3 Turbo code. Then, the generated parity bits and original CRC coded information bits are divided into two groups by puncturing. One group is transmitted immediately and the other is stored for retransmissions. The group of bits being transmitted are mapped to two antennas using the Alamouti code. This scheme uses BPSK. It was extended in [16] by using another Turbo code with rate 1/5. Due to the use of BPSK and the Alamouti code, the transmission of one information bit needs at least two time slots. So, the maximum achievable spectral efficiency is 0.5 bit/s/Hz, which is very low compared to the spectral efficiencies offered by the HARQ schemes proposed in Section 4.2.

A HARQ scheme using TCM was proposed in [57]. In this scheme, 16-PSK is used and the arriving information bits/symbols are split into n_t substreams. Each substream is CRC encoded and then TCM encoded, before they are transmitted. These n_t substreams are spatially interleaved to break up bursts of errors introduced by fading. The Viterbi algorithm is used by the decoder. If retransmission has been requested, all copies for the same information packet are stored and combined using Chase combining (as described in Chapter 2) before the decoding is carried out.

STTCs are also used extensively with HARQ schemes. The scheme proposed in [30] for SISO systems was extended to MIMO systems using STTCs in [36]. In this HARQ scheme, different STTCs are used for each retransmission in order to improve performance. A measure of the coding gain offered by a STTC was proposed. It was used to find an optimal STTC for each retransmission. The simulation results show that this

HARQ scheme outperforms a HARQ scheme using a fixed STTC (optimal for the first transmission) for all transmissions.

The last HARQ scheme introduced in this section is one proposed for LST systems [56]. In this scheme, each input packet is split into several substreams and each substream is encoded using a CRC code. These CRC coded subtreams are then encoded using other error control codes which provide coding gain. Each encoded substream is interleaved separately before it is transmitted from a single transmit antenna or a group of antennas. All substreams are transmitted simultaneously. The receiver decodes the received information substream by substream. It uses the CRC code attached to each substream to validate the contents of the substream. If a substream passes its CRC check, the receiver will then perform interference cancellation to remove the interference due to this substream. If the CRC check for a particular substream fails, a NACK for this substream is fed back and the corresponding substream is retransmitted. This scheme is more efficient and reliable compared to using the traditional TS-ARQ scheme for LST systems, in which the CRC encodes the entire packet. This is because, in the scheme of [56], substreams rather than packets get retransmitted and each substream is checked based on its CRC parity bits which results in a more reliable interference cancellation for this substream.

All the HARQ schemes discussed in this section suffer a performance loss when the MIMO channel is temporally correlated. They use error correction codes and powerful combining scheme in order to achieve good performance. The HARQ schemes designed for STTCs/TCM suggest that using different codes for each transmission may lead to good performance.

In this thesis, we focus on the HARQ design for STBCs. So, we have to choose a good STBC first. OSTBCs have simple decoding algorithms, but they cannot offer coding gain. Therefore, other error control codes should be used to provide error correction capability. But this limits the achievable rate. Quasi-orthogonal space-time block codes (QOSTBCs) with full diversity gain require the use of a rotated constellation [47], which leads to a smaller Euclidean distance between adjacent constellation points. QOSTBCs, which cannot achieve full diversity gain, are outperformed by OS-

TBCs at high SNRs [45]. This leads to higher dropped packet rate and higher average number of retransmissions for them at high SNRs. So, although they have higher code rate, the spectral efficiency of the HARQ system is degraded as well at high SNRs. This motivates us to use the non-orthogonal space-time block codes (NOSTBCs) proposed in [12] which have good error correction capability and offer higher rates. We design powerful HARQ schemes for them in Section 4.2 in order to achieve good performance and rate simultaneously.

4.1.2 Adaptive Modulation

If the transmitter can be reconfigured according to the current CSI, then more robust communication or a higher transmission rate can be achieved. This requires that the feedback channel be used to feedback more information with negligible number of errors and delay. Parts of the transmission scheme such as transmit power, modulation or coding are then adapted relative to the current CSI. This is referred to as adaptive modulation and coding (AMC) [6]. Although an AMC scheme can be designed to adaptively change many components of the transmission scheme simultaneously, it has been shown in [59] that adaptively changing one or two components yields close to the maximum spectral efficiency obtained by adaptively changing all components.

There are some practical issues associated with AMC. For example, if the channel is changing so rapidly that it can not be estimated reliably, or if the feedback channel is not error free and has significant delay, then AMC may not perform as well as expected [6]. Also, due to hardware limitations, only finite levels of power or rate are available for use by the AMC. This imposes some constraints on the achievable performance.

The water filling principle can be used to adaptively change the transmit power for each antenna such that the achievable channel capacity is maximized as introduced in Section 3.2. This principle requires infinitive levels of power to be available in the transmitter, which is not practical. But it can serve as the best case scenario. This principle is studied with the HARQ schemes proposed in this thesis.

A discrete rate adaptive modulation schemes was combined with an ARQ scheme

for a SISO system in [4]. In this scheme, the transmitter uses different constellations depending on the current CSI and stops transmission when the CSI is worse than a preset threshold. This scheme uses the TS-ARQ scheme and a set of constellations, of which each symbol can be used to modulate different numbers of bits ranging from 1-7 bits. The packet error rates for all different constellations were derived. They are functions of SNR. The allowed number of retransmissions[5], $m_{max} - 1$, is decided by the maximum allowed delay time for this application and the time for a round trip.

The dropped packet rates for different constellations can be calculated based on the packet error rate functions. A target dropped packet rate is specified according to the application under consideration. By inverting these dropped packet rate functions, a set of SNR values are obtained, which are used as the thresholds for choosing the constellation. The transmitter chooses a particular constellation to modulate the information bits by comparing the current CSI with the calculated set of thresholds. This AMC scheme can achieve a high average transmission rate while maintaining the required reliability. This scheme was extended to MIMO systems using the Alamouti code in [60] [61] [5].

The rate adaptive systems can also be designed based on the use of different FEC codes with different rates. Another rate adaptive scheme which only uses the past NACK/ACK was proposed in [62]. In this scheme, if all the past packets are positively acknowledged, the system increases the transmission rate. It assumes that the channel is constant over one packet and changes slowly over packets. This scheme feeds back the same amount of information as an ARQ system.

Normally, the design goal for rate adaptive AMC systems is different from that of the HARQ schemes introduced in the previous section and the HARQ schemes proposed in this thesis. The rate adaptive AMC schemes aim to improve the transmission rate. When the CSI is such that the target dropped packet rate can still be achieved even when the packets are transmitted at a higher rate, then the system uses a different transmission scheme (for example, using different modulation or error control codes) to increase the rate. Doing this results in a performance loss in terms of dropped

[5]It is set to three in [4].

packet rate compared with using only one constellation. But this is acceptable for some applications, which only require a certain dropped packet rate to be maintained. Hence, for rate adaptive AMCs, the spectral efficiency increases as the SNR increases while the dropped packet rate curve becomes flat. For HARQ schemes, the transmission scheme does not change with the CSI. It aims to achieve as low a dropped packet rate as possible. Therefore, the dropped packet rate decreases as the SNR increases while the spectral efficiency curve becomes flat. So, rate adaptive AMC schemes can not be compared with HARQ schemes directly.

For MIMO systems, the CSI can also be used to select a subset of transmit antennas to be used for transmission. This is done to minimize the instantaneous error probability, when not all transmit antennas transmit signals simultaneously [63] [64]. In [65], a hybrid selection scheme was proposed, which selects the best three antennas and one of two STBC candidates. The simulation results show that the performance can be improved by using the best transmit antennas and different codes for different CSI. In this thesis, it is assumed that the transmitter uses the exact number of transmit antenna required by the MIMO system. Therefore, antenna selection is not considered here. But it could be considered in future work.

4.2 PROPOSED HARQ SCHEMES

The goal of this thesis is to design HARQ schemes coupled with STCs to simultaneously offer high throughput and high reliability wireless network access. To achieve this purpose, the NOSTBCs of [12], which were described in Chapter 3, are used. They offer high rate and good performance. Several novel HARQ schemes designed for these NOSTBCs are proposed in this section. Some of the proposed HARQ schemes can also be applied to other space-time coding schemes. Because the NOSTBCs of [12] only provide error correction capability, the information bit stream needs to be CRC encoded first for error detection purposes as mentioned in Section 4.1.1. So, for all three HARQ schemes proposed here, the packet consists of information bits and CRC parity check bits.

The three HARQ schemes proposed in this thesis are briefly summarized below. Full descriptions of each scheme are given in the next three subsections.

1. HARQ-1: Only the parts of the previously received signals corresponding to the best CSI are kept to be combined with the current received signals.

2. HARQ-2: All previously received signals and corresponding CSI are kept to be combined with the current received signals. This has the effect of increasing the number of virtual receive antennas with each retransmission [66].

3. HARQ-3: For a given packet, the data is encoded using a systematic encoder and non-systematic encoder for odd and even numbered transmissions, respectively. The receiver combines the received soft information in a similar manner to HARQ-2.

4.2.1 HARQ-1 Scheme

Now we describe the HARQ-1 scheme in detail. For each transmission, the received symbols are decoded independently from previous transmissions. If the CRC fails, the current received soft information is combined with information from previous transmissions in the following manner:
For each frame, we calculate the total channel gain $\sum_{i=1}^{n_t} |h_{j,i}^t|^2$ from all transmit antennas to the j^{th} receive antenna for $j = 1, 2, \cdots, n_r$. Then, for each receive antenna, we compare this gain with the gain calculated during the previous transmissions. A new "combined channel matrix", $\mathbf{a}$, is formed, where the channel coefficients of each row equal the coefficients of the channel matrix for the transmission which had the biggest gain over these coefficients. The values in the j^{th} row, $\mathbf{a}_j^t$, for time slot t after m ($m \leq m_{max}$) transmissions can be calculated using

$$\mathbf{a}_j^t = \arg \max_{\mathbf{h}_j^\gamma, \gamma=1,2,\cdots,m} \left(\sum_{i=1}^{n_t} |h_{j,i}^{t^\gamma}|^2 \right) \quad \text{for } j = 1, 2 \cdots n_r, \tag{4.1}$$

where γ is the transmission index and $\mathbf{h}_j^m$ is the j^{th} row of the channel matrix for the m^{th} transmission. A combined received symbol matrix, $\mathcal{S}$, is also constructed. Each

row of $\mathcal{S}$ contains the received signals corresponding to the transmission chosen for $\mathbf{a}_j^t$. The combined channel matrix and received symbol matrix, $\mathbf{a}$ and $\mathcal{S}$, are then passed to the NOSTBC decoder and CRC decoder. If the CRC check still fails and $m < M_{max}$, these two matrixes are stored in a buffer for use with further transmissions of the same packet. A flow chart of HARQ-1 is shown in Fig. 4.2.

4.2.2 HARQ-2 Scheme

The proposed HARQ-2 scheme is now described. For each transmission, the received symbols are initially decoded without using information from previous transmissions. If the CRC fails, the current received soft information will be combined with the information from previous transmissions in the following manner:
The current received symbols and channel matrix will be treated as those for an additional n_r receive antennas. In other words, the receiver acts as if it has mn_r virtual receive antennas after m transmissions of the same information. This concept was used in a non-ARQ context in [66]. After m transmissions, the combined $(m \cdot n_r) \times n_t$ channel matrix is denoted $\mathbf{a}$ and the $(m \cdot n_r) \times T$ combined received symbol matrix is denoted $\mathcal{S}$, where T is the number of time slots for each frame. Recall, we are assuming a quasi-static channel model and so the same $n_r \times n_t$ channel matrix is used for an entire frame. We can write the combined received symbol matrix as

$$\mathcal{S} = \mathbf{a}\mathbf{X} + \mathcal{N} = \begin{bmatrix} \mathbf{Y}^m \\ \vdots \\ \mathbf{Y}^1 \end{bmatrix} = \begin{bmatrix} \mathbf{H}^m \\ \vdots \\ \mathbf{H}^1 \end{bmatrix} \mathbf{X} + \begin{bmatrix} \mathbf{N}^m \\ \vdots \\ \mathbf{N}^1 \end{bmatrix}, \tag{4.2}$$

where the superscript on $\mathbf{Y}, \mathbf{H}, \mathbf{X}$ and $\mathbf{N}$ represents the transmission number. HARQ-2 has a similar flow chart to HARQ-1 as shown in Fig. 4.2. Both schemes can be used with a wide range of STCs including OSTBCs, QOSTBCs, NOSTBCs and STTCs.

A modified HARQ-2 scheme is also proposed in this thesis. It is designed to improve performance under a very slow time varying channel or, in other words, a highly time correlated channel. In this scheme, the transmitted NOSTBC block for each transmission is generated by cyclic shifting the eRS/eBCH/sBCH codeword. The sim-

ulation results verify that this modification can improve performance when the channel is slowly varying in time. This is because the symbols are transmitted from different transmit antennas during each retransmission. This scheme may be represented as

$$\mathcal{S} = \mathbf{aX} + \mathcal{N} = \begin{bmatrix} \mathbf{Y}^m \\ \vdots \\ \mathbf{Y}^1 \end{bmatrix} = \begin{bmatrix} \mathbf{H}^m & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{H}^1 \end{bmatrix} \begin{bmatrix} \mathbf{X}^m \\ \vdots \\ \mathbf{X}^1 \end{bmatrix} + \begin{bmatrix} \mathbf{N}^m \\ \vdots \\ \mathbf{N}^1 \end{bmatrix}, \tag{4.3}$$

where $\mathbf{X}^m$ denotes the NOSTBC block obtained by cyclic shifting the eRS/eBCH/sBCH codeword m times.

4.2.3 HARQ-3 Scheme

Another novel HARQ scheme, HARQ-3, is now proposed. It is based on HARQ-2 and is specifically designed for the NOSTBCs of [12], which are defined using eRS, eBCH or sBCH codes. Both systematic and non-systematic eRS/eBCH/sBCH encodings are used in HARQ-3. In [12] only systematic encoding was used in order to reduce the decoding complexity. In order to keep decoding complexity manageable for higher rate codes, the systematic codes are used for the first transmission in HARQ-3. Then even numbered transmissions of the same packet use non-systematic encodings. The codewords considered by the decoder for the non-systematic transmissions are narrowed by using information from the odd numbered transmissions. This scheme is described as follows:

The first transmission of a given packet uses systematically encoded eRS, eBCH or sBCH codewords to define the NOSTBC. When decoding the first transmission, a list of probable information vectors, called the "decoding candidate list", is generated based on the calculated squared Euclidean distance and stored for future use. It is a list of possible information vectors and the corresponding squared Euclidean distance between the hypothesized received signal matrix (based on the systematic codewords) and the actual received signal matrix. If the CRC check fails, the same information bits are encoded by the transmitter using a non-systematic encoder and this non-systematic version is transmitted during the second transmission. It is decoded independently

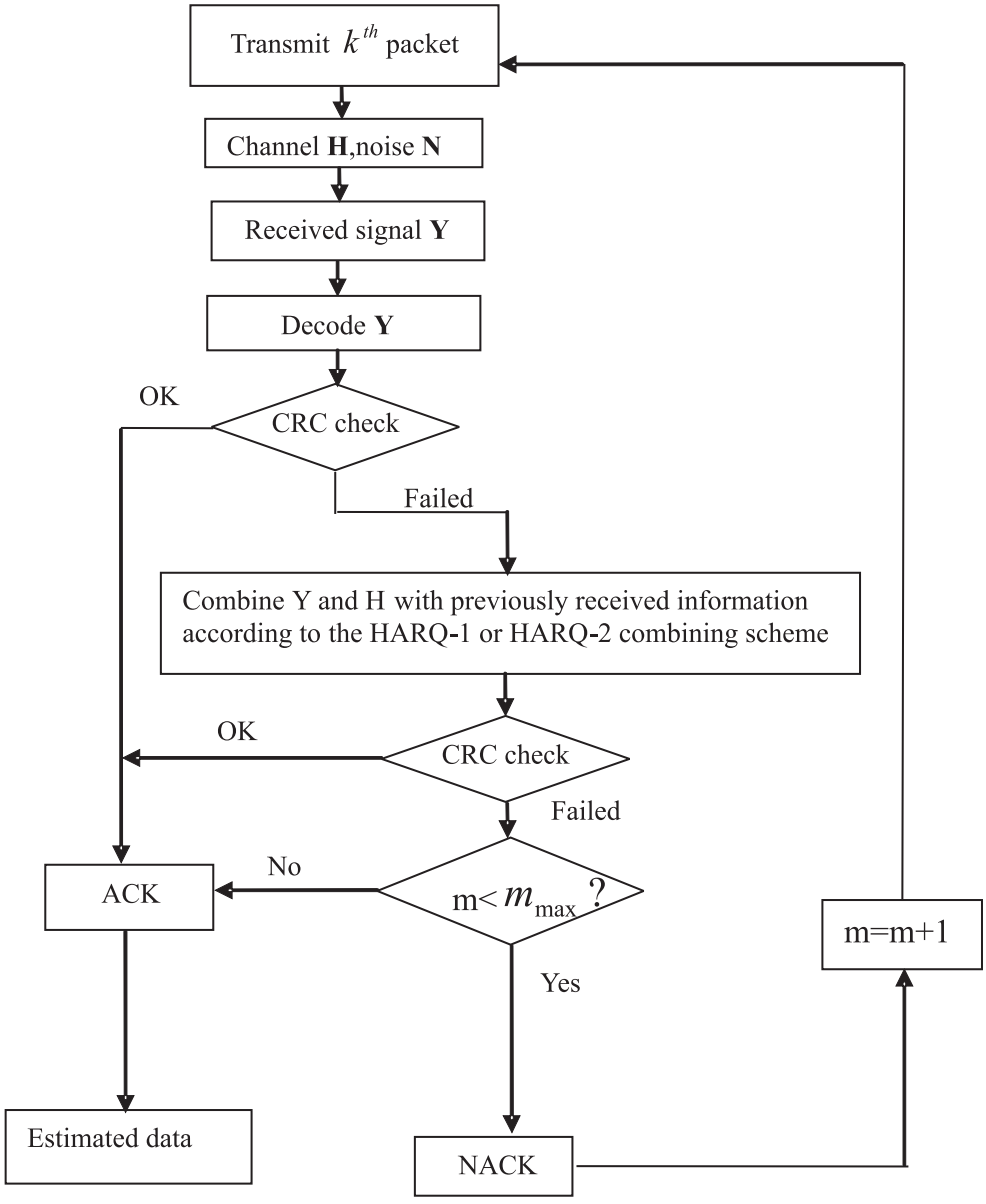

Figure 4.2 Flow chart for the HARQ-1 and HARQ-2 schemes.

from the first transmission. The decoder only considers codewords corresponding to information vectors in the decoding candidate list rather than to all possible information vectors. The squared Euclidean distance calculated during the first decoding is added to the calculated distance from the second decoding to give an overall distance measure. This is done to combine the information gained from the first two transmissions and to reduce the decoding complexity. If the CRC check still fails, the original systematically encoded codeword will be retransmitted (transmission 3) and the received symbols and channel matrix will be combined with those from the first transmission as in HARQ-2. This combined information will be decoded to generate a new decoding candidate list with corresponding Euclidean distance. This list is more reliable than the first one. The final decoded result from the third transmission is obtained by decoding the received symbols for the second transmission again using the new decoding candidate list. This candidate list will be stored in memory if the CRC check fails and further retransmission is required. During the fourth transmission, the non-systematic encoded codeword is retransmitted. It is combined in the receiver with the second transmission as in HARQ-2. It then uses the decoding candidate list generated from the third transmission.

Generally, for a given packet, the systematically encoded codewords will be transmitted during odd numbered transmissions and non-systematically encoded codewords will be transmitted during even numbered transmissions. A decoding candidate list and the associated Euclidean distance are generated during odd numbered transmissions and will be used for the decoding of all transmissions after the first transmission. Using a decoding candidate list reduces the detection/decoding complexity for this ARQ system using the NOSTBCs in [12]. A flow chart for the HARQ-3 scheme is shown in Fig. 4.3

4.2.4 Comparison to Existing MIMO ARQ Schemes

In this section, we will summarize the difference between the proposed HARQ schemes and other existing HARQ schemes. The HARQ schemes proposed in the previous section are quite different from the HARQ schemes introduced in Section 4.1. The

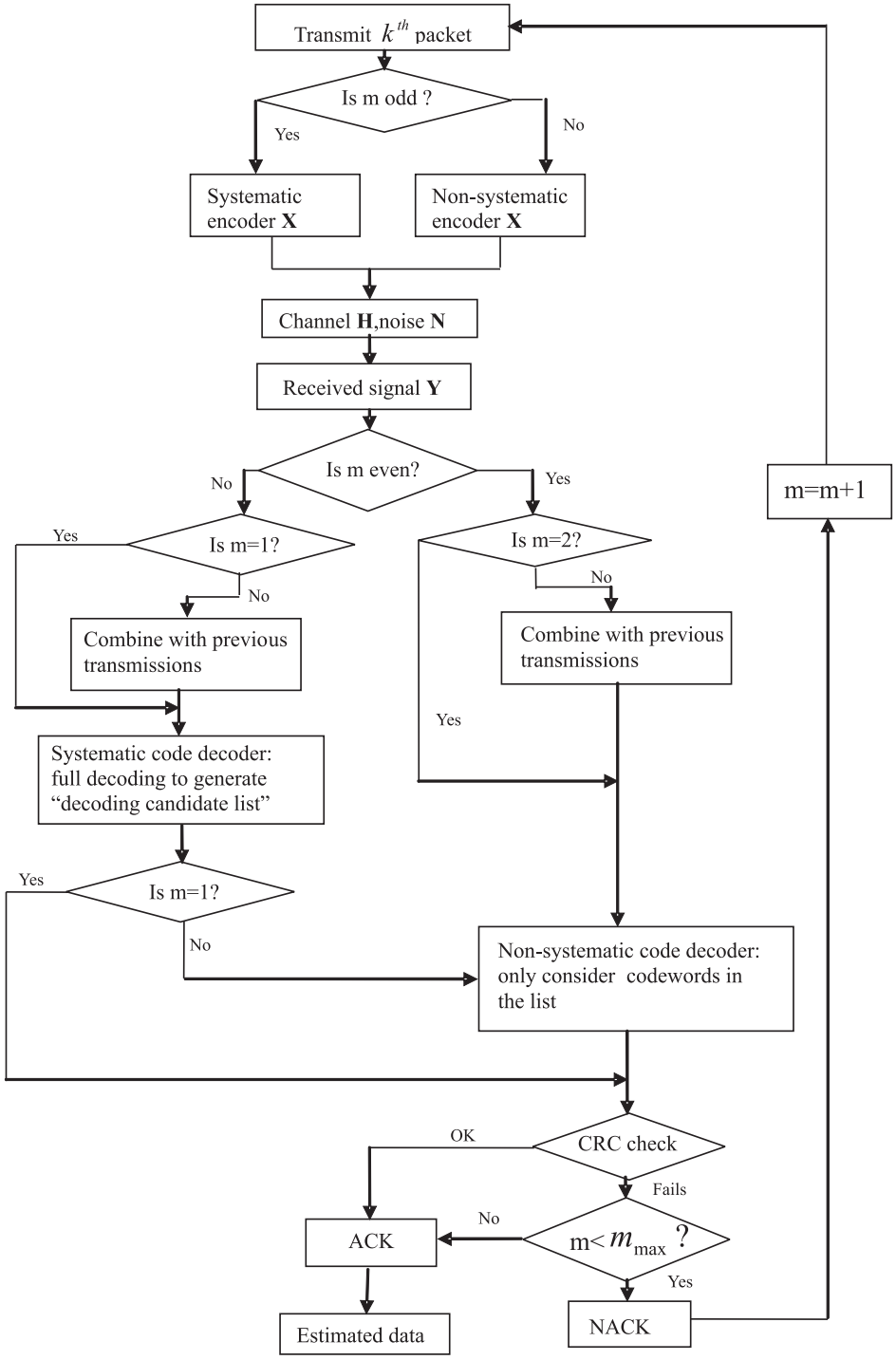

Figure 4.3 Flow chart for the HARQ-3 scheme.

main difference is the STCs, for which the HARQ schemes are designed. Most of the schemes proposed for MIMO systems use convolutional codes [13], TCM [57] or STTCs [36]. This is because these codes can provide good coding gain and offer error correction capability across the entire packet.

Schemes using STTCs and TCM normally conduct a code search to find good codes [57] [36]. As a comparison, the HARQ-3 scheme uses two encoding methods which are natural for all cyclic codes. Also, for STTCs and TCM with high rates and large constellations, the decoding complexity is normally higher than the schemes using STBCs.

The scheme of [13] uses the Turbo code concept. In this scheme, a packet and its interleaved version are encoded using a recursive convolutional code to obtain two encoded packets. They are transmitted alternatively if a retransmission is requested. The main drawback of the scheme of [13] is that it uses BPSK to modulate the coded bit stream. Therefore, the good performance shown in [13] is achieved at a very low spectral efficiency (< 1 bit/Hz/s). Comparing the simulation results in [13] with the simulation results in Chapter 5, it can be seen that the HARQ-3 scheme can achieve much higher spectral efficiency while having a better dropped packet rate performance. The HARQ schemes of [15] [16] are also based on Turbo codes. They use BPSK and Turbo codes concatenated with the Alamouti code. This results in a very low (≤ 0.5 bits/s/Hz) maximum achievable spectral efficiency.

The HARQ scheme proposed in [14] uses the Alamouti code with a convolutional code and a combining scheme. BPSK is used to modulate the bits in this scheme. The use of BPSK can provide good Euclidean distance between different symbols. The use of the Alamouti code results in a simple STC decoder. But they limit the spectral efficiency this HARQ scheme can offer. The combining scheme in [14] is designed based on BPSK, which means a larger constellation cannot be directly applied to this scheme.

A HARQ scheme has to use an error control code to provide coding gain in order to achieve good performance. So, the schemes, which use OSTBCs, have to use additional codes to provide coding gain because OSTBCs can not offer any coding gain. This limits the rates which can be achieved by these HARQ schemes since the rates of the OSTBCs

are less than or equal to 1 and using additional codes further degrades the rate.

As mentioned in Section 4.1.1, QOSTBCs are not considered in this thesis for the following reasons. QOSTBCs, which cannot achieve full diversity gain, are outperformed by OSTBCs at high SNRs [45]. This means that they will have higher dropped packet rate than the same scheme using OSTBCs. This leads to a higher average number of retransmissions, which degrades the rate of these HARQ schemes, at high SNRs. Also, they have higher decoding complexity than OSTBCs. The QOSTBCs in [47], which can achieve full diversity gain, use a rotated constellation. We do not consider rotated constellations in this thesis. To the best of our knowledge, no HARQ scheme specifically designed using QOSTBCs has been proposed in the literature to date.

As a comparison, the NOSTBCs proposed in [12] are good codes for ARQ design due to the large coding gain and the powerful combining scheme proposed in this thesis. In the HARQ schemes proposed in this thesis, the CRC coded information packets are only encoded using these NOSTBCs. Therefore, they can offer high throughputs. The simulation results presented in Chapter 5 show that using these NOSTBCs and the proposed HARQ schemes result in good performance at high rates.

4.3 THE PROPOSED HARQ SCHEMES USING THE WATER FILLING PRINCIPLE

Adaptively allocating the transmit power among all transmit antennas according to the water filling principle defined by (3.12) can further improve the performance of the proposed HARQ schemes. Therefore, it is also simulated in this thesis. This adaptive modulation scheme is applied to all proposed HARQ schemes for a quasi-static fading channel. The columns of the channel matrix $\mathbf{H}$ corresponding to each transmit antenna will be multiplied by the square root of the power calculated for each transmit antenna

to form a new channel matrix

$$\grave{\mathbf{H}} = \begin{bmatrix} \times\sqrt{P_1} & \times\sqrt{P_2} & \cdots & \times\sqrt{P_{n_t}} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ h_{1,1} & h_{1,2} & \cdots & h_{1,n_t} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r,1} & h_{n_r,2} & \cdots & h_{n_r,n_t} \end{bmatrix} \tag{4.4}$$

All HARQ schemes still work with the new channel matrix $\grave{\mathbf{H}}$. If any transmit antenna is calculated to have negative power, this transmit antenna is switched off because it is not possible to generate a signal with negative power. So, the power for this antenna is set to zero. Doing this effectively increases the total transmit power. In order to make a fair comparison, the power constraint has to be maintained. Therefore, the power for other transmit antennas has to be recalculated to meet the total power constraint. In the simulations conducted here, the extra power added resulting from switching off some transmit antennas is taken evenly from the remaining transmit antennas. The biggest transmit power is allocated to the transmit antenna which has the best total channel gain. This gain is calculated as $\sum_{j=1}^{n_r} |h_{j,i}|^2$ for the i^{th} transmit antenna. For example, in (4.4), if $P_1 > P_2 > \cdots > P_{n_t}$, this implies $\sum_{j=1}^{n_r} |h_{j,1}|^2 > \sum_{j=1}^{n_r} |h_{j,2}|^2 > \cdots > \sum_{j=1}^{n_r} |h_{j,n_t}|^2$.

4.4 SUMMARY

In this chapter, some recently proposed HARQ schemes designed for wireless MIMO communication systems and some AMC schemes were summarized. The main contribution of this thesis, three novel HARQ schemes, was presented. The advantages and disadvantages of these HARQ schemes compared with the HARQ schemes introduced in Section 4.1 were discussed. The proposed HARQ schemes can achieve both good performance and good throughput at the cost of increased decoding complexity. This will be shown by the simulation result presented in Chapter 5. We also described how

to apply the water filling principle to the proposed HARQ schemes.

Chapter 5

SIMULATION RESULTS

In this chapter, the simulation results for all hybrid ARQ (HARQ) schemes proposed in Chapter 4 are presented. The simulation environment considered for these HARQ schemes is introduced first. Then the simulation results obtained are presented and explained. The truncated selective-repeat ARQ (TS-ARQ) scheme with an orthogonal space-time block code (OSTBC) and a non-orthogonal space-time block code (NOSTBC) is used for comparison purposes for the reasons stated in Section 1.3.

5.1 SIMULATION ENVIRONMENT

In this section, we introduce the simulation environment including the channel models and the packet/frame structure used in the simulations presented.

5.1.1 Channel Model

The wireless communication channel is modeled as a Rayleigh fading channel. The channel coefficients, $\mathbf{H}$, are assumed to be iid Gaussian random variables with zero mean and variance of $1/2$ per dimension. The fading is considered to be flat because the transmitted signals are assumed to have a narrow bandwidth.

Two channel models are considered here. The first channel model is the quasi-static flat fading channel model, which assumes that the channel matrix $\mathbf{H}$ is constant over the entire packet and varies independently between adjacent packets. In the second channel model, the matrix $\mathbf{H}$ is no longer constant over the entire packet, but varies with time. In this environment, the receiver observes a bandwidth expansion due to the

Doppler effect as introduced in Section 1.2.2. This Doppler effect can be represented by normalizing the frequency of the Doppler spread with regard to the transmission rate, which is denoted as f_dT in this thesis, where T is the symbol duration and f_d is the Doppler spread. In this thesis, three different fade rates are considered, namely $f_dT = 0.0001, 0.001$ and 0.01. They represent a channel with slow to fast fading.

The time varying channel can be simulated using the following methods [11] [67] [68]

1. Filtered Gaussian Noise Fading Models: In this method, a filter having desired frequency response is used to filter a white Gaussian noise signal with zero mean to obtain the channel coefficients with the desired Doppler spectrum.

2. The Jakes Model: Another way of simulating a Rayleigh fading channel is to sum a sufficiently large number of sinusoids. These sinusoids are weighted such that the result can accurately approximate the desired channel Doppler spectrum.

3. Flat-Spectrum Fading Model: This model assumes that the Doppler spectrum is flat over a range of Doppler shifts symmetric about the carrier frequency.

In this thesis, the filtered Gaussian noise fading model is used because the Jakes model is computationally cumbersome (due to the large number of function calls during the simulation) and the flat-spectrum fading model can not approximate the required Doppler spectrum accurately. A filter with a cutoff frequency of f_dT and frequency response, $H(f)$, equal to the square root of the power spectral density (PSD) of the desired channel is used to filter an independent Gaussian random variable sequence. The resulting random sequence will have a PSD of $\mathcal{S}_{out}(f) = \mathcal{S}_{in}|H(f)|^2$, where $\mathcal{S}_{in}$ is the PSD of the input sequence. Because $\mathcal{S}_{in}$ is flat, the output sequence will have the desired PSD. This output sequence is band limited to f_dT. For some situations, this results in the required filter bandwidth being too narrow for implementation using a FIR filter because it requires too many taps. Therefore, an IIR filter is used in simulations[6].

[6]The Matlab code used to generate the channel coefficients for various fade rates is a modified version of the code written by Dr. Charalampos C. Tsimenidis from the University of Newcastle Upon Tyne,UK.

The envelope of the filtered Gaussian random variable sequences for fade rates of $f_dT = 0.0001$, 0.001 and 0.01 are shown in Fig. 5.1, 5.2, 5.3, respectively. These figures represent one SISO link for different fade rates. They clearly show that when the fade rate increases, the channel varies more rapidly. Visually, the autocorrelation between adjacent intervals is very high for the $f_dT = 0.0001$ case. This means the channel stays in a good or bad condition for a relatively long period of time. But, for $f_dT = 0.01$, the autocorrelation between adjacent time intervals is much smaller than that in the other two cases. The correlation between adjacent time intervals plays an important role in the performance of ARQ systems. This will be explained in Section 5.3.

For the quasi-static channel, it is assumed that the channel coefficients are perfectly known by the receiver. For the time varying channel cases, only the channel coefficients for the last time slot of every 4 time slots spanning a NOSTBC block are assumed perfectly known by the receiver. The coefficients for the last time slot are treated as the channel coefficients for all 4 time slots spanning one NOSTBC block. As a result, other time slots will suffer from a channel estimation error.

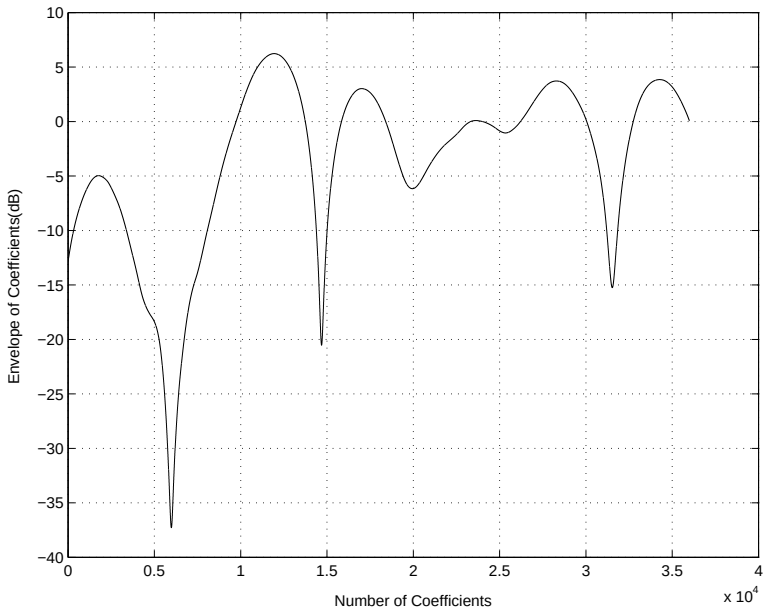

Figure 5.1 The envelope of the coefficients for a Rayleigh fading channel with fade rate $f_dT = 0.0001$.

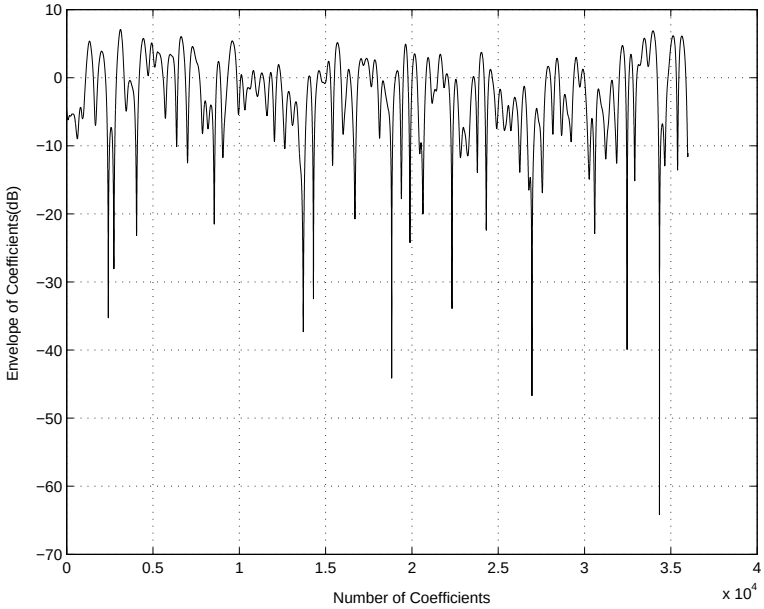

Figure 5.2 The envelope of the coefficients for a Rayleigh fading channel with fade rate $f_dT = 0.001$.

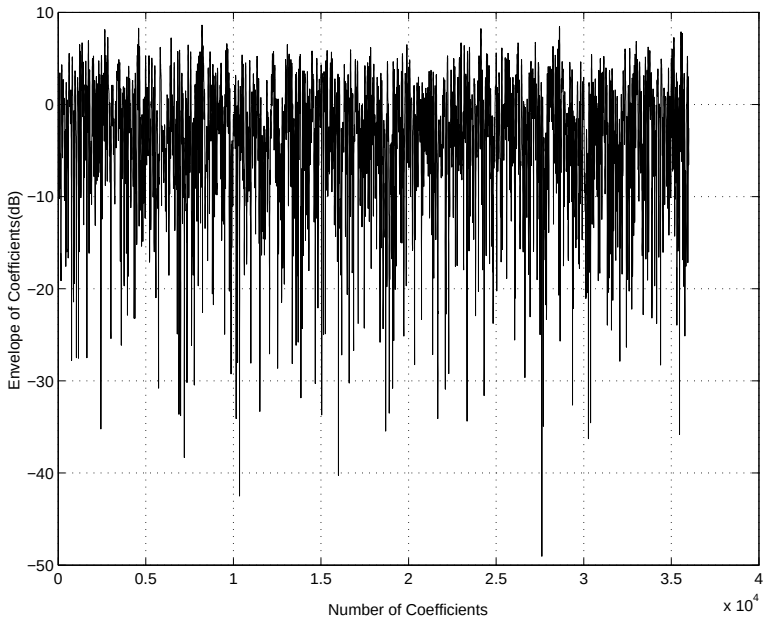

Figure 5.3 The envelope of the coefficients for a Rayleigh fading channel with fade rate $f_dT = 0.01$.

5.1.2 Simulation Parameters

The packet and frame structure of [4], [5] is used in the simulations. Each frame consists of several packets and control symbols and is formed by 1064 data bits. An overall cyclic redundancy check (CRC) code is used for ARQ error detection purposes. Here, the CRC-16 (USA) code [2] is used for error detection instead of assuming ideal error detection. Simulation results show that it gives performance close to ideal error detection and theoretical analysis shows it can provide 99.9985% error detection coverage [7]. Then the resulting 1080 CRC encoded bits are encoded using an OSTBC or NOSTBC. The resulting packet spans $n_t = 4$ transmit antennas and 360 time slots when the NOSTBC defined by the eRS (16,3) code is used (90 NOSTBC codewords).

The simulations presented use only one packet per frame for ease of exposition, but the proposed schemes can easily be used with multi-packet frames. The control symbols in the packet are ignored as we focus on the performance of the HARQ schemes. This packet structure is shown in Fig 5.4. As introduced in Section 3.1, a standard MIMO system with $n_t = 4$ transmit antennas and $n_r = 4$ receive antennas is considered here as shown in Fig. 3.1 and the baseband signals of this model are given by (3.2).

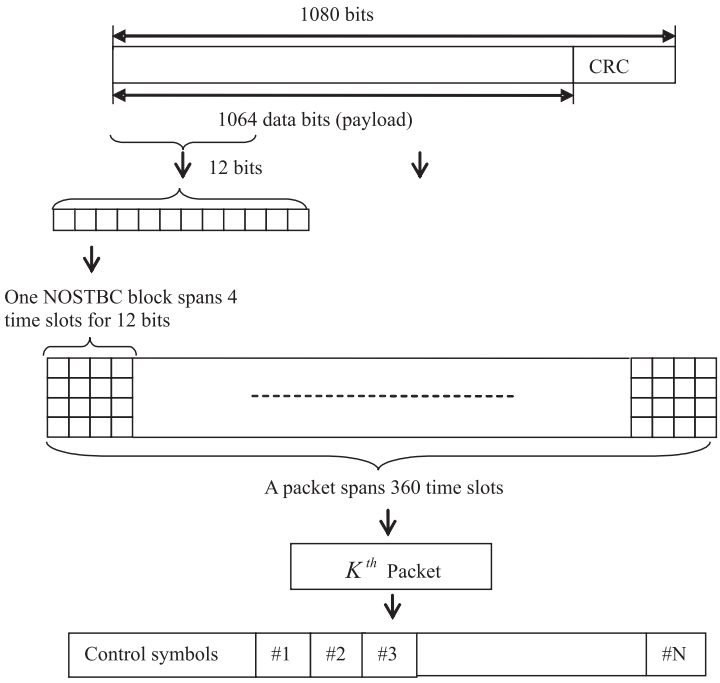

Figure 5.4 The packet and frame structure [4] [5].

Because the design of the HARQ scheme rather than the STC is the focus of this research, the NOSTBC defined on the eRS (16,3) code is used in simulations. The use of the eRS (16,3) NOSTBC makes the use of a brute force decoder, which is an optimal joint maximum likelihood decoder, feasible. This decoder searches all possible codewords to find the most likely transmitted codeword. The code book has 4096 codewords. This NOSTBC has a rate of $R_{stbc} = n_t \times R_{ecc} = 4 \times (3/16) = 3/4$, which is the same rate offered by the OSTBC introduced in Section 3.4 for 4 transmit antennas. Therefore, the proposed ARQ schemes, HARQ-1 and HARQ-2, are simulated with the OSTBC defined by (3.40) for comparison purposes. Recall that the HARQ-3 scheme is specially designed for the NOSTBCs and so cannot be simulated with the OSTBCs. Note that the proposed ARQ schemes can readily be used with the higher rate NOSTBCs defined in [12].

As in [13] [57] results are plotted against the data bit energy, E_b, divided by the noise variance for each receive antenna

$$N_0 = \frac{n_t E_s}{\log_2(M) R_{ecc} 10^{0.1SNR} R_{CRC}} \tag{5.1}$$

where E_s is the average energy of a M-QAM constellation point which is normalized to 1, $R_{CRC} = (1080 - 16)/1080 = 0.985$ is the rate of the CRC code and $R_{ecc} = 3/16$ is the rate of the eRS code used by the NOSTBC.

The performance of each ARQ scheme is measured using the dropped packet rate, spectral efficiency and throughput spectral efficiency. A dropped packet is a packet that still contains errors after the maximum allowed number of retransmissions. The dropped packet rate is computed as:

$$\mathcal{R}_{Drop} = \frac{P_{drop}}{P_{total}} \tag{5.2}$$

where P_{drop} is the number of dropped packets and P_{total} is the total number of transmitted information packets (including all dropped packets and successfully delivered

packets). The spectral efficiency is computed as

$$\text{Spectral} \quad \text{Efficiency} = \frac{\log_2(M) \times R_{stbc} \times R_{CRC}}{m_{ave}} \tag{5.3}$$

where m_{ave} is the average number of transmissions for each packet and M is the number of points in the constellation. The dropped packet rate indicates the quality of the wireless system and the spectral efficiency (bits/s/Hz) indicates how fast the information can be delivered to the user in a given bandwidth. The dropped packet rate and spectral efficiency can be combined to define the *throughput spectral efficiency*. It is defined as

$$\text{Throughput Spectral Efficiency} = \frac{\log_2(M) \times R_{stbc} \times R_{CRC} \times (1 - \mathcal{R}_{Drop})}{m_{ave}} \tag{5.4}$$

Because only finite delays are allowed for most practical system, a maximum of 4 transmissions are allowed for each packet (meaning 3 retransmissions) as used in [13]. The decoding candidate list used for the HARQ-3 simulations has 2000 or 250 codewords for the quasi-static channel case or 500 codewords for the time varying channel case, which are approximately a half, 1/16 and 1/8 of the total number of codewords in the (16,3) eRS code, respectively. Larger savings are expected for higher rate codes.

5.2 QUASI-STATIC CHANNEL PERFORMANCE

Fig. 5.5 shows that the (16,3) eRS NOSTBC offers significant coding gains compared to the rate 3/4 OSTBC [43] for the HARQ-1 and HARQ-2 schemes. All the performance curves representing HARQ-3 in this section are obtained using a decoding candidate list of 2000 codewords unless otherwise stated. Fig. 5.5, Fig. 5.6 and Fig. 5.7 clearly show that all the proposed HARQ schemes outperform the TS-ARQ scheme in terms of dropped packet rate, spectral efficiency and throughput spectral efficiency. They also show that the HARQ-3 scheme has the best performance among all the proposed ARQ schemes. For the OSTBC, HARQ-1 and HARQ-2 gain approximately 1 dB and 3.9 dB at a dropped packet rate of approximately 10^{-3} compared with TS-ARQ, respectively.

For the NOSTBC defined by the eRS(16,3), HARQ-1, HARQ-2 and HARQ-3 gain approximately 0.9 dB, 4.9 dB and 6 dB at a dropped packet rate of approximately 10^{-3} compared with TS-ARQ, respectively. The performance curve for HARQ-3 has a steeper slope than those for other HARQs. Fig. 5.6 also indicates that the average numbers of transmissions per packet for the HARQ schemes are much smaller than that for the TS-ARQ scheme for a wide range of SNRs, over which the spectral efficiency curve of TS-ARQ is not saturated.

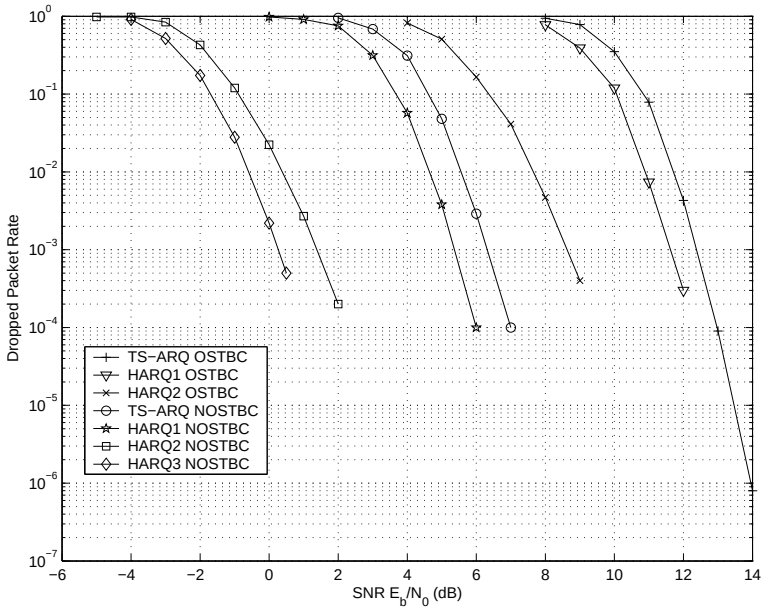

Figure 5.5 The dropped packet rate for different ARQ schemes and different STBCs.

All curves in Fig. 5.6 and Fig. 5.7 eventually approach

$$\log_2(M) \times R_{stbc} \times R_{CRC} = 4 \times (3/4) \times 0.985 = 2.95 \quad \text{bits/s/Hz}. \tag{5.5}$$

Note that if the $R_{stbc} = 9/4$ eRS(16,9) NOSTBC in [12] is used, the maximum achievable spectral efficiency is 8.85 bits/s/Hz. As can be seen all HARQ schemes proposed here, especially HARQ-3, combined with the NOSTBC of [12] can be used to provide reliable communications at a good throughput. But this improvement is achieved at the cost of requiring more memory and increasing the computational complexity.

Fig. 5.8, Fig. 5.9 and Fig. 5.10 compare the performance of HARQ-2 and HARQ-3 for different maximum allowed numbers of transmissions. Fig. 5.8 clearly shows

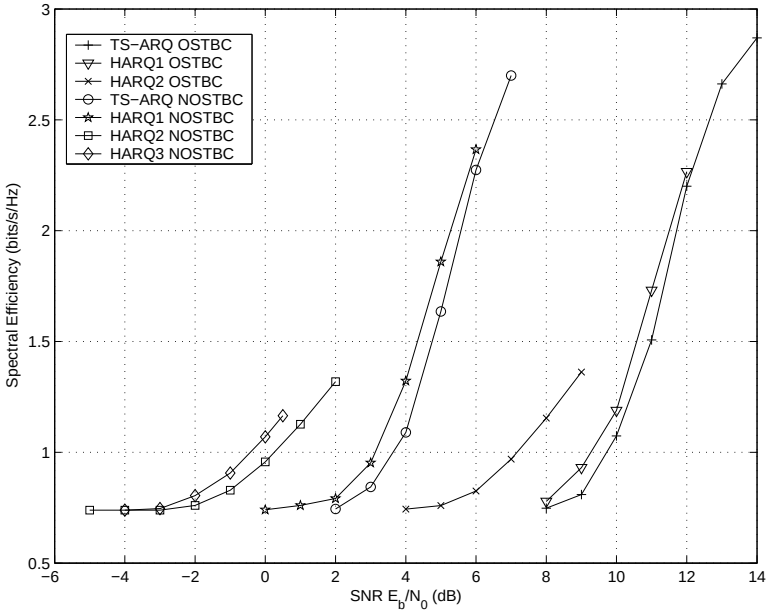

Figure 5.6 The spectral efficiency for different ARQ schemes and different STBCs.

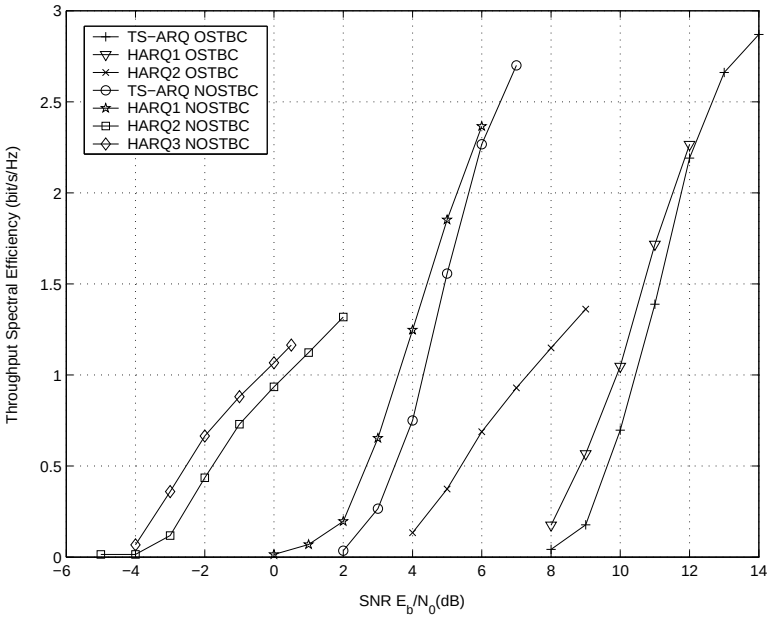

Figure 5.7 The throughput spectral efficiency for different ARQ schemes and different STBCs.

that, by using more retransmissions, the information can be successfully sent through a noisy channel at lower SNR compared to using fewer retransmissions. The curves in these figures are closely related to the maximum allowed number of transmissions, m_{max}. When the SNR is so small that the scheme fails to deliver information without errors, m_{ave} is equal to the maximum allowed number of transmissions. Therefore, in Fig. 5.9, the spectral efficiency curves for both HARQ-2 and -3 start from 0.7388, 0.985 and 1.4777 for systems with $m_{max} = 4, 3$ and 2, respectively. Fig. 5.9 also shows the average number of transmissions per packet, m_{ave}, because it is the only variable considered in (5.3). Fig. 5.10 provides an overview for the ARQ system under investigation by combining the information from both Fig. 5.8 and Fig. 5.9. In this figure, the throughput spectral efficiency is almost 0 when the dropped packet rate is close to 100%. As the SNR increases, the average number of transmissions per packet decreases and eventually becomes almost 1.

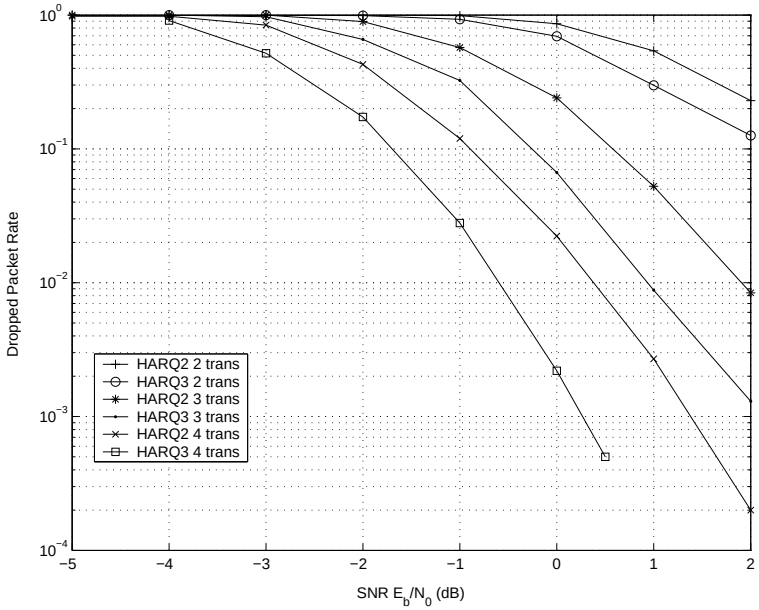

Figure 5.8 The dropped packet rate of HARQ-2 and HARQ-3 for a maximum of 2, 3 and 4 transmissions.

Fig. 5.11, Fig. 5.12 and Fig. 5.13 show that adaptively allocating power across transmit antennas (water filling) can improve the performance significantly at low SNRs, but it approaches the performance without water-filling as SNR increases.

Fig. 5.14 shows the performance of HARQ-3 when using decoding lists of vari-

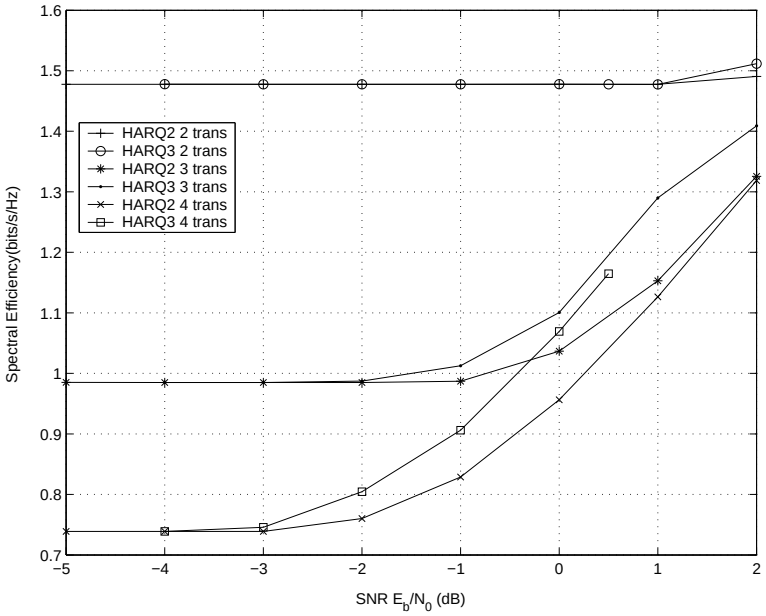

Figure 5.9 The spectral efficiency of HARQ-2 and HARQ-3 for a maximum of 2, 3 and 4 transmissions.

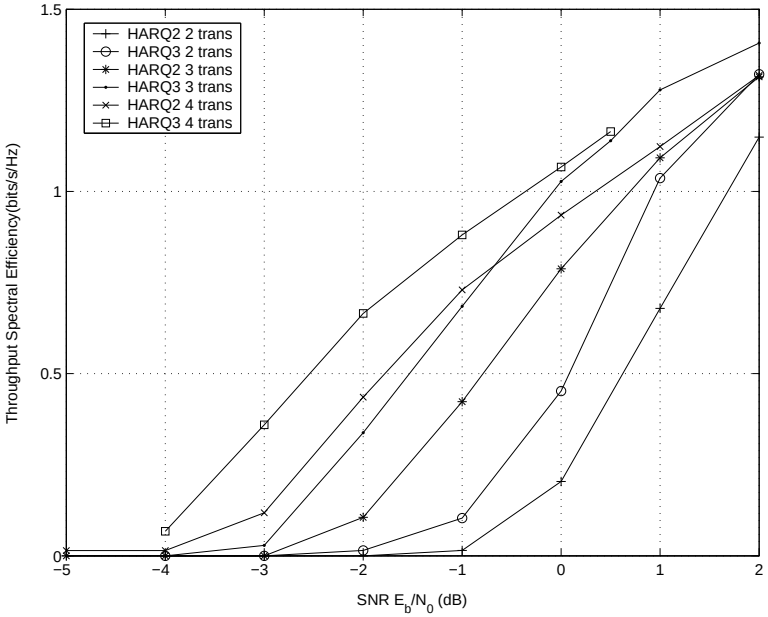

Figure 5.10 The throughput spectral efficiency of HARQ-2 and HARQ-3 for a maximum of 2, 3 and 4 transmissions.

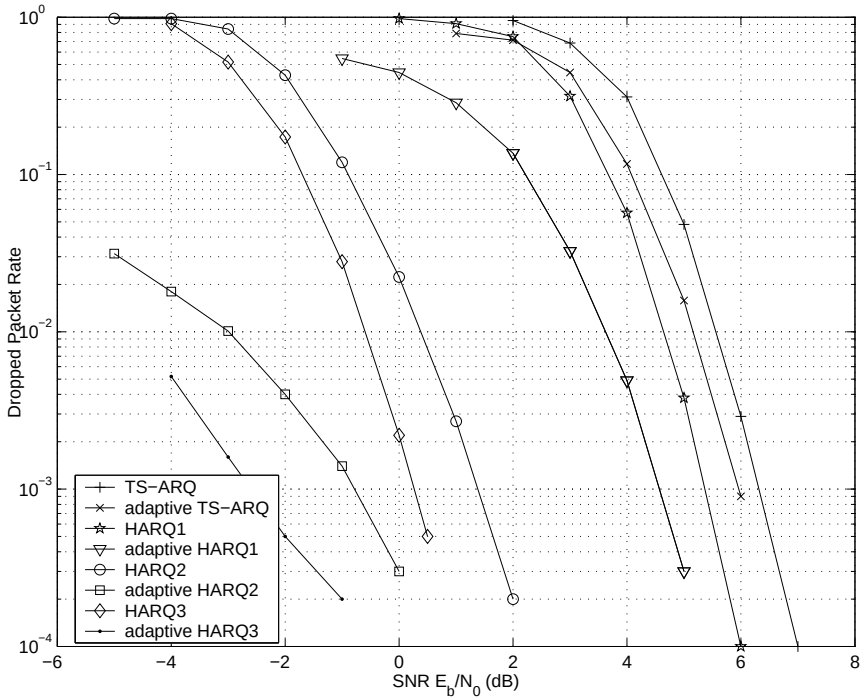

Figure 5.11 The dropped packet rate for ARQ schemes with and without water filling (adaptive) with a NOSTBC.

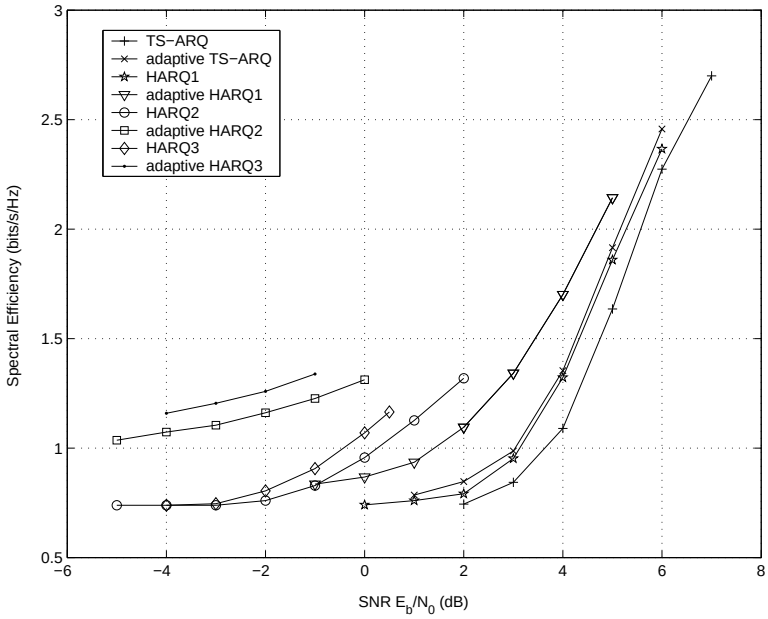

Figure 5.12 The spectral efficiency for ARQ schemes with and without water filling (adaptive) with a NOSTBC.

ous sizes. This scheme suffers a small performance loss when the list size decreases from 2000 to 500. The performance of HARQ-3 with a list of 500 codewords is almost identical to one with 250 codewords. In general, using a decoding candidate list can significantly reduce the computational complexity without suffering a significant performance loss compared with an optimal joint maximum likelihood decoder. Also, because the decoding candidate list is used to decode the information received during the retransmissions in the HARQ-3 scheme, the more retransmissions required for a packet, the more computational complexity is reduced compared to the case of not using a decoding candidate list. Assuming the decoding candidate list is used h times for a packet and the decoding list considers only $1/s$ of the total $\mathcal{T}$ codewords contained in the code book, then the number of tests reduced/saved by using this list is $h \cdot \mathcal{T} \cdot (1 - 1/s)$ for this packet. For example, we consider a decoding candidate list of 250 codewords, which is used twice to decode a NOSTBC block. This list contains slightly less than 1/16 of the total number of codewords contained in the eRS (16,3) code. For this block, by using this list, the number of tests is reduced by approximately $2 \times 4096 \times (1 - 1/16) = 7680$. For a whole packet, assuming the same packet structure shown in Fig. 5.4 is considered, the number of tests is reduced by approximately

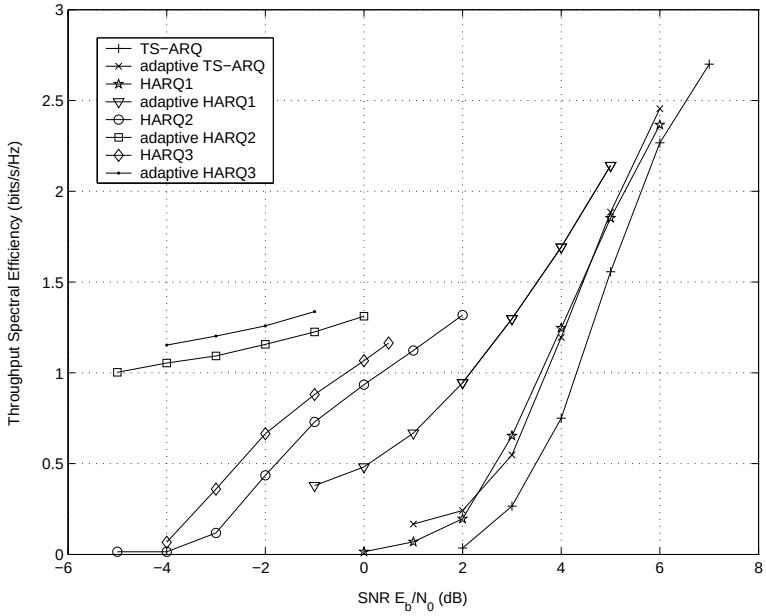

Figure 5.13 The throughput spectral efficiency for ARQ schemes with and without water filling (adaptive) with a NOSTBC.

$90 \times 7680 = 691200$.

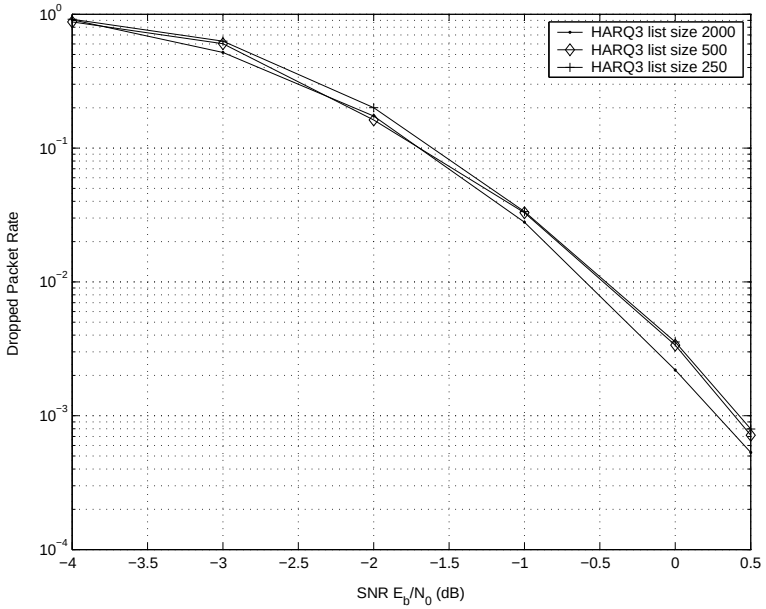

Figure 5.14 Performance of the HARQ-3 scheme for decoding lists of various sizes.

We now discuss why good performance is achieved by the HARQ-3 scheme. From all presented figures, it can be clearly observed that the HARQ-3 scheme has the best performance. This is because, by combining the soft information provided by the systematic and non-systematic eRS (16,3) codewords, the minimum squared Euclidean distance between two information vectors is increased. Each retransmission has the effect of increasing the number of virtual receive antennas [66]. Therefore, if a packet is transmitted $0 < m \leq m_{max}$ times, this MIMO system will have a diversity gain $rn_r \geq mn_r$ assuming CSI is independent for each transmission. If $mn_r \geq 4$, the design criteria for this type of MIMO system is to maximize the minimum Euclidean distance between all codewords as introduced in Chapter 3. By conducting an exhaustive search, it can be found that the minimum squared Euclidean distance for both systematic and non-systematic eRS (16,3) codewords is[7] 6.4. Therefore, for a non-ARQ system, the NOSTBCs defined by either systematic or nonsystematic eRS (16,3) codewords have the same performance, as shown in Fig. 5.15. It is clear that the combined minimum squared Euclidean distance is $2 \times 6.4 = 12.8$ for HARQ-2 with two transmissions. But,

[7]This assumes the 16-QAM constellation (shown in Fig. 1.4) is used and that all points are normalized to have unit average energy (by multiplying each symbol by $1/\sqrt{10}$).

for the HARQ-3 scheme, the combined minimum squared Euclidean distance, which is the total Euclidean distance between the systematic codewords and non-systematic codewords for a pair of information vectors is 20.6 rather than the 12.8 offered by the HARQ-2 system. This means that the information vectors, which are mapped close together by using a systematic encoding, are mapped further apart when using a non-systematic encoding. If a packet is transmitted 4 times, the upper bounds for the minimum distance is 41.2 and 25.6 for the HARQ-3 and HARQ-2 schemes, respectively. It is clear that the minimum distance is significantly increased by transmitting both systematic and non-systematic codewords for a given packet. So, the HARQ-3 scheme has the best performance among all ARQ schemes considered in this chapter.

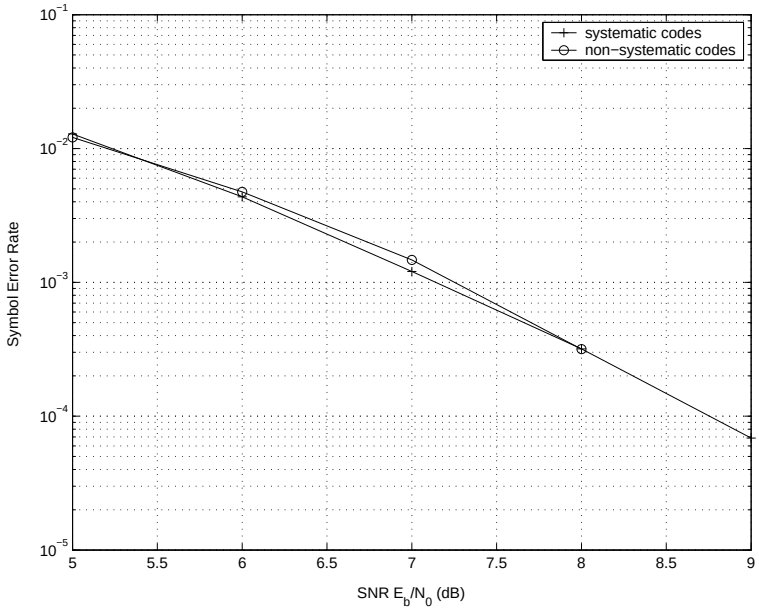

Figure 5.15 Performance of eRS(16,3) NOSTBC using systematic codes and non-systematic codes.

5.3 TIME VARYING CHANNEL PERFORMANCE

In this section, we consider the performance of the HARQ schemes in a time varying channel with fade rates of 0.0001, 0.001 and 0.01. Fig. 5.16, 5.17, 5.18 and 5.19 show the performance for TS-ARQ, HARQ-1, HARQ-2 and HARQ-3, respectively. All performance curves representing HARQ-3 in this section are obtained using a decoding candidate list of 500 codewords. Each retransmission is separated by 5 packet lengths

for the time varying channel. The quasi-static channel assumes each transmission is independently faded.

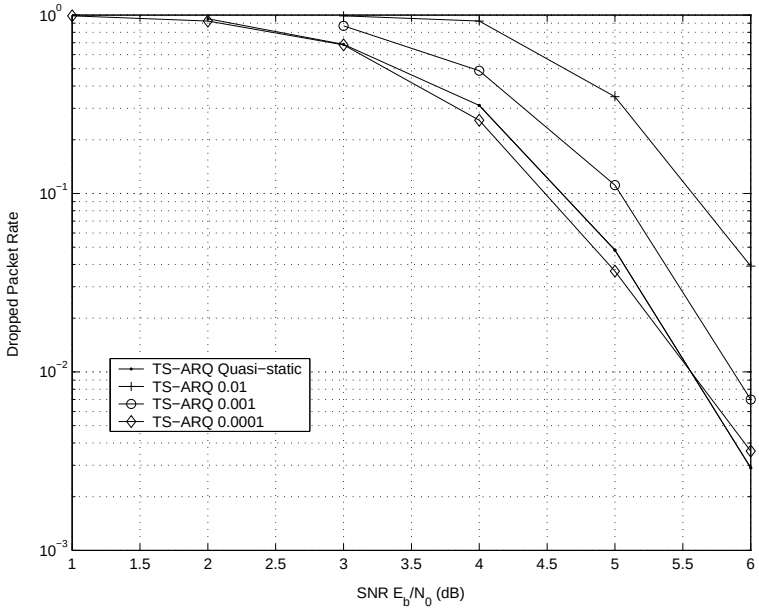

Figure 5.16 Performance of the TS-ARQ scheme for a quasi-static channel and a time varying channel with $f_dT = 0.01, 0.001$ and 0.0001 with a NOSTBC.

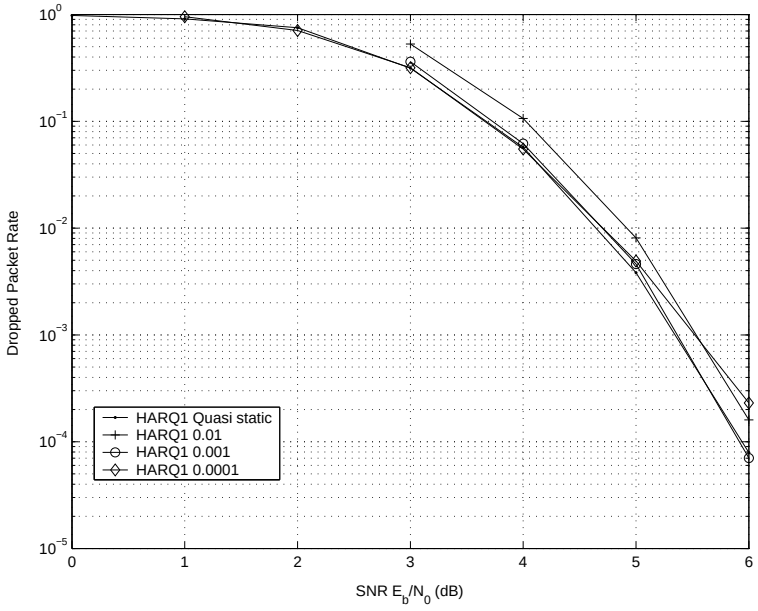

Figure 5.17 Performance of the HARQ-1 scheme for a quasi-static channel and a time varying channel with $f_dT = 0.01, 0.001$ and 0.0001 with a NOSTBC.

Fig. 5.16 shows that the TS-ARQ scheme cannot cope very well with the rapidly time varying channel. It performs the best when the channel has the slowest fade rate

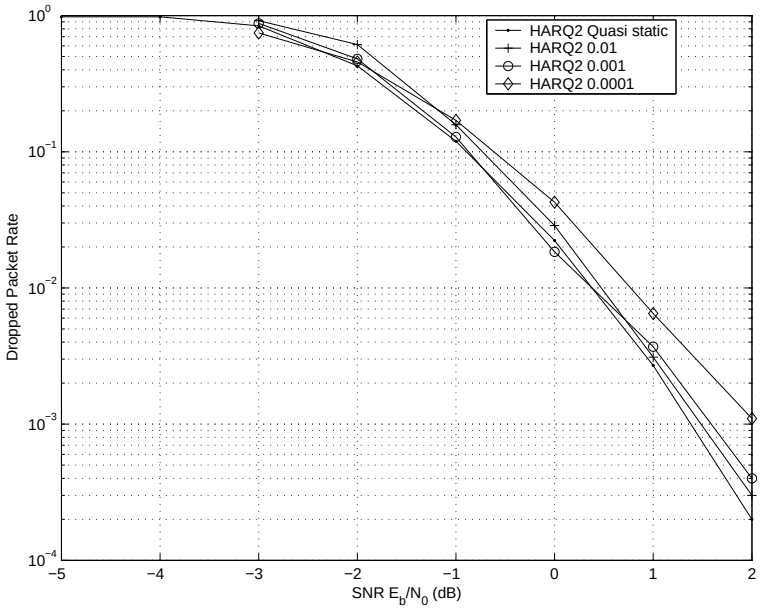

Figure 5.18 Performance of the HARQ-2 scheme for a quasi-static channel and a time varying channel with $f_dT = 0.01, 0.001$ and 0.0001 with a NOSTBC.

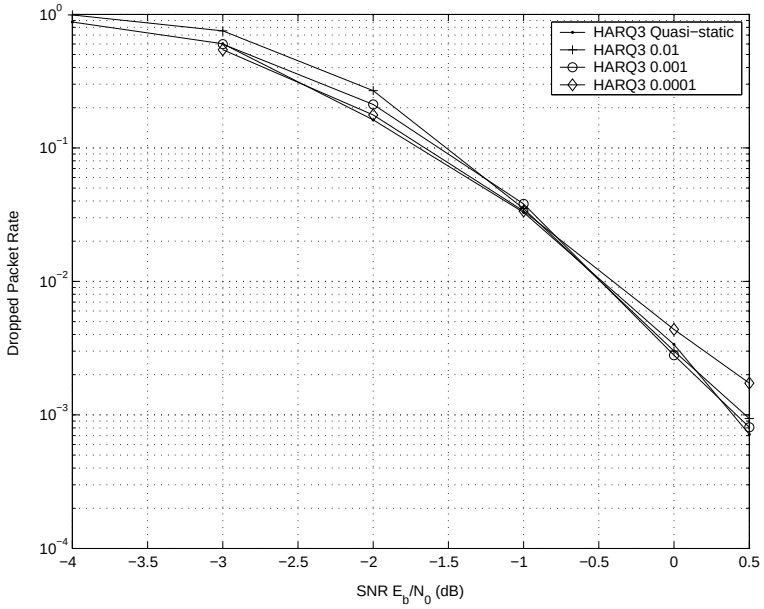

Figure 5.19 Performance of the HARQ-3 scheme for a quasi-static channel and a time varying channel with $f_dT = 0.01, 0.001$ and 0.0001 with a NOSTBC.

and the performance gets worse when the fade rate increases. The HARQ schemes have much more robust performance, especially HARQ-3, as shown in Fig. 5.17, 5.18 and 5.19. Again, these figures show that HARQ-3 has the best performance among all ARQ schemes. For all three HARQ schemes, the best performance can be obtained when the channel is quasi-static. The performance is similar to the quasi-static case when $f_dT = 0.001$. The performance becomes worse when $f_dT = 0.01$ and $f_dT = 0.0001$.

When the channel has $f_dT = 0.0001$, the performance curves of all HARQ schemes have a less steep slope than for faster fade rates or a quasi-static channel. HARQ-2 is modified in order to improve the performance for this case. In this modified HARQ-2 scheme, a systematic encoded NOSTBC is generated by cyclic shifting the corresponding NOSTBC block used in the previous transmission for each retransmission. This systematic encoded NOSTBC, which is cyclic shifted by m positions, is transmitted for the m^{th} retransmission instead of transmitting the same systematic encoded NOSTBC for all transmissions. Fig. 5.20 shows the performance of the modified HARQ-2 scheme. It is better than the original HARQ-2 scheme when the channel has $f_dT = 0.0001$. This modification can readily be applied to the other HARQ schemes to improve the performance when the channel is highly correlated. The performance of all HARQ schemes under time varying channels is shown in Fig. 5.21.

Now, we discuss the performance of HARQ schemes under time varying channels. Fig. 5.17, 5.18 and 5.19 clearly show that the best performance is obtained when the channel is assumed to be quasi-static. This is because the quasi-static assumption offers uncorrelated channel coefficients for all transmissions. When the channel fade rate is 0.001, the performance is very close to the quasi-static case. When the fade rate is equal to 0.01, the dropped packet rate-SNR curve has the same slope as the quasi-static case, but it is shifted to the right. When the fade rate equals 0.0001, all schemes suffer a diversity loss as seen by the change in slope. This means that the performance of $f_dT = 0.0001$ will be worse than that of other fade rates at high SNRs, although the channel is more accurately estimated due to the slow variation of the channel.

The slope degradation for $f_dT = 0.0001$ is due to the channel coefficients for a given

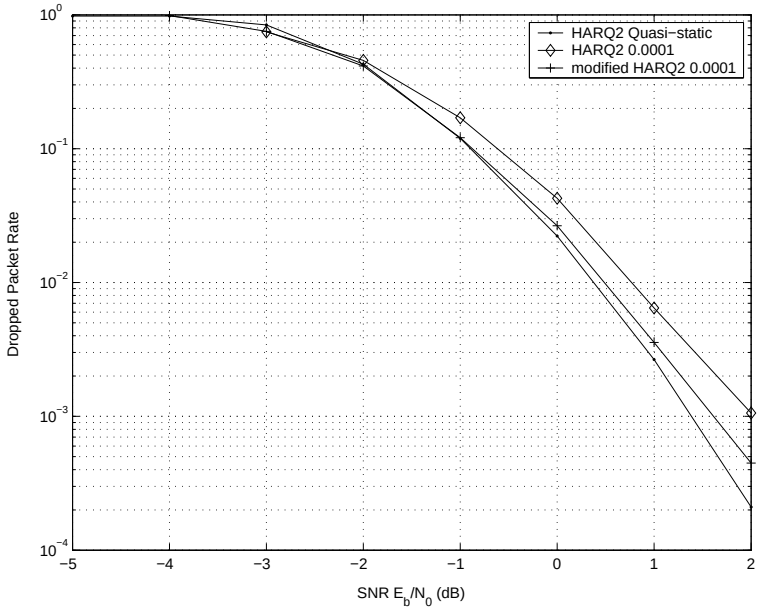

Figure 5.20 The performance of the HARQ-2 and modified HARQ-2 schemes for a quasi-static channel and a time varying channel with $f_dT = 0.0001$ with a NOSTBC.

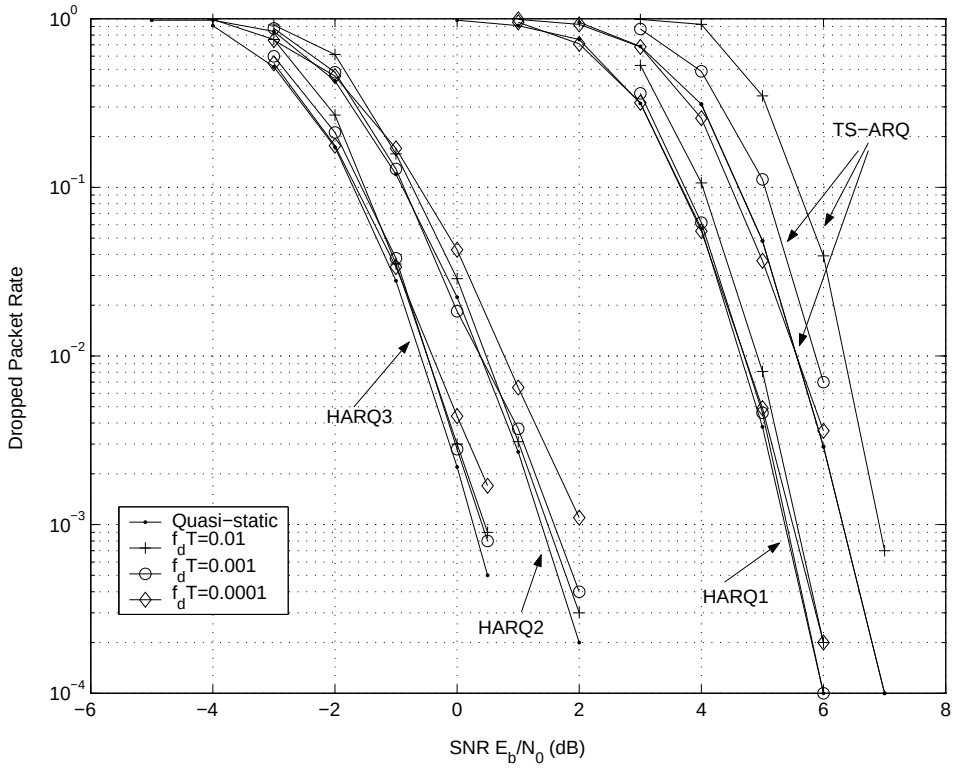

Figure 5.21 The dropped packet rate for all ARQ schemes for a quasi-static channel and time varying channel with $f_dT = 0.01, 0.001$ and 0.0001 with a NOSTBC.

path connecting a transmit and a receive antenna being highly correlated in time. Under the current assumptions and the packet structure introduced in Section 5.1, the number of time slots between adjacent transmissions is $5 \times 360 = 1800$ which is less than half the number of time slots contained in one division in Fig 5.1, which is $0.5 \times 10^4 = 5000$. Fig 5.1 shows that the channel coefficients are still highly correlated within these time slots. This means, if the first transmission is not decoded correctly due to a deep fade, the second transmission is very likely to also be unsuccessful. So, although the allowed maximum number of transmissions is 4, the erroneous packet, in fact, has fewer transmissions which can experience independent channel conditions compared to the quasi-static channel. In contrast, although the quasi-static channel is fixed for each packet, it varies independently for each transmission.

Fig. 5.2 shows that when $f_dT = 0.001$, each transmission can have almost independent channel coefficients for each transmission and the change of these coefficients within one NOSTBC block is not severe when the fade rate is 0.001. Hence, the channel coefficients for the last time slot can approximate the entire block quite well. This leads to the performance for $f_dT = 0.001$ being very close to that of the quasi-static case. For $f_dT = 0.01$, the channel coefficients change rapidly. This causes some loss in coding gain. Therefore, it can be observed that from all three figures, the performance curves for $f_dT = 0.01$ have the same slope as the curves for the quasi-static case and are shifted to the right of the curves for the quasi-static case.

For the TS-ARQ scheme, since the previous information is not stored and used for all f_dT cases, the performance comparison is not very sensitive to this time diversity loss. Therefore, the channel with $f_dT = 0.0001$ gives the best performance. The performance loss is increased as f_dT increases. The time diversity loss can be solved by extending the time delay between retransmissions, but doing this effectively increases the round trip delay and leads to having fewer allowed retransmissions, if the allowed total delay is fixed.

Another way to compensate for this time diversity loss to some extent is to transmit different symbols from each transmit antenna for each retransmission. This is obtained by cyclic shifting the previously transmitted eRS(16,3) codeword to form a NOSTBC

as in the modified HARQ-2 scheme. Fig. 5.20 clearly shows that the modified HARQ-2 offers better performance. By cyclic shifting symbols, the symbols experiencing bad channels during previous transmissions may be sent through a good channel for the current transmission because the fading for each path is still independent. Transmitting different symbols from each transmit antenna for adjacent transmissions is also part of the reason for getting better performance using HARQ-3 under time varying channels. The performance of HARQ-3 can be further enhanced by applying the same technique used in the modified HARQ-2 scheme to both systematic and non-systematic codewords.

5.4 SUMMARY

In this chapter, simulation results for the proposed HARQ schemes were presented. The simulation environment considered was introduced. Simulation results show that using the NOSTBCs proposed in [12] can offer a significant coding gain compared with OSTBCs. The simulation results also show the HARQ-3 scheme offers the best performance among all HARQ schemes for all channels considered in this thesis. If the water filling principle is used with these HARQ schemes, then performance can be further improved. The proposed HARQ schemes suffer a performance loss when the channel is correlated in time. In order to enhance the performance for this channel condition, a modified version of HARQ-2 was proposed. The simulation results show it has better performance than the original HARQ-2 scheme when the channel is correlated. This modification can easily be used with the other HARQ schemes.

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

Automatic Repeat-reQuest (ARQ) schemes can help provide a reliable communication link for many types of communication systems, especially for wireless communication systems. Using hybrid ARQ (HARQ) schemes can greatly improve the performance compared to traditional ARQ schemes, because they reduce the possibility of further retransmission if errors occur. They provide good performance when used with codes that provide good coding gains.

There are many different error correction codes which could be used including convolutional codes and trellis coded modulation (TCM). In order to achieve good performance, a powerful packet combining scheme should be designed for the particular codes chosen as part of the ARQ scheme. This can greatly improve the system performance.

Error detection codes are also a necessary component of an ARQ scheme. They are used to decide whether further retransmissions are required. CRC codes are normally used for error detection purposes because they can detect a wide range of errors by adding small amounts of redundancy.

In order to combat channel fading, space-time codes (STCs) for MIMO systems have attracted lots of research attention. In this thesis, we focus on space-time block codes (STBCs), which can be classified into three subfamilies, namely orthogonal space-time block codes (OSTBCs), quasi-orthogonal space-time block codes (QOSTBCs) and non-orthogonal space-time block codes (NOSTBCs). OSTBCs and QOSTBCs have

simple decoding algorithms, but can not provide coding gain. Therefore, they are normally concatenated with other codes such as trellis codes and Turbo codes to achieve good performance. This concatenation limits the achievable throughput. On the other hand, some NOSTBCs can provide good rate and good coding gains at the cost of increased decoding complexity. The simulation results in [12] show that the NOSTBCs proposed in [12] are very powerful STBCs. They can offer high rate and good coding gain. Hence, a good HARQ scheme can be designed for them to achieve reliable communication and high rate simultaneously.

In this thesis, we propose three HARQ schemes, namely HARQ-1, HARQ-2 and HARQ-3, for the NOSTBCs in [12]. In the HARQ-1 scheme, only part of the previously received information is kept and combined with the current received information. In the HARQ-2 scheme, information for all received versions are combined together. In HARQ-3, different codewords are used for retransmissions. Simulation results presented in Chapter 5 show that all three HARQ schemes outperform the TS-ARQ scheme. Among these schemes, the HARQ-3 scheme has the best performance for all the channel models considered. This is because the HARQ-3 scheme provides the largest minimum Euclidean distance between different codewords out of all the three HARQ schemes. This satisfies the trace criterion, the design criterion for STC systems with diversity gain of 4 or more. Some of the proposed HARQ schemes (HARQ-1 and HARQ-2) can also be applied to other STCs.

Adaptive modulation and coding is another powerful technique which can be used to achieve high channel capacity or high transmission rate. The water filling principle is used in this thesis in conjunction with the proposed HARQ schemes. It is used to adaptively allocate the transmit power among transmit antennas. Simulation results show that using this principle with the proposed HARQ schemes can improve performance, especially for low SNRs. These simulations are conducted assuming an error free, no-delay feedback channel is available. We also assume that the transmitter can precisely generate the power calculated using water filling principle.

6.2 FUTURE WORK

Possible future work is briefly described as follows. There are two questions associated with this research, which may be answered analytically.

1. The minimum list size required without significantly degrading performance may be analyzed by calculating the ratio of the union bound of the block error rate contributed by only nearest codewords in the list to the union bound obtained using maximum likelihood decoding. It could be expected that, at the SNRs of interest, the ratio is close to 1.

2. The performance bounds could be analyzed. The key of this analysis is to work out the block error rates[8] for each transmission, which, again, may resort in the calculation of the union bound. Assuming the block error rates for each transmission are available and denoted by $\mathcal{P}(b^n)$, where n is the number of transmissions, the drop packet rate after the fourth transmission may be represented as

$$\mathcal{P}_{Drop} = (1-(1-\mathcal{P}(b^1))^\eta)(1-(1-\mathcal{P}(b^2))^\eta)(1-(1-\mathcal{P}(b^3))^\eta)(1-(1-\mathcal{P}(b^4))^\eta),$$

where η is the number of NOSTBCs a packet contains.

Another possible future project is to add antenna selection into the proposed HARQ scheme. The transmitter chooses the best four transmit antennas for transmission. This requires the transmitter to have more antennas than required by the NOSTBCs in [12].

A rate adaptive modulation and ARQ scheme could be developed for these NOSTBCs. It would be easy to have a set of NOSTBCs defined by eRS(16,k) codes with a range of different rates by simply changing the number of information symbols rather than changing the constellation size, which is what the adaptive modulation scheme proposed in [4] does. This rate adaptive scheme could be designed as follows:

1. Choose the rates required by the system.

[8] Each space-time block code codeword is considered to be one block.

2. Derive the error packet rate functions based on the NOSTBCs defined by eRS (16,k) codes and the packet/frame structure adopted for this system. Then, the dropped packet rate functions can be derived from these error packet rate functions based on the allowed maximum number of transmissions used by this system.

3. Specify a desired dropped packet rate and then, by inverting the dropped packet rate functions obtained during previous steps, a set of SNRs can be obtained. This set of SNRs are the thresholds based on which the transmitter switches to different NOSTBCs upon receiving the current fed back SNR information.

The goal of this adaptive system is to achieve higher rate at the cost of having lower reliability than that of a non adaptive ARQ scheme. This loss can be tolerated by this system because the required dropped packet rate is still maintained. This adaptive design can be directly applied to the TS-ARQ scheme. It is much more difficult to design a HARQ scheme for a rate adaptive system because the transmissions for the same packet may be encoded using different codes. Therefore, received signal matrices may have different sizes. How the proposed HARQ schemes work for this system remains an open problem.

Adaptively changing the transmit power among all antennas according to the water filling principle is also studied in this thesis. Simulation results show that, by doing this, the performance of the HARQ schemes can be improved, especially when the SNR is low. But these simulation results are obtained when optimum conditions are considered. These optimum conditions include a large amount of information (bits) being fed back to the transmitter without error. In addition, the transmitter can precisely generate the power required for each antenna. These assumptions are not realistic in the real world. Therefore, an open problem is to design schemes which reduce the amount of information required to be fed back. For example, assuming there are four power levels available for the transmitter, then two bits are enough to represent the required power level for one transmit antenna. The total number of bits required to represent the power level for all four transmit antennas is eight, which is a realistic number for the error free feedback channel assumption to hold true. This

certainly will cause some performance loss. Simulations could be conducted to see how much will be lost and design some schemes to minimize this loss. Also, as mentioned in the thesis, the water filling is only optimal for Gaussian inputs, the performance for adaptive power allocation may be improved by using other schemes, for example, the mercury/water filling principle in [38].

Appendix A

GF(16) TABLES

The " $\cdot$ " operation table:

$\cdot$	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}
α	0	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1
α^2	0	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α
α^3	0	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2
α^4	0	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3
α^5	0	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4
α^6	0	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5
α^7	0	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6
α^8	0	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7
α^9	0	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8
α^{10}	0	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9
α^{11}	0	α^{11}	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}
α^{12}	0	α^{12}	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}
α^{13}	0	α^{13}	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}
α^{14}	0	α^{14}	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}

The " + " operation table:

+	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}
0	0	1	α	α^2	α^3	α^4	α^5	α^6	α^7	α^8	α^9	α^{10}	α^{11}	α^{12}	α^{13}	α^{14}
1	1	0	α^4	α^8	α^{14}	α	α^{10}	α^{13}	α^9	α^2	α^7	α^5	α^{12}	α^{11}	α^6	α^3
α	α	α^4	0	α^5	α^9	1	α^2	α^{11}	α^{14}	α^{10}	α^3	α^8	α^6	α^{13}	α^{12}	α^7
α^2	α^2	α^8	α^5	0	α^6	α^{10}	α	α^3	α^{12}	1	α^{11}	α^4	α^9	α^7	α^{14}	α^{13}
α^3	α^3	α^{14}	α^9	α^6	0	α^7	α^{11}	α^2	α^4	α^{13}	α	α^{12}	α^5	α^{10}	α^8	1
α^4	α^4	α	1	α^{10}	$\alpha 7$	0	α^8	α^{12}	α^3	α^5	α^{14}	α^2	α^{13}	α^6	α^{11}	α^9
α^5	α^5	α^{10}	α^2	α	α^{11}	α^8	0	α^9	α^{13}	α^4	α^6	1	α^3	α^{14}	α^7	α^{12}
α^6	α^6	α^{13}	α^{11}	α^3	α^2	α^{12}	α^9	0	α^{10}	α^{14}	α^5	α^7	α	α^4	1	α^8
α^7	α^7	α^9	α^{14}	α^{12}	α^4	α^3	α^{13}	α^{10}	0	α^{11}	1	α^6	α^8	α^2	α^5	α
α^8	α^8	α^2	α^{10}	1	α^{13}	α^5	α^4	α^{14}	α^{11}	0	α^{12}	α	α^7	α^9	α^3	α^6
α^9	α^9	α^7	α^3	α^{11}	α	α^{14}	α^6	α^5	1	α^{12}	0	α^{13}	α^2	α^8	α^{10}	α^4
α^{10}	α^{10}	α^5	α^8	α^4	α^{12}	α^2	1	α^7	α^6	α	α^{13}	0	α^{14}	α^3	α^9	α^{11}
α^{11}	α^{11}	α^{12}	α^6	α^9	α^5	α^{13}	α^3	α	α^8	α^7	α^2	α^{14}	0	1	α^4	α^{10}
α^{12}	α^{12}	α^{11}	α^{13}	α^7	α^{10}	α^6	α^{14}	α^4	α^2	α^9	α^8	α^3	1	0	α	α^5
α^{13}	α^{13}	α^6	α^{12}	α^{14}	α^8	α^{11}	α^7	1	α^5	α^3	α^{10}	α^9	α^4	α	0	α^2
α^{14}	α^{14}	α^3	α^7	α^{13}	1	α^9	α^{12}	α^8	α	α^6	α^4	α^{11}	α^{10}	α^5	α^2	0

REFERENCES

[1] I. Widjaja, *Communication Networks Fundamental Concepts and Key Architectures*. Mc Graw Hill, 2 ed., 2004.

[2] S. Haykin, *Communication Systems*. Wiley, 4 ed., 2001.

[3] B. Vucetic and J. Yuan, *Space-time Coding*. Wiley, 2003.

[4] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combing of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1746–1755, Sep. 2004.

[5] A. Maaref and S. Aissa, "Combined adaptive modulation and truncated ARQ for packet data transmission in MIMO systems," in *Proc. GLOBECOM*, pp. 145 – 149, 2004.

[6] A. Goldsmith, *Wireless Communiation*. Cambridge University Press, 2005.

[7] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.

[8] S. Lin, D. J. C. Jr., and M. J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Magazine*, vol. 22, pp. 5–17, Dec 1984.

[9] H. Jafarkhani, *Space-time Coding Theroy and Practice*. Cambridge University Press, 2005.

[10] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected GRAY code," *IEEE Trans. Infor. Theory*, vol. 50, pp. 3170–3182, Dec. 2004.

[11] K. Pahlavan and A. Levesque, *Wireless Information Networks*. Wiley, 1995.

[12] P. A. Martin and D. P. Taylor, "High-throughput error correcting space-time block codes," *IEEE Commun. Lett.*, vol. 8, July 2004.

[13] A. V. Nguyen and M. A. Ingram, "Hybrid ARQ protocols using space-time codes," in *Proc. VTC*, vol. 4, pp. 2364–2368, Oct. 2001.

[14] M. K. Oh, Y. H. Kwon, and D. J. Park, "Efficient hybrid ARQ with space-time coding and low-complexity decoding," in *Proc. ICASSP*, vol. 4, pp. iv–589 – iv–592, May 2004.

[15] W. T. Kim, S. J. Bae, J. G. Kim, and E. K. Joo, "Performance of STBC with turbo code in HARQ scheme for mobile communication system," in *Proc. ICT*, vol. 1, pp. 85 – 89, Feb. 2003.

[16] W. T. Kim, S. J. Bae, J. G. Kim, and E. K. Joo, "Transmission schemes of STBC with turbo code for HARQ system," in *Proc. TENCON*, vol. B, pp. 191 – 194, Nov. 2004.

[17] P. A. Martin, *Error Control Coding,* ENEL *670 Lecture Notes.* University of Canterbury, New Zealand, 2005.

[18] S. Lin and D. J. Costello, *Error Control Codeing.* Pearson prentice hall, 2 ed., 2004.

[19] P. S. Sindhu, "Retransmission error control with memory," *IEEE Trans. Commun.*, vol. com-25, pp. 473–479, May 1977.

[20] J. J. Metzner and D. Chang, "Effective selective repeat ARQ strategies for very noisy channel," in *Proc. GLOBECOM*, pp. 35.2.1–35.2.8, Dec 1983.

[21] Y. W. Wang and S. Lin, "Modified selective repeat type-II hybrid ARQ system: Performance analysis," *IEEE Trans. Commun.*, vol. com-25, pp. 473–479, May 1977.

[22] M. L. B. Riediger and P. K. M. Ho, "Application of Reed-Solomon codes with erasure decoding to type-II hybrid ARQ transmission," in *Proc. GLOBECOM*, vol. 1, pp. 55 – 59, Dec. 2003.

[23] L. J. Zhang, V. O. K. Li, and Z. G. Cao, "Throughput analysis of nonbinary type-II hybrid ARQ," in *Proc. PIMRC*, vol. 1, pp. 149 – 152, Sep. 2003.

[24] S. B. Wicker and M. J. Bartz, "Type-II hybrid-ARQ protocols using puncturing MDS codes," *IEEE Trans. Commun.*, vol. 42, pp. 1431 – 1440, Feb./Mar./April 1994.

[25] S. B. Wicker, "High reliability data transfer over the land mobile radio channel using interleaved hybrid-ARQ error control," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 48–55, Feb. 1990.

[26] S. B. Wicker and M. J. Bartz, "Reed-solomon error control coding for rayleigh fading channels with feedback," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 124 – 133, May 1992.

[27] E. J. Weldon, "An improved selection repeat ARQ strategy," *IEEE Trans. Commun.*, vol. 3, pp. 480–486, Mar. 1982.

[28] D. Chase, "Code combining-a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. com-33, pp. 385–393, May 1985.

[29] Q. Zhang and S. A. Kassam, "Hybrid ARQ with selective combining for fading channels," *IEEE J. Select Areas Commun.*, vol. 17, pp. 867–880, May 1999.

[30] J. X. Yu, Y. Li, H. Murata, and S. Yoshida, "Hybrid-ARQ scheme using different TCM for retransmission," *IEEE Trans. Commun.*, vol. 48, pp. 1609 – 1613, Oct. 2000.

[31] H. Samra, Z. Ding, and P. M. Hahn, "Symbol mapping diversity design for packet retransmissions through fading channels," in *Proc. GLOBECOM*, vol. 4, pp. 1989 – 1993, Dec. 2003.

[32] H. Samra, Z. Ding, and P. M. Hahn, "Optimal symbol mapping diversity for multiple packet transmissions," in *Proc. ICASSP*, vol. 4, pp. IV–181–4, April 2003.

[33] H. Samra, Z. Ding, and P. M. Hahn, "Symbol mapping diversity design for multiple packet transmissions," *IEEE Trans. Commun.*, vol. 53, pp. 810 – 817, May 2005.

[34] H. Samra and Z. Ding, "Symbol mapping diversity in iterative decoding/demodulation of ARQ systems," in *Proc. ICC*, vol. 5, pp. 3585 – 3589, May 2005.

[35] G. Caire, G. Taricco, and E. Biglieri, "The mimo arq channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.

[36] Y. S. Jung and J. H. Lee, "Hybrid-ARQ scheme employing different space-time trellis codes in slow fading channels," in *Proc. VTC*, vol. 1, pp. 247 – 251, Sep 2002.

[37] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, Mar. 1998.

[38] A. Lozano, A. M. Tulino, and S. verdu, "Optimum power allocation for parallel Gaussian channels with arbitary input distribution," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3033–3051, Jul. 2006.

[39] J. Yuan, Z. Chen, B. Vucetic, and W. Firmanto, "Performance and design of space-time coding in fading channels," *IEEE Trans. Commun.*, vol. 51, Dec. 2003.

[40] E. G. Larsson and P. Stocia, *Space-time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[41] S. M. Alamouti, "A simple transmit diversity techique for wireless communications," *IEEE J. Select Areas Commun.*, vol. 16, pp. 3072–3082, Oct. 1998.

[42] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from othogonal design," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456 – 1467, July 1999.

[43] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451 – 460, Mar 1999.

[44] X.-G. Xia, "Differentially en/decoded orthogonal space-time block codes with apsk signals," *IEEE Commun. Lett.*, vol. 6, April 2002.

[45] H. Jafarkhani, "A quasi-orhogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, Jan. 2001.

[46] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal non-orthogonality rate 1 space-time block code for 3+ Tx antennas," in *Proc. ISSSTA*, pp. 429–432, Sep. 2000.

[47] W. Su and X.-G. Xia, "Quasi-orthogonal space-time block codes with full diversity," in *Proc. GLOBECOM*, vol. 2, pp. 1098–1102, 17-21 Nov. 2002.

[48] E. N. Onggosanusi, A. G. Dabak, and T. M. Schmidl, "High rate space-time block coded scheme: Performance and improvement in correlated fading channel," in *Proc. WCNC*, vol. 1, pp. 194–199, Mar. 2002.

[49] L. He and H. Ge, "A new full-rate full-diversity orthogonal space-time block coding scheme," *IEEE Commun. Lett.*, vol. 7, Dec. 2003.

[50] L. Xian and H. Liu, "Space-time block codes from cyclic design," *IEEE Commun. Lett.*, vol. 9, Mar. 2005.

[51] H. Huang and H. Viswanathan, "Multiple antenan enhancements to high data rate (HDR) CDMA system," in *Proc. VTC*, May 2000.

[52] N. S. J. Pau, D. P. Taylor, and P. A. Martin, "Rate 2 quasi-orthogonal space time block codes using parallel interference cancellation," in *Proc. VTC spring*, vol. 5, pp. 2324–2328, Dec 2006.

[53] M. Uysal and C. N. Georghiades, "Non-orthogonal space-time block codes for 3 Tx antennas," *IEEE Electronics Lett.*, vol. 38, pp. 1689–1691, Dec. 2002.

[54] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-time Wireless Communications*. Cambridge University Press, 2003.

[55] A. Naguib, N. Seshadri, and A. Calderbank, "Increasing data rate over wireless channel," *IEEE Signal Processing Mag*, pp. 76–92, May 2000.

[56] H. Zhang, A. Lozano, and M. Haleem, "Multiple ARQ processes for MIMO systems," *EURASIP Applied Signal Processing*, pp. 772–782, May 2004.

[57] T. Koike, H. Murata, and S. Yoshida, "Hybrid ARQ scheme suitable for coded MIMO transmission," in *Proc. ICC*, vol. 5, pp. 2919 – 2923, June 2004.

[58] H. E. Gamal, G. Caire, and M. O. Damen, "The mimo arq channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3601–3621, Aug. 2006.

[59] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, pp. 1561–1571, Sep. 2001.

[60] A. Maaref and S. Aissa, "A cross-layer design for MIMO Rayleigh fading channels," in *Proc. Canadian Conference on Electrical and Computer Engineering*, 2004.

[61] A. Maaref and S. Aissa, "Adaptive modulation using orthogonal STBC in MIMO Nakagami fading channels," in *Proc. ISSSTA*, vol. 6, pp. 3818 – 3822, Nov. 2004.

[62] C. K. Ho and J. Oosteveen, "Rate adaptation in time varying channels using acknowledgement feedback," in *Proc. VTC-spring*, vol. 4, pp. 1683 – 1687, 2006.

[63] D. Gore and A. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Trans. Signal Processing*, vol. 50, pp. 2580–2588, Dec. 2002.

[64] W. H. Wong and E. G. Larsson, "Orthgonal space-time block coding with antenna selection and power control," *IEE Electronics Lett.*, vol. 39, pp. 379–381, Feb. 2003.

[65] R. Machado and B. F. Uchoa-Filho, "A hybrid transmit antenna / code selection scheme using space-time block codes," in *Proc. WCNC*, March 2004.

[66] D. M. Rankin, D. P. Taylor, and P. A. Martin, "Improved information outage rate in certain MIMO system," *IEEE Signal Processing Lett.*, vol. 13, pp. 393 – 396, July 2006.

[67] W. C. Jakes, *Microwave Mobile Communications*. IEEE Press, 1974.

[68] C. Komninakis, "A fast and accurate Rayleigh fading simulatior," in *Proc. GLOBECOM*, vol. 6, pp. 3306 – 3310, Dec. 2003.