MIMO Receiver Structures with Integrated Channel Estimation and Tracking

Yau Hee Kho
B. Eng. (Hons. I)

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

March 2008

ABSTRACT

This thesis looks at the problem of channel estimation and equalization in a multiple-input multiple-output (MIMO) dispersive fading environments. Two classes of MIMO receiver structure are proposed with integrated channel estimation and tracking. One is a symbol-by-symbol based receiver using a MIMO minimum mean square error (MMSE) decision feedback equalizer (DFE), and the other is a sequence-based receiver using a partitioned Viterbi algorithm (PVA) which approaches the performance of maximum likelihood sequence estimation (MLSE).

A MIMO channel estimator capable of tracking the time and frequency selective channel impulse responses, known as the vector generalized recursive least squares (VGRLS) algorithm, is developed. It has comparable performance and a similar level of complexity as the optimum Kalman filter. However, it does not require any knowledge of the channel statistics to operate and as such it can be employed in a Rician fading channel readily. A reduced complexity form of the estimator, known as the vector generalized least mean squares (VGLMS) algorithm, is also developed. This is achieved by replacing the online recursive computation of the VGRLS algorithm's 'intermediate' Riccatti matrix with an off-line pre-computed matrix. This reduces the complexity of the algorithm by an order of a magnitude, but at the expense of degraded performance.

The estimators are integrated with the above-mentioned equalizers in a decision directed mode to form a receiver structure that can operate in continuously time-varying fading channels. Due to decision delays, the outputs from the equalizer are delayed and this then produces 'delayed' channel estimates. A

simple polynomial-based channel prediction module is employed to provide up-to-date channel estimates required by the equalizers. However, simulation results show that the channel prediction module may be omitted for a very slowly fading channel where the channel responses do not vary much. In the case of the PVA-receiver, the zero-delay tentative decisions are used as feedback to the channel estimators with negligible loss.

ACKNOWLEDGEMENT

I am very grateful to my supervisor Professor Desmond Taylor for the opportunity to be his research student. His guidance and support throughout the course of my research has been indispensable. His valuable insights and comments have no doubt make this thesis a lot better.

This thesis would not have been possible without the financial funding from the University of Canterbury in the form of a Doctoral Scholarship. I am also grateful to the Institution of Engineering and Technology (IET), UK, for awarding me a Hudswell International Research Scholarship. I would also like to thank my supervisor for funding my presentation at PIMRC'07 in Athens, Greece.

In conducting this research I have learnt something from almost everyone that I come into contact with. Special thanks are due to Dr. Wing Seng Leon for his technical consultancy. I would also like to thank Dr. Peter Green, Dr. Peter Smith, Dr. Lee Garth and Dr. Philippa Martin for their advice. The companion of fellow Comms. Lab. students in room A221 has no doubt made postgraduate life more interesting and I appreciate that very much.

On a personal front, I am deeply indebted to my parents and parents-in-law for their love and support. I am very grateful to my wife, Vivy, for believing in me. Her encouragement, love and companion has made life a lot more meaningful, colourful and joyful. Thanks for the lovely kids - William, Nicholas and Clarissa - for the daily joy and laughter. I rejoice in blissful gratitude that God has always guided me onto the right path in His wisdom and everything happens for the best!

GLOSSARY

1G	first generation
2G	second generation
3G	third generation
4G	fourth generation
AMPS	Advanced Mobile Phone Services
AR	auto-regressive
ASIC	application specific integrated circuit
AWGN	additive white Gaussian noise
BER	bit error rate
BLAST	Bell Laboratories Layered Space Time Architecture
CCI	co-channel interference
CIR	channel impulse response
DDFSE	delayed decision feedback sequence estimation
DFE	decision feedback equalizer
DFMLSE	decision feedback MLSE
DPLL	digital phase-locked loop
DSP	digital signal processing
FBF	feedback filter
FFF	feed-forward filter
FIR	finite impulse response
FPGA	field programmable gate arrays
GLMS	Generalized Least Mean Squares
GRLS	Generalized Recursive Least Squares
GSM	Global System for Mobile Communications

IMTS	Improved Mobile Telephone Service
IIR	infinite impulse response
IP	internet protocol
IS-95	Interim Standard 95
IS-136	Interim Standard 136
ISI	intersymbol interference
LMS	least mean squares
LS	least squares
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum likelihood
MLSE	maximum likelihood sequence estimation
MMSE	minimum mean squared error
MRC	maximum ratio combining
MSD	mean square deviation
MSE	mean square error
NLOS	non line of sight
NMT	Nordic Mobile Telephone
OFDM	orthogonal frequency division multiplexing
PAM	pulse amplitude modulation
POTS	plain old telephone service
PSP	per-survivor processing
PVA	partitioned Viterbi algorithm
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
RLS	recursive least squares
RSSE	reduced-state sequence estimation
RW-RLS	rectangular windowed RLS
SER	symbol error rate
SIC	successive interference cancellation

SIMO	single-input multiple-output
SISO	single-input single-output
SMS	Short Message Service
SNR	signal-to-noise ratio
STBC	space time block codes
STC	space time coding
STTC	space time trellis codes
TDL	tapped delay line
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
US	uncorrelated scattering
VA	Viterbi algorithm
VAR	vector autoregressive
VBLAST	Vertical-BLAST
VGLMS	Vector Generalized Least Mean Squares
VGRLS	Vector Generalized Recursive Least Squares
VVA	vector Viterbi algorithm
WCDMA	Wideband Code Division Multiple Access
WLAN	wireless local area network
WLMS	Wiener LMS
WMF	whitened matched filter
WSS	wide sense stationary
WSSUS	wide sense stationary with uncorrelated scattering
ZF	zero-forcing

CONTENTS

Chapter 1

INTRODUCTION

The telecommunications industry, in particular the wireless sector, has gone through very rapid growth in recent years [1],[2],[3]. This has been largely driven by the increasing demand for high-quality digital communications at ever increasing data rates. During the past 15 years several so-called 'generations' (G) of wireless system have emerged and each generation is perhaps best epitomized by the distinct technological advances associated with it.

When the first generation (1G) systems [4], such as the Advanced Mobile Phone Services (AMPS) in North America and Nordic Mobile Telephone (NMT) in Europe, were developed in the 1970s and 1980s, *analogue* transmission techniques were used. In general terms, these systems are very similar to the older "0G" Improved Mobile Telephone Service (IMTS) service, but use considerably more computing power in order to select frequencies, hand off conversations to plain old telephone service (POTS) lines, and handle billing and call setup.

In the 1980s and 1990s when the second generation (2G) systems [4],[5], such as the Global System for Mobile Communications (GSM), Interim Standard (IS)-95, IS-136, were developed, *digital* technology was employed. The key advantage of these systems[1] to consumers has been better voice quality and the availability of low-cost alternatives to making calls, such as the Short Message Service (SMS).

Currently with third generation (3G) systems [6],[7],[8], such as Wideband Code Division Multiple Access (WCDMA), CDMA2000 and Time Division-Synchronous

[1]GSM and IS-136 employ time division multiple access (TDMA) technique while IS-95 uses code division multiple access (CDMA) technique.

Code Division Multiple Access (TD-SCDMA), the industry is undergoing a revolutionary transformation from low-date-rate, voice-dominated to *high-data-rate*, *multimedia-rich* system.

Even as the current deployment of 3G systems is still in its infancy, many research advances for fully internet protocol (IP)-based integrated systems, known as the fourth generation (4G) systems [6],[9], have already been developed. These include the like of Wi-Fi, WiMAX, etc. True 4G deployment will be achieved when wired and wireless technologies converge and will be capable of providing 100 Mb/s and 1 Gb/s speeds both indoors and outdoors, with both high quality and security.

Beyond being simply a replacement for the fixed-line telephony system, wireless technology has revolutionized many aspects of our lives - the way we communicate, conduct business, socialize and entertain ourselves [3]. Today's services promise to deliver many innovative applications beyond voice communication. These include 'killer' applications such as wireless broadband internet, picture messaging, live video streaming, location-aware services etc, and all with ubiquitous, '24-7', always-on connection. As exciting as it may sound, in reality, however, many technical challenges remain in designing robust systems to meet the requirement of these emerging applications [1].

Currently wireless local area networks (WLANs) employing the IEEE 802.11g/a standard offer up to 54 Mb/s [6]. However, even this is barely sufficient when faced with the demands of multimedia content. Therefore, high-data-rate wireless communications in the gigabit region has generated enormous interest in both the wireless industry and research community [10]. In theory, it is possible to build such systems using the current single antenna technology known as the single-input single-output (SISO) system. However, this would put tremendous pressure on the wireless system, and as we will explain, it is simply impossible to do so in practice.

Two fundamental elements that govern the performance of a wireless system are transmitted power and system bandwidth. It is intuitive that with higher

transmitted power giving higher system signal-to-noise ratio (SNR), a wireless system is better equipped to combat the distortion and noise encountered at the mobile radio channel. Consequently, fewer errors occur during transmission which results in better error rate performance. However, in reality the transmitted power near any human beings is capped by regulatory bodies at less than 1 W for the indoor environment due to biohazard considerations [10]. Even without this constraint, in practice the upper limit would be limited by the linearity of the power amplifiers currently used at the transmitter - the maximum power available without significantly distorting the data. Furthermore, the performance is also limited by co-channel and adjacent channel interferences.

Larger system bandwidth will enhance system capacity. However, the licensed frequency spectrum in which a wireless system operates is finite and limited, making bandwidth a very valuable commodity - billions of dollars were spent to occupy a slice of the 'free air' in the auctions of 3G spectrum in recent years [11]. In addition, any increase in bandwidth is hard to obtain particularly in the frequency band below 6 GHz where non line of sight (NLOS) networks are feasible [10]. This is the region where most wireless systems operate, including GSM, currently the most popular system in term of the number of subscribers worldwide. Note that NLOS is important to ensure wide coverage.

In order to create a 1 Gb/s link with a SISO system [10], either very high transmitted power or very large bandwidth is required, both of which far exceed the regulatory limits. Consequently new technical advances are needed.

1.1 OVERVIEW OF MIMO SYSTEMS

The information-theoretic papers of [12],[13] have shown that the capacity of a multiple-input multiple-output (MIMO) system can increase linearly with the number of antenna elements. Increased spectral efficiency can thus be achieved without putting further demands on the transmitted power and system bandwidth albeit at the expense of increased hardware cost and design complex-

ity. However, with the ever increasing power of field programmable gate arrays (FPGA), digital signal processing (DSP) chips and application specific integrated circuits (ASIC), the cost is beginning to drop significantly, making it affordable for mass deployment in the near future.

Defined simply, a MIMO system is a wireless communication system where there are multiple antenna elements at both the transmitter and the receiver. Its operation depends on space-time signal processing in which both the time and spatial domains are utilized. The time domain has been the traditional natural dimension in digital communication where signals are transmitted at allocated times; while the spatial domain is inherent in the use of multiple antenna elements where multiple cochannel transmission links are established.

A key feature of a MIMO system is its ability to exploit the multipath propagation often viewed as a pitfall in conventional wireless communication [14]. It does so by leveraging the intrinsic diversity provided by the multipath fading to improve the error rate performance. This improvement is most significant if the MIMO system is employed in a dense multipath environment. This is because the multipath propagation in such an environment creates many independent and almost uncorrelated 'virtual' links where the effect of fading in one link does not seriously affect the other links. Hence when one received signal is bad, there is an increased chance of detecting good signals in the other links.

The original approach to using multiple transmit and receive antennas was proposed by Foschini *et al* and is known as the Bell Laboratories Layered Space Time Architecture (BLAST) [15]. Together with Vertical-BLAST (VBLAST) [16], a simplified version of BLAST, such schemes are designed to maximize the system throughput in terms of bits per second per Hertz (b/s/Hz). Specifically they seek to improve spectral efficiency, by transmitting independent signals from multiple transmit antennas. A BLAST scheme typically relies on successive interference cancellation (SIC) [15] at the receiver to detect the signals. In doing so, however, it loses diversity gain due to the interference cancellation process. Moreover the scheme requires at least the same number of receive antennas as

transmit antennas. This constraint is overcome in [17] by proposing a new transmission format known as space time block codes (STBC). It provides transmit diversity in a 2 transmit and 1 receive antenna system by using an orthogonal signal design. This results in a diversity gain with no loss of system throughput[2]. Besides STBC, space time trellis codes (STTC) [18] were also proposed. These coding schemes typically provide good error rate performance achieved through diversity and/or coding gain at the expense of system throughput. They are collectively known as space time coding (STC) [19],[20],[21] and there have been much recent research aimed at improving their throughput. MIMO systems will be discussed in more detail in Chapter 2.

1.2 MOTIVATION FOR THESIS

A common feature of many of the above described MIMO techniques is the frequency flat fading environment in which they are assumed to operate, i.e. they were designed for a narrowband environment. Together with the assumption of perfect channel knowledge at the receiver, relatively simple system designs are possible, which are useful in studying the feasibility of the concept. In reality, however, channel knowledge is rarely available at the receiver and must be estimated. In addition, in a time-varying environment where the mobile terminals may be moving, continuous channel estimation or tracking is necessary.

As the data rate increases in a wideband system, the channel response becomes frequency selective and its delay spread relative to the symbol period becomes significant. This causes intersymbol interference (ISI) which is a critical impediment as it degrades system performance and introduces error floors even at high SNR. Orthogonal frequency division multiplexing (OFDM) [22] can be used to combat ISI as it transforms the frequency selective channel into parallel flat fading channels. Alternatively equalization can be used at the receivers, and this together with continuous channel estimation are the focus of this thesis.

[2]Orthogonal STBC with more than 2 transmit antennas suffer from a loss of throughput.

In [23], Bello proposes *Power Series* modelling of the mobile channel which includes the like of f-, t-, ft- and tf-power series expansions[3]. Recently, f- and t-power series models have been employed in receiver designs [24],[25],[26],[27],[28]. In particular, channel estimation techniques using t-power series model have been proposed in [25],[29],[30]. However these are to-date limited to SISO systems.

The power series models can represent the underlying fading channel efficiently using a reduced number of parameters. This is attractive for frequency selective MIMO systems because not only are there multiple channel links to estimate (and track), each of the links will have significant channel delay spread. The overall effect is a large set of parameters to estimate compared to SISO systems. The utilization of these channel models therefore allows a reduced complexity approach in channel estimation and tracking. With the channel estimator, it is then necessary to integrate it into appropriate receiver structures and to evaluate the overall performance.

The motivation for the thesis is therefore to develop channel estimation and tracking technique for both time and frequency selective MIMO fading channels[4], based on the t-power series techniques of [25],[29],[30], and to use the channel estimates to drive equalizer structures in a VBLAST-type (spatial multiplexing) transmission system.

1.3 THESIS OUTLINE

We have so far discussed the constraint of current SISO technology and explained the potential of a MIMO system in meeting the demands of high-data-rate applications. A brief overview of MIMO systems has been provided to motivate the work of the thesis. In the following we provide an outline of subsequent chapters.

In Chapter 2, required background information and literature survey is presented. This includes discussion of the building blocks of a typical SISO com-

[3]This refers to the expansion of a channel function, eg $C(f,t)$, in the f or t parameter, or sequentially in both the f and t, or t and f parameters. For details see [23].

[4]Also known as doubly selective channels.

munication system, followed by an extension to a MIMO system. The wireless channel is discussed in more detail. This includes methods to parameterize the channel and some common channel estimation techniques. The chapter concludes with a discussion of equalization methods.

In Chapter 3, we develop a vector polynomial predictor based channel estimator which we call a Vector Generalized Recursive Least Squares (VGRLS) estimator. We start by presenting the polynomial series model and the minimum noise gain polynomial predictor used. This is followed by a description of the statistics-based state transition matrix and how polynomial prediction can be used in the VGRLS estimator. We then analyze the effects of some parameters on its performance and evaluate it in terms of mean square deviation (MSD).

We present a reduced complexity form of the VGRLS estimator in Chapter 4 and call the resulting estimator a Vector Generalized Least Mean Squares (VGLMS) estimator. We describe the simplification process and analyze the complexity of the estimator. This is followed by an analysis on the effect of some parameters on the estimator's performance and concludes with a performance evaluation of the estimator.

In Chapter 5, we present an integrated symbol-by-symbol-based MIMO receiver design by combining a vector decision feedback equalizer (DFE) [31] with the VGRLS and VGLMS estimators. We start by describing the design of the vector DFE and how a polynomial predictor is used to bridge the time gap created by the equalizer's decision delay. This is followed by an evaluation through simulation of the error rate performance of the integrated receiver.

In Chapter 6, we present a reduced complexity sequence based equalizer with integrated channel estimation and tracking. This is achieved by combining the partitioned Viterbi algorithm (PVA) [32] with the VGRLS and VGLMS estimators. We describe the extension of the original PVA structure to accommodate adaptive channel tracking. A polynomial predictor is also used to bridge the time gap created by the equalizer's decision delays. We evaluate the performance of the resulting PVA-based receiver and compare it with one using a full vector

Viterbi algorithm (VVA) [33],[34],[35]. We also further investigate using tentative decisions as feedback data in the PVA receiver.

In Chapter 7 we present conclusions to the thesis by providing a brief review of the previous chapters and summarizing the contributions of the thesis. Possible future work arising from the thesis is also discussed.

1.4 THESIS CONTRIBUTIONS

The work of the thesis has resulted in MIMO receiver structures using symbol-by-symbol and sequence-based equalization techniques with integrated channel estimation and tracking for time and frequency selective MIMO fading channels.

Channel estimation is based on polynomial prediction by expanding the underlying channel using a Taylor's series in the time domain. This is known as the t-power series expansion. The resulting polynomial coefficients, i.e. the tap weights of a polynomial predictor, are derived *a priori* without requiring channel statistics. These are used as the coefficients of a state transition matrix of an approximate channel state model having *unforced dynamics*. Together with the measurement equation, these are then employed in a vector form of the generalized recursive least squares (GRLS) algorithm [29], known as the *vector GRLS* (VGRLS) estimator. This estimator is Kalman-like in structure and its performance can be made comparable

Further simplification of the VGRLS algorithm, following the approach of the *generalized least mean squares* (GLMS) algorithm [30], results in a reduced complexity form of the channel estimator which we call *vector GLMS* (VGLMS). This estimator has similar features to the VGRLS estimator but it offers much reduced complexity, achieved by replacing the online computation of the 'intermediate' matrix with an offline pre-computed matrix. This complexity reduction becomes significant in a large MIMO system with many antennas. However, it is achieved at the expense of degraded performance especially in a fast fading environment. Nevertheless, as shown in Chapter 4, the performance of the inte-

grated receiver shows that it still offers a nice trade-off in terms of complexity and performance in a slow fading environment, or in a Rician fading with strong mean components.

These estimators are integrated with both symbol-by-symbol and sequence-based equalizers in order to assess the error rate performance of a typical receiver system. For symbol-based equalization, we have employed the DFE which is attractive in terms of performance and complexity for practical applications. For sequence-based equalization, the PVA algorithm has been adapted to operate in a continuously fading environment. In both receivers, due to the decision delay of the equalizers, the estimators can only produce delayed estimates of the channel impulse response (CIR), whereas up-to-date estimates are required for the equalizer to operate properly. We address this problem by proposing the use of a polynomial prediction based channel predictor, and in the case of the PVA, using the tentative decisions (i.e., zero delay) as well. Although the equalization techniques used here have been well reported, the integration with the proposed channel estimation and tracking has resulted in novel MIMO receiver structures.

The work of this thesis has been published/accepted for publication in the following papers:

1. Y. H. Kho and D. P. Taylor, "MIMO Channel Estimation and Tracking Based on Polynomial Prediction with Application to Equalization", *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1585 - 1595, May 2008.

2. ——, "Reduced Complexity MIMO Channel Estimation and Equalization Using A Polynomial-Predictor Based Vector GLMS Algorithm", *Proceedings of IEEE Personal, Indoor and Mobile Radio Conference*, pp. 348 - 352, 3 - 7 Sept. 2007, Athens, Greece.

3. ——, "MIMO Receiver using Reduced Complexity Sequence Estimation with Channel Estimation and Tracking", *accepted for publication in IEEE Transactions on Vehicular Technology, 9 pages, to appear in Mar. 2009.*

Related to this thesis, the work involved in the investigation of the GLMS algorithm has also been presented in the following papers:

4. Y. H. Kho, "Reduced Complexity Channel Estimation and Tracking Based on Polynomial Prediction with Application to Equalization", 3^{rd} *prize, Postgraduate section, IEEE Region 10 Student Paper Contest*, Jul. 2007.

5. Y. H. Kho and D. P. Taylor, "A Reduced Complexity Kalman-like Algorithm for Channel Estimation and Equalization", *accepted for oral presentation at the* 2^{nd} *IET International Conference on Wireless, Mobile and Multimedia Networks, 12 - 15 Oct. 2008, Beijing, China.*

Chapter 2

BACKGROUND

In this chapter we present information required for a better understanding of subsequent chapters. A general overview of the architectures and building blocks for the communication system considered in the thesis work is presented. We start by describing a general single-input single-output (SISO) communication system where the information presented will be employed in a multiple-input multiple-output (MIMO) context in the following chapters. A detailed discussion of the physical mobile wireless channel is presented. This is followed by a discussion and literature survey of equalization techniques, and some general MIMO systems.

2.1 COMMUNICATION SYSTEM AND SIGNAL MODEL

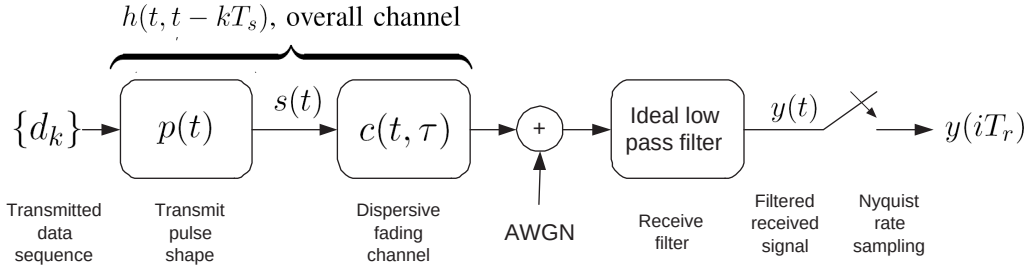

Figure 2.1 A general block diagram of a discrete-time SISO communication system at time k.

In the thesis we assume independent SISO communication systems for each of the subchannels in an overall MIMO system. We describe a SISO system where a

complex baseband model is shown in Fig. 2.1. The noiseless linearly modulated signal with bandwidth B at the transmitter can be expressed as

$$s(t) = \sum_{k=0}^{K-1} d_k p(t - kT_s) \tag{2.1}$$

where d_k is the k-th transmitted symbol from an M-ary complex constellation with a sequence length of K, T_s is the symbol period and $p(t)$ is the transmit pulse shape. The signal is transmitted over a fading dispersive channel with an instantaneous, time-varying impulse response $c(t, \tau)$ and maximum Doppler frequency f_D. The channel is assumed to be a zero-mean, complex Gaussian fading channel. At the front-end of the receiver, the signal is corrupted by additive white Gaussian noise (AWGN) with variance, σ_n^2. As the channel is random and unknown, matched filtering cannot be implemented. Instead an ideal low pass filter with bandwidth $B_{lpf} > B + f_D$ is employed. The filtered received signal is

$$y(t) = \sum_{k=0}^{K-1} d_k h(t, t - kT_s) + n(t) \tag{2.2}$$

where $n(t)$ is the filtered noise term and

$$h(t, t - kT_s) = \int_{-\infty}^{\infty} p(t - \tau - kT_s) c(t, \tau) d\tau \tag{2.3}$$

is the composite channel impulse response.

Although the response $p(t)$ is unlimited in time, we can in general assume that most of its energy is concentrated within a few symbol intervals and the truncated impulse response spans L_p symbol periods for $0 \leq t < L_p T_s$. We further assume that $c(t, \tau)$ is non-zero over the delay range $0 \leq \tau < \tau_{max}$ where the maximum delay spread $\tau_{max} = L_c T_s$. Therefore $h(t, t - kT_s)$ may be assumed non-zero only on the interval $(t/T_s) - L < k < (t/T_s)$ where $L = \lceil L_p + L_c \rceil$. The received signal $y(t)$ is sampled at the Nyquist rate of $f_s = \frac{1}{T_r} = 2B_{lpf}$ where $T_r = \frac{T_s}{N_r}$ and N_r is the number of samples per symbol interval. Usually $N_r = 2$ is sufficient. This ensures that the noise samples are uncorrelated. The fractionally-spaced received

samples are defined as $\mathbf{Y} = \{y(0), y(T_r), \cdots, y\left((K+L-1)N_r - 1\right)T_r\}$ where

$$y(iT_r) = \sum_{k=-L+1+\lfloor \frac{i}{N_r} \rfloor}^{\lfloor \frac{i}{N_r} \rfloor} d_k h\left(iT_r, (i-kN_r)T_r\right) + n(iT_r). \tag{2.4}$$

We will employ the above SISO description of the communication system and extend it into a MIMO equivalent in the subsequent chapters.

2.2 MOBILE WIRELESS CHANNEL

The mobile wireless channel is a primary source of performance degradation in any wireless communication system due to multipath fading. This is also observed in communication channels such as HF shortwave ionospheric, UHF troposcatter and VHF ionospheric forwrad scatter [36] which causes attenuation, delay and phase shift in the transmitted signal. The aim of a communication receiver is to recover the transmitted signal as faithfully and reliable as possible. An understanding of the characteristics of the channel is therefore important in designing a receiver that can do so. Here we briefly discuss the characterization of a mobile wireless channel. Detailed information on this topic can be found in [37].

2.2.1 Fading mechanism

Multipath fading occurs due to interferences from multiple propagation paths in the channel, or medium, in which the signal is transmitted and received. These paths arise due to reflection, refraction, or diffraction encountered in the channel, as shown in Fig. 2.2. The amplitude and phase of each path vary in time due to changes in the structure of the medium. As a result, the received signal consists of the sum of multiple time-variant versions of the original transmitted signal, delayed and scaled by the multipath channel. The interferences may be constructive or destructive, depending on the relative amplitudes and phases of the multiple paths, and this results in change in the received signal level. The

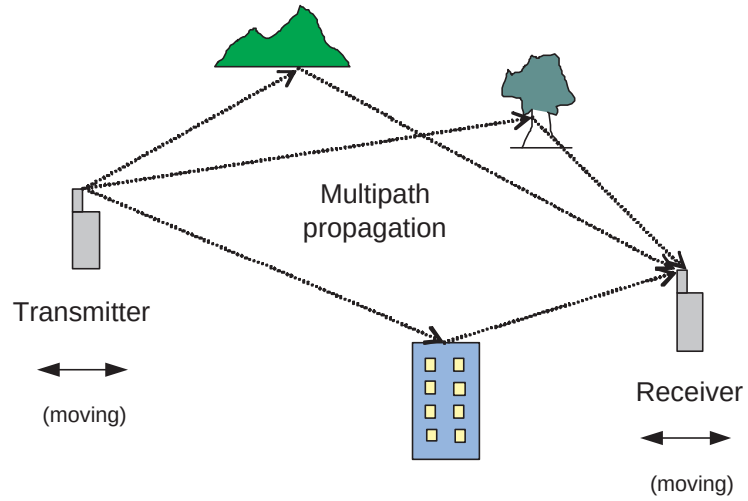

Figure 2.2 The phenomenon of multipath fading where multiple copies of the transmitted signal are received.

received signal may also experience dispersion, defined as spreading of the signal in time or frequency. As such the fading is also known as *dispersive fading*.

Consider a general bandpass transmitted signal written as

$$\bar{s}(t) = Re\{s(t)exp(j2\pi f_c t)\} \tag{2.5}$$

where $s(t)$ is the linearly modulated signal given in (2.1) and f_c is the carrier frequency. Assuming there are P propagation paths in the channel where each path is characterized by a gain (or attenuation) element, a_p, and an associated path delay, τ_p, the resultant bandpass received signal is the sum of the multiple attenuated and delayed version of the transmitted signal, may be written as

$$\begin{aligned} \bar{y}(t) &= Re\left(\sum_{p=1}^{P} a_p(t)s\left(t-\tau_p(t)\right)exp\left(j2\pi f_c\left(t-\tau_p(t)\right)\right)\right) \\ &= Re\left(\left[\sum_{p=1}^{P} a_p(t)exp\left(-j2\pi f_c\tau_p(t)\right)s\left(t-\tau_p(t)\right)\right]exp(j2\pi f_c t)\right) \end{aligned} \tag{2.6}$$

where the equivalent complex baseband form of the received signal is

$$y(t) = \sum_{p=1}^{P} a_p(t) exp(-j2\pi f_c \tau_p(t)) s(t - \tau_p(t)). \tag{2.7}$$

We define the channel impulse response $c(t, \tau)$ as

$$c(t, \tau) = \sum_{p=1}^{P} a_p(t) exp(-j2\pi f_c \tau_p(t)) \delta(t - \tau_p(t)) \tag{2.8}$$

such that the received signal can be expressed as

$$y(t) = \int_{-\infty}^{\infty} s(t - \tau) c(t, \tau) d\tau. \tag{2.9}$$

This equation describe the time-varying nature of the multipath effects on the transmitted signal through $c(t, \tau)$ where $c(t, \tau)$ denotes the response of the channel at time t due to an impulse applied at time $t - \tau$.

The modelling of a dispersive fading channel is based on the use of mathematics to describe the physical or observed properties of the channels. Bello [23] described a channel model using a tapped delay line representation which is based on knowledge of the correlation properties of the channel. With respect to the delay index τ, $c(t, \tau)$ is considered as the time-varying tap gain at delay τ. In term of various paths, $c(t, \tau)$ is the sum of the complex gains of all paths with delay τ and if the terminal is moving, the sum will be time-varying.

Since the fading process is random and not known *a priori*, a statistical description of the channel is necessary by viewing it as a stochastic process. It is assumed that a large number of paths exist such that central limit theorem can be applied. The channel impulse response can then be represented as a complex Gaussian process, $c(t, \tau)$, which captures the time-varying nature of the channel.

Under the complex Gaussian assumption, the channel is fully characterized by its ensemble mean

$$\overline{c(t, \tau)} = E[c(t, \tau)] \tag{2.10}$$

and its second moment or autocovariance

$$R_c(t_1, t_2; \tau_1, \tau_2) = E[\{c(t_1, \tau_1) - \overline{c(t_1, \tau_1)}\}\{c(t_2, \tau_2) - \overline{c(t_2, \tau_2)}\}^*] \tag{2.11}$$

where * denote complex conjugation.

By taking the Fourier transform with respect to τ, the time-varying transfer function of the channel, which is the dual representation of $c(t, \tau)$ in the frequency domain is obtained as

$$C(f, t) = \int_{-\infty}^{\infty} c(t, \tau) e^{-j2\pi f \tau} d\tau. \tag{2.12}$$

As the channel is a zero mean complex Gaussian random process, and given that the Fourier transform is a linear operation [38], (2.12) retains the same statistics as $c(t, \tau)$.

When the complex gains due to different scatterers have similar amplitudes and there is no direct or dominant path in the channel, the function $c(t, \tau)$ (and $C(f, t)$) is Gaussian with zero mean, i.e., $\overline{c(t, \tau)} = 0$. The received envelope then has a Rayleigh distribution and the fading channel is known as a Rayleigh fading channel. Most treatments of digital communications over fading channels have focussed primarily on Rayleigh fading channels. This is mostly due to the wide acceptance of the model in describing the fading effects on many radio channels and its mathematical tractability. When a dominant or line-of-sight path is present, a non-zero mean is present and the envelope will have a Rician distribution. The resulting channel is then a Rician fading channel. $c(t, \tau)$ can be decomposed into a specular (non-random) and diffuse (random) component, where

$$c_s(t, \tau) = E[c(t, \tau)] \tag{2.13}$$

which is the channel mean and

$$c_d(t,\tau) = c(t,\tau) - E[c(t,\tau)] \tag{2.14}$$

which is Rayleigh faded.

It is further usually assumed that the fluctuations of the channel are *wide sense stationary* (WSS). This assumption is based on the fact that on a sufficiently small time scale and bandwidth usually associated with short-term, or small-scale fading, the fluctuations of the channel in time and frequency can be assumed to be stationary. Hence, we can assume $c(t,\tau)$ to be stationary in a time sense. Further the channel may be modelled as a continuum of uncorrelated scatters such that $c(t,\tau)$ is independent for different values of delay, τ and the channel is said to exhibit *uncorrelated scattering* (US). When the time-varying impulse response is assumed to have stationary fluctuation in time and frequency, the channel is considered to be wide sense stationary with uncorrelated scattering (WSSUS).

2.2.2 Channel functions

The WSSUS assumption enables the channel to be completely determined statistically through its second order statistics, i.e., auto-correlation function or its Fourier transform the power spectral density. Due to the time and frequency duality nature of the channel, there exist several autocorrelation functions or power spectral densities that are used to characterize the channel.

One of them is the *tap gain correlation function* given by

$$R_c(t, t+\triangle t; \tau_1, \tau_2) = R_c(\triangle t, \tau_1)\delta(\tau_1 - \tau_2). \tag{2.15}$$

By setting $\triangle t = 0$, (2.15) becomes $R_c(0,\tau) = \sigma_c^2(\tau)$ which is the multipath intensity profile or delay power profile. This describes the distribution of the average power of the channel with respect to the delay parameter and allows the evaluation of multipath delay spread[1] τ_d which is the interval over which $R_c(0,\tau)$

[1] Also known as the maximum delay spread τ_{max} of the channel.

is effectively non-zero. Under the WSSUS assumption, (2.15) can usually be written in a product form as

$$R_c(\triangle t, \tau) = \sigma_c^2(\tau) R_c(\triangle t). \tag{2.16}$$

where $\sigma_c^2(\tau)$ has the same meaning as $R_c(0, \tau)$ and $R_c(\triangle t)$ is the normalized autocorrelation function. The Fourier transform of $R_c(\triangle t)$ gives $S_C(\upsilon)$ which is the Doppler spectrum of the channel.

The other function is the time-frequency correlation function defined as

$$\begin{aligned} R_C(t_1, t_2; f_1, f_2) &= E\left[C(t_1, f_1) C^*(t_2, f_2)\right] \\ &= E\left[C(t, f_1) C^*(t + \triangle t, f_2)\right] \\ &= \int_{-\infty}^{\infty} R_c(\triangle t, \tau) e^{-j2\pi \triangle f \tau} d\tau \\ &= R_C(\triangle t, \triangle f) \end{aligned} \tag{2.17}$$

which is the autocorrelation function of the time-varying transfer function (2.12). It represents the cross-correlation function between the complex envelopes of received carriers $\triangle f$ apart. Since the transfer function is assumed stationary with uncorrelated scattering, (2.17) is dependent only on the frequency and time separation. Note that (2.17) is the Fourier transform of (2.15) in τ.

The Fourier transform of (2.15) with respect to t yields the *channel scattering function* which is defined as the power spectrum of the complex gain fluctuation at delay τ

$$S_c(\upsilon, \tau) = \int_{-\infty}^{\infty} R_c(\triangle t, \tau) e^{-j2\pi \upsilon \triangle t} d\triangle t. \tag{2.18}$$

It exhibits the delay and Doppler spreading characteristics of the dispersive channel. The width in υ is the two-sided Doppler bandwidth B_D (or Doppler spread) and is equal to $2f_D$.

In the same manner, the *spaced-frequency Doppler spread correlation* is given

by

$$S_C(\upsilon, \triangle f) = \int_{-\infty}^{\infty} R_C(\triangle t, \triangle f) e^{-j2\pi\upsilon\triangle t} d\triangle t. \tag{2.19}$$

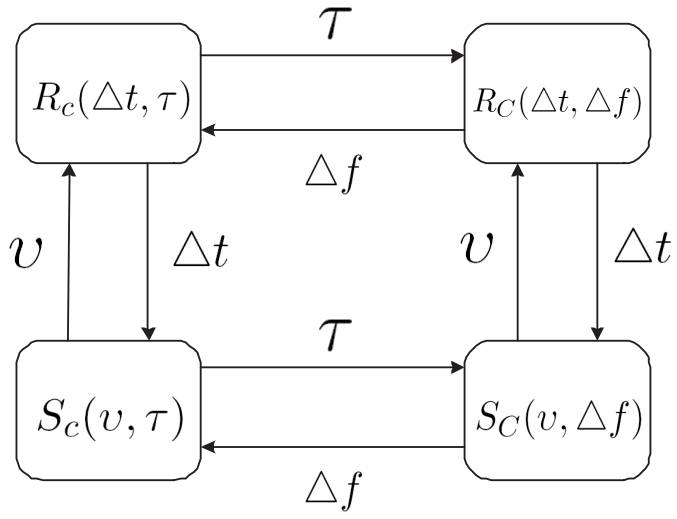

Figure 2.3 The relationships of various channel correlation functions by Fourier transformation as indicated by arrow with respect to the variable next to it.

The Fourier transform relationships among the four correlation functions are shown in Fig. 2.3 and more detail can be found in [23],[36].

2.2.3 Channel Classification

From (2.19), if $\triangle f = 0$, we have

$$S_C(\upsilon) = \int_{-\infty}^{\infty} R_C(\triangle t) e^{-j2\pi\upsilon\triangle t} d\triangle t. \tag{2.20}$$

which is the Doppler power spectrum of the channel and gives the power at the output of the channel as a function of the Doppler variable υ. The range over which $S_C(\upsilon)$ is non-zero is called the Doppler spread, B_D, of the channel. The Fourier transform of $S_C(\upsilon)$ gives $R_c(\triangle t)$ the normalized autocorrelation function which is depicted in Fig. 2.4. The channel is also characterized by a *coherence time*, T_c, which represents the time over which the received signal

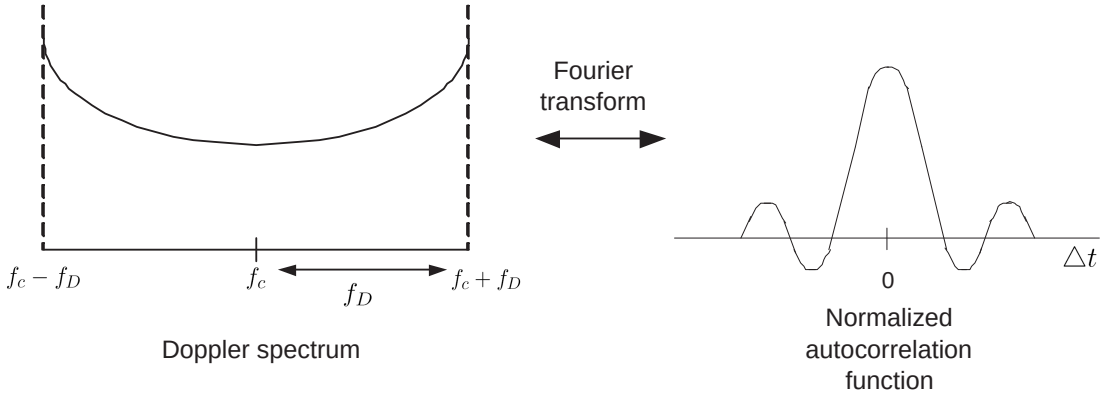

Figure 2.4 Correlation and power density function for Rayleigh fading channel: Doppler spectrum and normalized autocorrelation function.

can be considered coherent and it is roughly equal to the inverse of the Doppler spread, $\frac{1}{B_D}$. Within T_c the channel is effectively time-invariant.

Similarly, by setting $\triangle t = 0$ in (2.17), we have

$$R_C(\triangle f) = \int_{-\infty}^{\infty} R_c(\tau) e^{-j2\pi f\tau} d\tau. \tag{2.21}$$

The range of frequency over which $R_C(\triangle f)$ is non-zero is known as the *coherence bandwidth*, B_c, of the channel. Within this bandwidth, the multipath fading is flat as all frequency components of the signal are sent through the channel with equal gain and change linearly in phase. Numerically, it is roughly equals the inverse of the multipath delay spread of the channel, $\frac{1}{\tau_d}$. We note that the Fourier transform of $R_C(\triangle f)$ gives $\sigma_c^2(\tau)$, the delay power profile of the channel and the relationship is depicted in Fig. 2.5.

Using these parameters, a channel can be classified as fast or slow, frequency flat or selective fading in relation to the transmitted signal. The channel will exhibit fast fading if the symbol period T_s of the signal is greater than the coherence time of the channel and the bandwidth of the signal B is smaller than the coherence bandwidth of the channel, i.e. $T_s > T_c$ and $B < B_D$, and slow fading if $T_s \ll T_c$ and $B \gg B_D$. On the other hand, the channel will exhibit frequency flat fading if the signal bandwidth is greater than the coherence bandwidth and

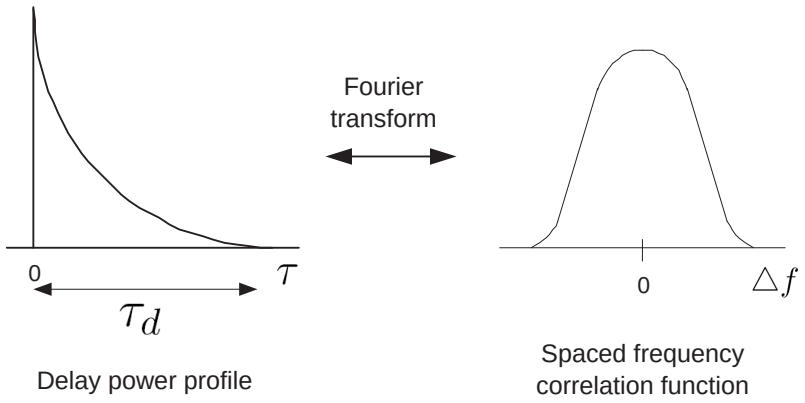

Figure 2.5 Correlation and power density function for Rayleigh fading channel: Delay power profile and spaced frequency correlation function.

the symbol period is smaller than the multipath delay spread, i.e. $B > B_c$ and $T_s < \tau_d$, and frequency selective fading if $B \ll B_c$ and $T_s \gg \tau_d$.

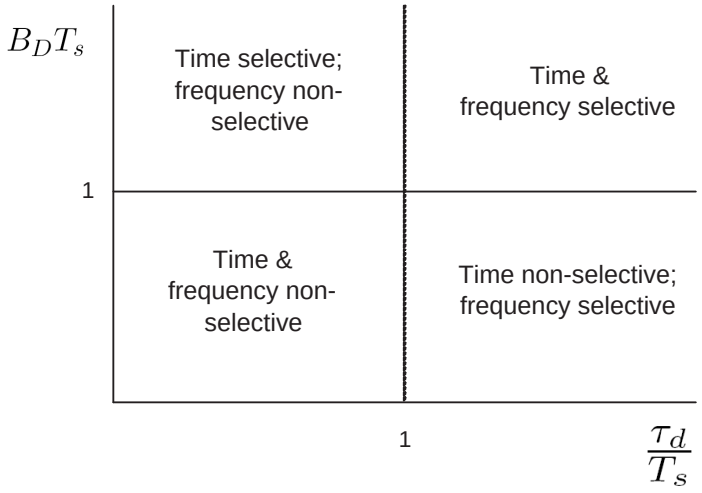

Figure 2.6 The classification of the fading channel in relation to the values of $B_D T_s$ and $\frac{\tau_d}{T_s}$.

The above parameters by themselves are rather meaningless. It is common practice to normalize the parameters with respect to the symbol period as $\frac{T_c}{T_s}$, $B_D T_s$ (or $f_D T_s$), $B_c T_s$ and $\frac{\tau_d}{T_s}$ for ease of comparing the effect of different parameter values. In general, based on the values of $B_D T_s$ and $\frac{\tau_d}{T_s}$ the channel can be classified as belonging to one of four channel classes [39],[37] as shown in Fig. 2.6. In the first, the channel is both time and frequency non-selective ($\frac{\tau_d}{T_s} \ll 1$, $B_D T_s \ll 1$). The received signal is scaled by a complex gain, so equalization con-

sists merely of estimating the channel gain and phase, where $c(t,\tau) = c\delta(t)$. The second class comprises time-invariant ($B_D T_s \ll 1$), frequency-selective channels. The channel response varies with frequency across the bandwidth of the transmitted signal but it changes slowly compared to the symbol rate. The impulse response may be considered as that of a linear, time-invariant filter $c(t,\tau) = c(0,\tau)$, which causes quasi-constant ISI between adjacent symbols. Equalizers have been historically developed for such channels. In the third class, the channels are time-selective but frequency non-selective ($\frac{\tau_d}{T_s} \ll 1$). The main effect is a time-varying complex attenuation, $a(t)$, which affects all frequency components equally and $c(t,\tau) = a(t)\delta(\tau)$. The fourth class comprises of both time and frequency selective (known as doubly selective) channels. The response varies significantly across the signal bandwidth and is time-varying. In this thesis, we are concerned with the fourth class.

Another parameter $B_D\tau_d$ [36] is used to define a channel that exhibits both time and frequency selectivity, known as a doubly selective channel. If $B_D\tau_d > 1$ the channel is overspread, otherwise it is underspread [40]. Overspread channels, such as underwater channels, are extremely difficult if not impossible to estimate, whereas for underspread channels the impulse response may be estimated although the difficulty increases as $B_D\tau_d$ nears unity.

2.2.4 Channel Models

With (2.9), the channel can be interpreted as a continuum of scatterers [23], each scatterer being associated with a complex attenuation $c(t,\tau)$ corresponding to delays in the range $(\tau, \tau + d\tau)$. However, as shown in an earlier section, since the transmitted signal and channel process are bandlimited, the received signal is also bandlimited. It follows from the sampling theorem that there is no loss of information if the received signal is sampled every T_r seconds, where $T_r = T_s/r$ and r is chosen large enough to satisfy the Nyquist sampling criterion. From (2.9), the sampled received signal can be written as

$$y(iT_r) = \int_{-\infty}^{\infty} s(iT_r - \tau)c(iT_r, \tau)d\tau. \tag{2.22}$$

The bandlimited transmitted signal can also be expressed as

$$s(t) = \sum_{m=-\infty}^{\infty} s_m sinc\left(\frac{1}{T_r}(t - mT_r)\right) \tag{2.23}$$

where s_m is the sample of $s(t)$ at time $t = mT_r$. From (2.22) and (2.23),

$$\begin{aligned} y(iT_r) &= \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} s_m sinc\left(\frac{1}{T_r}(iT_r - mT_r - \tau)\right) c(iT_r, \tau)d\tau \\ &= \sum_{m=-\infty}^{\infty} s_{i-m} \int_{-\infty}^{\infty} c(iT_r, \tau) sinc\left(\frac{1}{T_r}(mT_r - \tau)\right) d\tau \\ &= \sum_{m=-\infty}^{\infty} s_{i-m} c_{i,m} \end{aligned} \tag{2.24}$$

where we define

$$c_{i,m} = \int_{-\infty}^{\infty} c(iT_r, \tau) sinc\left(\frac{1}{T_r}(mT_r - \tau)\right) d\tau. \tag{2.25}$$

Equation (2.24) shows that a transversal filter can be used to represent the channel as shown in Fig. 2.7. This model is commonly known as the tapped delay line (TDL) model.

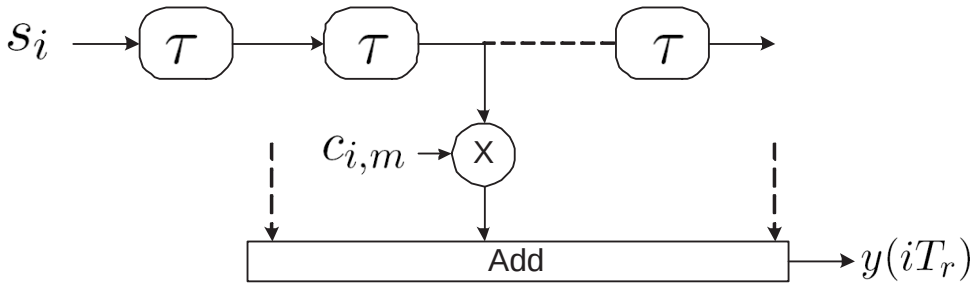

Figure 2.7 Sampled time tapped delay line model for a multipath fading channel with T_r-spaced taps.

Besides the TDL model, alternative methods using *power series* based models

have been reported [23] and these include t-, f-, tf- and ft- power series models. A t-power series model expands the channel response in the time domain as a N-th order polynomial in a small finite interval of interest $|t - \eta T_s| \leq \varphi T_s$ about $t = \eta T_s$ using Taylor's theorem [25],[23],[26] to obtain

$$c(t,\tau) = \sum_{n=0}^{N} a_\eta^{(n)}(\tau) b_\eta^{(n)}(t) + R_N(t,\tau) \tag{2.26}$$

where the coefficients are given by

$$a_\eta^{(n)}(\tau) = \frac{T_s^n}{n!} \left[\frac{d^n c(t,\tau)}{dt^n} \right]_{t=\eta T_s} \tag{2.27}$$

with the elementary basis functions,

$$b_\eta^{(n)}(t) = \left(\frac{t - \eta T_s}{T_s} \right)^n \tag{2.28}$$

and $R_N(t,\tau)$ is the remainder term, or residual (modeling) error, of the Taylor series, given by

$$R_N(t,\tau) = \frac{(t - \eta T_s)^{N+1}}{(N+1)!} \frac{d^{N+1}}{dt^{N+1}} c(t,\tau). \tag{2.29}$$

Given that the Doppler spread is normally much smaller than the sampling rate, only the first N terms in the expansions (3.10) are significant [25],[23],[26]. The remaining higher order terms can be ignored as long as the expression is used only in a small vicinity of ηT_s. This enables the channel response to be approximated using a small number of parameters.

The f-power series model is the dual of the t-power series model and expands the channel response in the frequency domain. It is also known as the reduced dimensionality model in [41]. If the delay spread of the channel is not too long, it is found that only the first few terms of the series are necessary to give a good approximation of the channel [23],[27],[41].

2.2.5 Channel Simulations

A statistical description of the fading channel was reported in [42] which is further adopted by [43] in developing the classical Jakes' fading model. This model has been widely used and referenced in many research publications. It is a deterministic method for simulating a time-correlated Rayleigh fading waveform [43] with the following autocorrelation function assuming isotropic scattering [42]

$$R_c(\triangle t, \tau) = \sigma_c^2(\tau) J_0(2\pi f_D |\triangle t|) \tag{2.30}$$

where σ_c^2 is the total average power of the channel and $J_0(\cdot)$ is the zeroth order Bessel function of the first kind.

The corresponding normalized Doppler spectrum is the classical U-shaped spectrum given by

$$S_C(\upsilon) = \begin{cases} \frac{1}{\pi f_D \sqrt{1-\frac{\upsilon^2}{f_D^2}}} & |\upsilon| \leq f_D \\ 0 & \text{otherwise} \end{cases} \tag{2.31}$$

The Rayleigh fading process can be generated by filtering zero-mean complex white noise using a low pass correlation filter [44], also known as a Doppler filter as it produces correlated Doppler fading. The autocorrelation of the fading process is given by (2.30) and the frequency response of the filter is given by $\sqrt{S_C(\upsilon)}$ where $S_C(\upsilon)$ is given by (2.31). The corresponding impulse response of the filter is [44]

$$h_d(t) = \begin{cases} \frac{\sqrt[4]{f_D \pi}}{\Gamma(5/4)} & \text{t} = 0 \\ \frac{J_{1/4}(2\pi f_D t}{\sqrt[4]{t}}) & \text{otherwise} \end{cases} \tag{2.32}$$

where $J_{1/4}(\cdot)$ is the one-fourth order Bessel function of the first kind. Due to its implementation as an FIR filter, the impulse response of (2.32) has heavy tails and a Hanning window is used to produce a smooth fading process. A plot of the Rayleigh fading process generated using the above method is shown in Fig. 2.8.

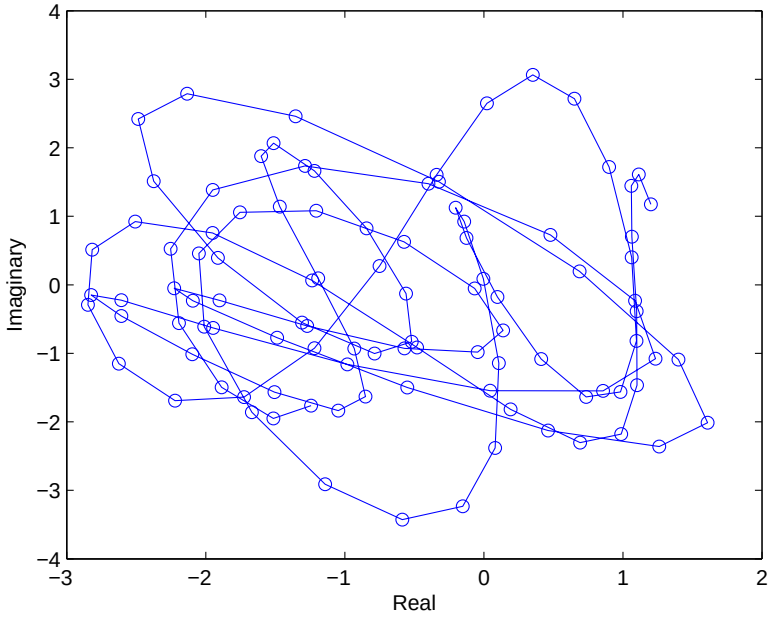

Figure 2.8 Rayleigh fading process with a normalized fade rate $f_D T_s = 0.1$. The markers represents symbol intervals.

A frequency selective fading channel may be implemented using the TDL model of Fig. 2.7 where each of the complex gains $h(t, \tau_p)$ is implemented as a Rayleigh fading generator. The normalized fading process is first weighted according to the delay power profile of the channel before being used as the tap gain of the CIR. A Rician fading channel may be implemented by using a constant tap gain for the first branch of the TDL model.

2.2.6 MIMO Channel Models

In general, the above characterizations of the SISO channel are applicable to a MIMO channel. However, some degree of signal correlation normally exists due to the multiple antennas at the transmitter and receiver. This correlation is a complicated function of the scattering environment and the antenna spacing. We assume here sufficiently spaced antennas that the signal correlation between adjacent antenna is rendered insignificant, and a rich scattering environment such that the channel gains become independent and identically distributed (uncor-

related). In this case, each of the MIMO subchannels can be described by the SISO characterization. Detailed information on MIMO channel modelling can be found in Chapter 6 of [45] and the references within.

2.3 EQUALIZATION

As we have seen in the previous section, the radio channel in which a wireless system operates is usually a multipath fading channel where there is a large performance penalty compared to an AWGN channel. For an AWGN channel, the asymptotic decrease of the bit error rate (BER) performance in relation to SNR has an exponential relationship, but for a Rayleigh flat fading channel, the asymptotic decrease follows an inverse law [36]. This means that a large SNR is necessary to achieve acceptable BER. In frequency selective fading, the performance degradation is even more severe as ISI causes an irreducible error floor at high SNR. Equalization is needed to reduce the error floor.

As signal processing provides a powerful mechanism to mitigate the effect of ISI, in a broad sense, equalization can be thought of as any signal processing techniques used at the receiver to mitigate the ISI problems. For time-varying channels, it is necessary to employ adaptive equalizers that track the time-varying channel response.

Fig. 2.9 shows a typical equalization scheme in a digital communication system. Although adaptive equalizers that do not require explicit channel estimates explicitly, as shown in Fig. 2.9(a), can be used, in this thesis we are concerned with equalizers that operate with explicit channel estimates as shown in Fig. 2.9(b). These equalizers usually have better performances. For example a DFE designed using explicit channel estimates is more robust against channel time variations [46] compared to one that is implemented with adaptive tap coefficients tracking. A comprehensive review of the extensive research for the problem of adaptive equalization for a SISO dispersive channel has been conducted by [39].

In the following we briefly describe equalization techniques using a linear

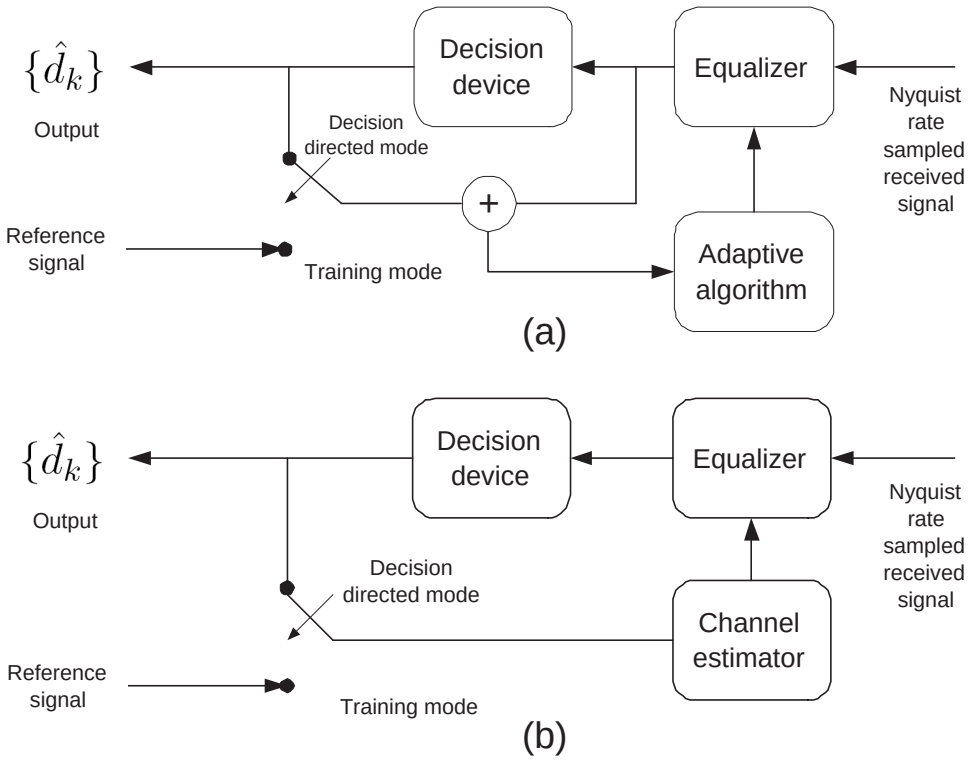

Figure 2.9 Equalization techniques in a digital communication system: (a) using adaptive algorithm and (b) using explicit channel estimates.

transversal filter, a DFE, an optimal sequence-based technique using maximum likelihood sequence estimation (MLSE) and a reduced complexity sub-optimal MLSE method using the partitioned Viterbi algorithm (PVA).

2.3.1 Linear Equalization

Linear equalization is a popular per-symbol equalization technique that can be implemented as a tapped delay line filter structure with optimized tap weights [47]. It attempts to compensate the distortions imposed on the received signals by the time variations of the channel. In practice, at start-up, the receiver does not have knowledge of the channel dynamics, and the channel may also vary in time. Hence, a linear equalizer must be implemented as an adaptive filter which can converge to a solution that satisfactorily reduces the error rate, and can then track time variations in this solution as the channel varies in time.

The first adaptive equalizer or filter design is often credited to [48] for the design of a zero-forcing (ZF) equalizer in 1966. A ZF linear equalizer attempts to cancel the effect of the channel by approximating the inverse of the channel frequency response. The tap weights are chosen such that all but one of the combined channel and equalizer samples are zero. It is relatively simple to design and implement. However, it suffers from noise enhancement as it does not take into account the effect of additive noise in its design. When the channel experiences a spectral null (which results in relatively small response), the inverse of the response tend to be large and this will excessively enhance the noise.

This problem of noise enhancement may be alleviated by using the mean square error (MSE) design criterion under which ISI mitigation is balanced with noise enhancement. It is defined as the mean square value of the error between the desired output and the actual equalizer output. It is shown [36] that when the additive noise approaches zero, the MSE criterion and the ZF criterion yield the same set of tap coefficients, which results in complete elimination of ISI. On the other hand, in the presence of noise, the MSE criterion produces better optimized equalizer coefficients than the ZF criterion.

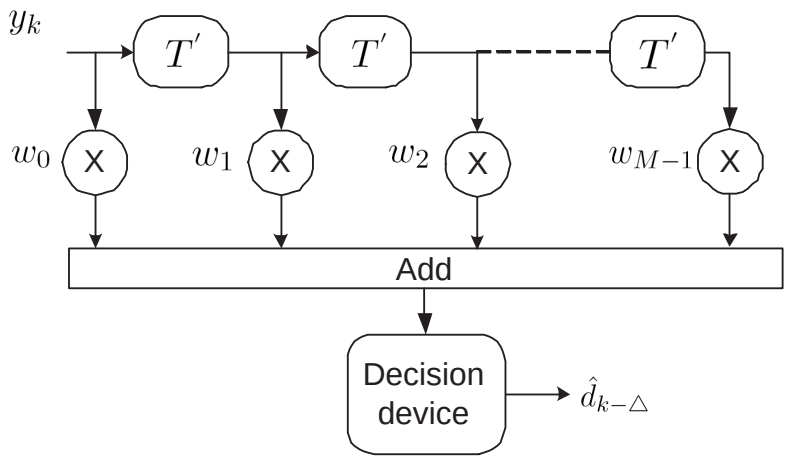

Figure 2.10 A linear equalizer of length M.

Consider a linear equalizer of length M with a vector of tap coefficients $\mathbf{w} = [w_0, w_1, \cdots, w_{M-1}]^T$ as shown in Fig. 2.10. The input vector at time kT' is $\mathbf{y}_k = [y_k, y_{k-1}, \cdots, y_{k-M}]^T$ where $y_k = y(kT')$ and T' is the delay between successive

taps. The objective is to minimize the mean squared error

$$\varepsilon_{mse} = E\left[|z_k - d_{k-\triangle}|^2\right] \tag{2.33}$$

where the output of the equalizer is $z_k = \mathbf{w}^H \mathbf{y}_k$, the desired response is $d_{k-\triangle}$ and $\triangle$ is the decision delay.

The MMSE tap weights are the solution to the Wiener-Hopf equation [49] which is expressed as

$$\mathbf{w} = \mathbf{R}_y^{-1} \mathbf{P}_\triangle \tag{2.34}$$

where $\mathbf{P}_\triangle = E[d^*_{k-\triangle}\mathbf{y}_k]$ and $\mathbf{R}_y = E[\mathbf{y}_k \mathbf{y}_k^H]$. The MMSE may be expressed by

$$\varepsilon_{mmse} = 1 - \mathbf{P}_\triangle^H \mathbf{R}_y^{-1} \mathbf{P}_\triangle \tag{2.35}$$

2.3.2 Decision Feedback Equalization

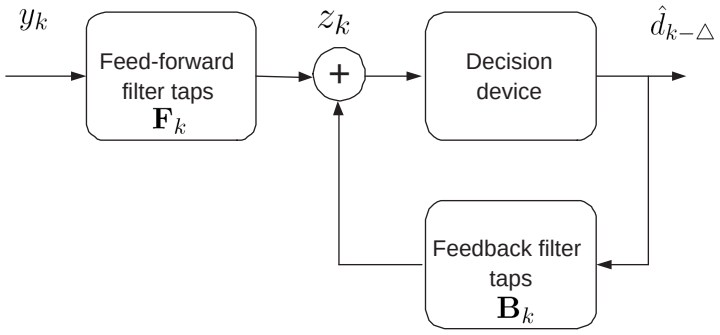

Figure 2.11 A vector MMSE DFE.

The DFE [50] is a well-known receiver structure for communication channels with severe amplitude distortion and other bad channel characteristics. The DFE decodes channel inputs on a symbol-by-symbol basis and uses past decisions to remove trailing ISI. It contains a feed-forward filter (FFF) and a feedback filter (FBF) as shown in Fig. 5.1. The FFF is a linear transversal equalizer. The detected symbols are assumed correct and fed back to cancel the ISI which is

attributed to them. The tap coefficients of the FBF are in fact the tail of the overall response of the channel and FFF. The coefficients of the FFF and FBF are optimized jointly.

We define $\mathbf{F}_k$ as the tap weight vector of the FFF of length N_f and $\mathbf{B}_k$ the FBF taps of length N_b. The input vector to the FFF at interval $kT^{'}$ is $\mathbf{y}_k = [y_k, y_{k-1}, \cdots, y_{k-N_f}]^T$ and the input vector to the FBF at the same interval is $\mathbf{d}_k = [d_{k-\triangle-1}, d_{k-\triangle-2}, \cdots, d_{k-\triangle-N_b}]^T$ where $\triangle$ is defined as the decision delay. Let $\mathbf{w}_{dfe} = [\mathbf{F}_k^T, \mathbf{d}_k^T]^T$ be the combined tap vector and $\mathbf{y}_{dfe,k} = [\mathbf{y}_k^T, \mathbf{d}_k^T]^T$ the joint input vector. Similar to the linear equalizer, the objective is to minimize

$$\varepsilon_{mse} = E\left[|z_k - d_{k-\triangle}|^2\right] \tag{2.36}$$

where $z_k = \mathbf{w}_{dfe}^H \mathbf{y}_{dfe,k}$ and the optimum tap weights for minimizing the mean squared error are given by the Wiener-Hopf equations

$$\mathbf{w}_{dfe} = \mathbf{R}_{dfe}^{-1} \mathbf{P}_{dfe,\triangle} \tag{2.37}$$

where $\mathbf{R}_{dfe}$ is the autocovariance matrix of the input vector $\mathbf{y}_{dfe,k}$ and $\mathbf{P}_{dfe,\triangle} = E[d_{k-\triangle}^* \mathbf{y}_{dfe,k}]$. The MMSE may be expressed as

$$\varepsilon_{mmse} = 1 - \mathbf{P}_{dfe,\triangle}^H \mathbf{R}_{dfe}^{-1} \mathbf{P}_{dfe,\triangle} \tag{2.38}$$

When no constraint is placed on filter length, the optimal DFE filters generally have infinite-length [51],[52] corresponding to infinite impulse response (IIR) designs. In [52], the transmitter and receiver of the infinite-length DFE for pulse amplitude modulation (PAM) systems were jointly optimized. The result were extended to quadrature amplitude modulation (QAM) systems in [53]. To reduce complexity, improve stability, or allow adaptability, however, most designs use finite impulse response (FIR) filters in both the feed-forward and feedback sections [54],[55].

The designs of the DFE structure either estimate the CIR and then compute

the FFF/FBF taps [54],[55], or directly adapt the FFF/FBF taps using adaptive algorithms such as the least mean squares (LMS) or recursive least squares (RLS) algorithms [49].

MIMO DFEs have attracted much attention recently. The MIMO finite length minimum mean squared error (MMSE) DFE was developed in [31] and optimized for decision delay $\triangle \geq 0$. The choice of $\triangle > 0$ improves performance for a wide range of channels, as shown in [56]. Here only decisions on temporally preceding symbols are fed back into the detection process of each stream, therefore co-channel interference (CCI) contributions from undetected future and current-time symbols are not cancelled.

In contrast, for the ordered successive interference cancellation in [57], the data streams are successively detected in an ordered manner using a multiple-input single-output (MISO) DFE based BLAST where a MISO DFE is used at each stage. Each stream is detected with the entire CCI contribution from every previously detected stream already cancelled out. In [58], a similar scheme using a MIMO DFE based BLAST is found to offer a performance advantage over the MISO DFE case.

DFE is also known as successive cancellation in multiuser detection [59]. In [60], multiuser detection using a DFE that simultaneously detects all incoming signals is compared to interference rejection using a DFE that detects one signal and rejects the remaining signals as interference. It was found that multiuser detection in general provide better performance than interference rejection, especially when the power levels of users differ substantially.

2.3.3 Maximum Likelihood Sequence Estimation

Assuming perfect knowledge of the channel, an optimum receiver in the presence of ISI and Gaussian noise is a maximum likelihood sequence estimator (MLSE) [33] using a whitened matched filter (WMF) and a Viterbi algorithm (VA) [34]. In linear and decision feedback equalization, the receiver first attempts to suppress the channel impairments and then makes a decision on the transmitted data on

a symbol-by-symbol basis. In MLSE, the receiver attempts to recover the entire transmitted sequence of symbols using the maximum likelihood (ML) detection criterion

$$\hat{\mathbf{d}} = \arg\max_{\tilde{\mathbf{d}}} \quad \mathrm{p}(\mathbf{Y}/\tilde{\mathbf{d}}) \tag{2.39}$$

where $\hat{\mathbf{d}}$ is the detected symbol sequence, $\tilde{\mathbf{d}}$ is the hypothesized sequence and $\mathrm{p}(\mathbf{Y}/\tilde{\mathbf{d}})$ is the pdf of the vector of received samples $\mathbf{y} = [y_1, y_2, \cdots, y_K]$ conditioned on the hypothesized sequence. Based on (2.39), the ML detector selects the hypothesized sequence that maximizes the conditional pdf as the transmitted sequence.

If the elements of the received sample vector $\mathbf{y}$ are conditionally independent of one another, then the joint conditional pdf may be expressed as

$$\mathrm{p}(\mathbf{Y}/\tilde{\mathbf{d}}) = \prod_{k=1}^{K} \mathrm{p}(y_k/\mathcal{Y}_{k-1}) \tag{2.40}$$

where $\mathcal{Y}_k = \{[y_1, y_2, \cdots, y_k, \tilde{\mathbf{d}}\}$. By taking logarithm, the expression is simplified to

$$\ln \mathrm{p}(\mathbf{Y}/\tilde{\mathbf{d}}) = \sum_{k=1}^{K} \ln \mathrm{p}(y_k/\mathcal{Y}_{k-1}). \tag{2.41}$$

For Gaussian distributed received samples the negative log likelihood metric of (2.41) reduces to a Euclidean squared distance metric. Therefore MLSE selects the sequence which minimizes the Euclidean distance between the received samples and the hypothesized sequence.

For MLSE, the estimated channel response is required and this must be provided by channel estimators. There have been several proposed MLSE receivers incorporating channel estimation. The work of [61] proposed a MLSE receiver for a general fading channel using the VA and a bank of Kalman filters to estimate the channel continuously. Reference [62] further developed this explicitly for the time dispersive Rayleigh fading channel. In [63], an innovations-based

MLSE receiver was proposed with the channel estimates supplied by a set of time-invariant linear filters. All these employ the TDL channel model. In [27] and [41], the f-power series channel model was used. In particular, [27] proposed a MLSE receiver using the VA and per-survivor processing (PSP) [64] with the channel estimates provided by Kalman filters using the f-power series model. On the other hand, a t-power series channel model is used by the MLSE receivers in [25] and [26].

The complexity of MLSE is often prohibitive as it grows exponentially with the length of the channel memory. For a given modulation size, M, and channel memory, L, complexity increases as M^L. For a MIMO system, the optimum MLSE receiver using a vector Viterbi algorithm (VVA) was developed by [65], [35]. The complexity further increases exponentially as a product of the number of transmit antennas, T, and the length of the channel memory, L, according to M^{TL}. Reduced-state trellis-based equalizers, for example delayed decision-feedback sequence estimation (DDFSE) [66], reduced-state sequence estimation (RSSE) [67] and the M-algorithm [68] can be used. In general, the benefits of reduced-state equalization can only be realized if the channel impulse response which has to be equalized has a minimum-phase characteristic. This can usually be achieved by using a prefilter, and with infinite filter order and high SNR, it was found that the feedforward filter of a MMSE DFE tends to be the optimum one [69] (more on prefilter in the following section).

For array measurements at the receiver ($R > 1$), an adaptive approach based on PSP is proposed in [70], but when more transmitters are sharing the bandwidth, CCI exists. There are two broad classes of techniques to combat CCI at the receiver. One is to suppress interference, possibly in an adaptive fashion, as in [71]. Another strategy is to decode all T data sequences simultaneously [72], possibly with a blind/adaptive approach [73].

2.3.4 Partitioned Viterbi Algorithm

The partitioned Viterbi algorithm (PVA) of [32] is a suboptimal form of MLSE and incurs approximately a 2dB SNR penalty in achieving the same BER performance compared to using a VVA [35] in a quasi static fading environment. However, unlike the VVA, the complexity of the PVA increases only linearly with the number of transmit antenna according to TM^L. This offers significant complexity reduction in the total number of trellis searches required. We provide a brief description of the PVA algorithm here and the interested reader can refer to [32] for more detail.

A major component of the PVA is a length L_f prefilter used to provide an estimate of the T transmitted signals. Ideally it should be a vector WMF. However, in reality the WMF does not always exist in the MIMO case [69]. Because of this, the feedforward filter of an MMSE DFE is used instead [32], and is shown to approach the WMF as the SNR and number of taps used in the filter tend to infinity [69]. The prefilter compensates pre-cursor ISI and decouples the received signal vector into T outputs. Parallel VA are used to process these outputs. Tentative decisions are made in each interval and these are exchanged among the parallel processors. For each transmitted signal stream, feedback terms estimated using the tentative decisions obtained in the previous interval from other processors are used to cancel the 'cross-interference'.

In order to calculate the prefilter coefficients, channel estimates are needed. A least squares (LS) channel estimator is used in [32] where the channel fading is assumed to be quasi static (constant within a frame but changing randomly from frame to frame[2]). Training symbols at the beginning of a frame are used to obtain an estimate of the CIR which is then used for calculating the perfilter coefficients. The estimated CIR and the received signal vector are then passed through the prefilter, after which the prefiltered CIR matrix, $\tilde{\mathbf{H}}_k = [\tilde{\mathbf{H}}_{k,0}, \tilde{\mathbf{H}}_{k,1}, \cdots, \tilde{\mathbf{H}}_{k,L_f-1}]$, and prefiltered received signal vector, $\tilde{\mathbf{y}}_k$, are used as inputs to the parallel VAs.

[2]The fading from frame to frame is correlated with the fading coefficients generated randomly for the the duration of the frames.

After some decoding delay $\triangle$, an estimate of the transmitted vector $\hat{\mathbf{d}}_{k-\triangle}$ emerges as the PVA output.

2.4 MIMO SYSTEM

The MIMO system considered in the thesis contains the same number of, or more, receive antennas, R, as the number of transmit antennas, T, that is $R \geq T$. This configuration utilizes spatial multiplexing by transmitting multiple independent signal streams from each of the T antennas. This increases the overall system throughput and hence data rate. Optimum receiver structures for such a multiple antenna system were developed in the 1970's by [65], [35].

In the special case when $R = T = 1$, the system reduces to a conventional single-input single-output (SISO) system. For $T = 1$ and $R > 1$, which is known as the single-input multiple-output (SIMO) system, performance gain can be achieved through *receive diversity* techniques, for example as in a RAKE receiver in a CDMA system [36]. Special signal combining and processing techniques, such as maximum ratio combining (MRC) [74], are used at the receiver to detect the transmitted signals. However, using multiple receive antennas in a cellular system, for example on a mobile handset, may not be practical due to space and power constraints. It will be more convenient and also cost-effective to have the multiple antennas at the transmitter, which provides *transmit diversity*. With this technique, coding on the transmit signals is required so that they can be detected effectively at the receiver. The first system using transmit diversity and coding techniques was proposed in [75].

It was then realized that further performance gain and increased system capacity can be achieved with more than one antenna at both the transmit and receive sides. Recent research by [13],[15] has shown that the capacity increases according to the number of transmit antennas as long as there are at least as many receive antennas. This suggests a linear increase in data rate with the number of transmit antenna without any increase in transmission bandwidth or

power. This also improves the BER performance of the system where the slope of the BER curve against the SNR changes according to

$$BER \propto \frac{1}{SNR^d} \tag{2.42}$$

where d is the diversity order of the system.

2.4.1 BLAST

Foschini *et al* proposed the Bell Laboratories Layered Space Time Architecture (BLAST) in [15]. Together with Vertical-BLAST (VBLAST) [16], a simplified version of BLAST, such schemes are designed to maximize the system throughput in terms of bits per second per Hertz (b/s/Hz). Specifically they seek to improve spectral efficiency, by transmitting independent signals from each of multiple transmit antennas. A BLAST scheme typically relies on successive interference cancellation (SIC) at the receiver to detect the signals. In doing so, however, it loses diversity gain due to the interference cancellation process. Moreover the scheme requires at least the same number of receive antennas as transmit antennas.

VBLAST [16] has been shown to achieve the theoretically proven linear capacity increases and has been demonstrated to achieve capacity of 20 - 40 bits/s/Hz in an indoor environment with realistic SNR and error rate [76]. Various research has been pursued with attempt to improve the performance by coding [77],[78],[79], or by different detection architecture [80],[81],[82], or by reducing the receiver complexity [83],[84],[85].

2.4.2 Space Time Trellis Codes

Space time trellis codes (STTC) [18] are a type of space time code (STC) that introduce redundancy in the transmitted signal by using a trellis encoder which is similar to a convolutional encoder. This redundancy provides coding gain which improves the bit error rate performance. The coding gain is dependent on the

construction of the trellis code and the memory length used in the trellis. Being based on trellis codes, a STTC is complex to decode, especially in a frequency selective fading channel. STTC requires multiple transmit antennas but suffice with one receive antenna to operate; nevertheless, usually more than one antenna is employed since it improves performance. The use of multiple transmit and/or receive antennas further provides diversity gain.

There has been much research on developing design rules for STTC and evaluating their performance. Among the literature, [18] first proposed the construction on trellis codes based on determinant and rank criteria, while [86] proposed using the trace criteria which is similar to maximizing the Euclidean distance of the STTC.

2.4.3 Space Time Block Codes

Space time block codes (STBC) [17], [87] act on a block of data at once, in a way similar to block codes. As such it can be viewed as a simple variant of the STTC in a similar way that block codes are related to trellis codes. As with STTC, the operation of STBC necessarily requires multiple transmit antenna but suffice with one receive antenna (although more than one is usually used). Although STBC provides diversity gain, unlike STTC it does not provide coding gain.

A simple orthogonal STBC [17] requires 2 transmit and 1 receive antenna. Data is mapped to a encoding matrix in a block structure for transmission. All columns of the encoding matrix are orthogonal to each other which enables the signal to be easily and linearly separated at the receiver, hence requiring only a very simple decoding scheme. It provides diversity gain and is closely related to MRC. However, it suffers a 3-dB performance penalty under equivalent transmit power constraint where the transmitted power is reduced by P_T/T.

For quasi-orthogonal STBC [88],[89], the block structure is divided into pairs of columns, and each pair is orthogonal to each other. Decoding is done on a pair-wise basis and is more complex. Non-orthogonal STBC [90] are possible but are much more complex to decode.

2.5 SUMMARY

We have presented some background information in this chapter, which include discussions on communication system, fading channel, equalization techniques and some general MIMO systems.

Chapter 3

THE VECTOR GENERALIZED RECURSIVE LEAST SQUARES ALGORITHM

3.1 INTRODUCTION

Channel estimation is an integral part of modern receiver implementations as many classes of equalization techniques require estimates of the channel impulse response (CIR) to operate. These include maximum likelihood sequence estimation (MLSE) [33],[36] where the CIR estimate is needed in the computation of the likelihood metrics. In addition, for equalization techniques that can be implemented adaptively by adjusting filter tap coefficients, such as those employed in decision feedback equalization (DFE), direct computation of the tap coefficients using channel estimates is found to be more robust against time variation of the channel [46] compared to adaptive tap adjustment.

Conventional adaptive algorithms, such as the least mean squares (LMS) [91],[49] and recursive least squares (RLS) algorithms [49] are often employed in channel estimators for slowly time-varying fading environments as they are simple to implement and lead to good estimates of the CIR. Typically the RLS algorithm has a convergence rate an order of a magnitude faster than the LMS algorithm[1] and is not sensitive to variation in the eigenvalue spread of the correlation matrix of the input vector. With a finite training period, in steady state the ensemble averaged squared error of RLS algorithm is also lower than the LMS algorithm. However, this is achieved at a higher level of complexity.

[1] In an environment with sufficiently high signal-to-noise ratio [49].

With its faster convergence rate and lower steady state error, it is tempting to conclude that the RLS algorithm can track a fading environment better than the LMS algorithm. Nevertheless, it has been reported that with properly chosen parameters, the LMS algorithm actually has superior tracking performance [92],[93]. This may come as a surprise. However, according to [94],[95], the RLS algorithm can be interpreted as a special case of a Kalman filter with a state transition matrix equal to a constant multiple of an identity matrix and without a process noise vector. Therefore, the RLS algorithm can be considered as *model dependent* whereas the LMS is model independent. Unless the multiparameter regression model assumed in the derivation of the standard RLS algorithm closely matches the underlying model of the environment in which it operates, there will be a degradation in its tracking performance due to model mismatch. Furthermore, the assumption of a constant state transition matrix is not normally considered as a way to solve the tracking problem in a fading environment [94].

Even though the LMS and RLS algorithms are relatively simple to implement, their tracking ability is rudimentary at best in that in a faster fading environment, their performances tend to degrade [96]. The tracking ability can be enhanced by incorporating information about the channel dynamics, such as the temporal evolution of the CIR and an appropriate mathematical model describing its evolution.

A Kalman filter is one such estimation algorithm that is widely known and used in many applications. It computes the estimates recursively. Prior information about the statistics of the channel is a prerequisite for its design. The Kalman-filter-based estimator (known as the *Kalman estimator* hereafter) assumes that the channel dynamics follow an auto-regressive (AR) model of some appropriate order, where there is a trade-off between the accuracy of the model and the variance of the modelling errors [97]. Assuming that the underlying statistics are Gaussian, a Kalman filter is optimum in that it produces estimates with minimum variance [49].

A Kalman estimator requires a state-space model of the channel response

process with state transition matrix coefficients that depend on an underlying AR model of the channel dynamics. It thus requires knowledge of the second order channel statistics. In addition, knowledge of the signal-to-noise ratio (SNR) is needed to compute the process noise autocovariance. These requirements are critical drawbacks to using a Kalman estimator because the acquisition of channel statistics usually requires a long observation and measurement time [97],[98],[99] which may not be possible in practice. In addition, solving the resulting Yule-Walker equations to obtain the AR coefficients adds a layer of complexity in the estimation process.

For simplicity, in the literature the required statistics are often assumed known. For example, in [100], a Kalman estimator is used to estimate and track the frequency selective channel responses where the AR parameters required by the Kalman filter and the noise statistics are assumed known. On the other hand, additional algorithms for obtaining channel statistics can be used, as for example in [101] where a noise covariance estimation algorithm and a noise whiteness test are developed to estimate the noise covariance needed by the Kalman filter. In [97] an RLS algorithm is used to track the AR parameters.

To overcome these problems, a fixed state transition matrix where the coefficients are derived *a priori* without needing channel statistics is proposed in [25]. It uses the theory of polynomial prediction by assuming a polynomial series expansions of the underlying channel impulse responses using a Taylor's series in the time domain (also known as a t-power series expansion [23]). The coefficients for various predictor lengths and polynomial orders can easily be derived offline. However, the *rectangular windowed RLS* (RW-RLS) estimator in [25] is cumbersome as it involves constant 'downdating' of the observation window size and may experience numerical problems. Furthermore, its overall complexity is about twice that of a Kalman estimator. Note that the fixed state transition matrix can be used in a Kalman filter structure, therefore overcoming the requirement for second order channel statistics. However, the SNR is still required to compute the process noise autocovariance.

In [102] an estimator using a RLS-Kalman algorithm, also known as the *generalized RLS* (GRLS) algorithm, was developed. It incorporates a simplified state-space model of the channel process and the conventional RLS algorithm is a special case of this algorithm. The estimator models each sample of the CIR as a two term t-power series [23]. Instead of estimating the channel response coefficients, the coefficients of the t-power series are estimated. A two term t-power series is suitable only for linearly time-varying channels [25], and the resulting state-space model is limited to channels that vary linearly with time. In [103], the GRLS algorithm is applied to the derivation of variable loop gains of a digital phase-locked loop (DPLL). However, there is no mention of the derivation of the AR parameters, which are assumed to be known.

To avoid the need for channel statistics, an estimator is proposed in [29] that incorporates the non-statistics-based state transition matrix of [25] into the GRLS algorithm of [102]. The algorithm arrives at an approximate channel state model that has *unforced dynamics*, thereby avoiding the need for the process noise autocovariance. Its performance is found to be comparable to that of a Kalman estimator when appropriate predictor lengths and polynomial orders are used [29]. The fixed non-statistics-based state transition matrix enables the estimator to operate without modification in a Rician fading environment. Note that statistics-based methods such as the Kalman estimator require specific reconfiguration of the state transition matrix [97] to take into account a Rician channel model. The resulting estimator has a complexity of $\mathcal{O}\big((\mathcal{N})^3\big)$ real operations per iteration, which is the same as that of the Kalman estimator, $\mathcal{N}$ being the dimension of the channel state vector.

The GRLS algorithm of [29] pertains to a single-input single-output (SISO) environment. In this chapter we develop a vector form of the GRLS algorithm that can be used in a multiple-input multiple-output (MIMO) transmission environment. We call the resulting algorithm a *Vector GRLS* (VGRLS) algorithm. In the following sections we describe its application to channel estimation and evaluate its performance in a MIMO environment.

3.2 THE GENERAL SYSTEM MODEL

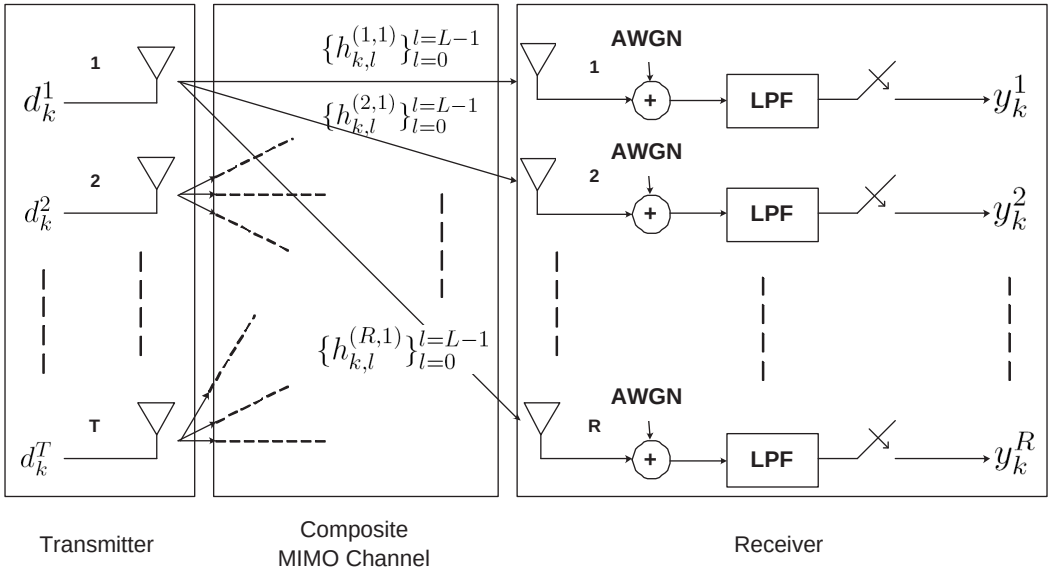

Figure 3.1 A general block diagram of a symbol-spaced discrete-time MIMO communication system at time k for T transmit and R receive antennas.

In order to develop the estimator structure, we assume a MIMO system transmitting independent signals from each of T antennas to $R \geq T$ receive antennas using a VBLAST-type[2] transmission format [16]. Figure 3.1 shows a discrete-time model for the system. At the receiver, each of the R antennas observes a linear combination of the transmitted signals. The symbol-rate sample of the complex baseband received signal at time k may be written at the j-th receive antenna as

$$y_k^{(j)} = \sum_{i=1}^{T} \sum_{l=0}^{L-1} d_{k-l}^{(i)} h_{k,l}^{(j,i)} + n_k^{(j)} \;\; ; \;\; j = 1, 2, \cdots, R \tag{3.1}$$

where $d_k^{(i)}$ is the k-th transmitted complex baseband M-ary data symbol from the i-th antenna, $\{h_{k,l}^{(j,i)}\}_{l=0}^{l=L-1}$ is the sampled fading dispersive composite[3] channel impulse response between the i-th transmit and j-th receive antennas at time k with delay spread of L symbol periods, and $n_k^{(j)}$ is sampled additive white Gaussian noise (AWGN) with variance, σ_n^2.

[2]Throughout the present work no space-time coding is employed.

[3]Assumed to be the convolution of the transmit pulse shape and physical channel response.

With an oversampling factor of $N_r \geq 1$ so that sampling occurs every T_s/N_r seconds T_s being the symbol period, we define vectors of N_r samples in the k-th symbol period as

$$\mathbf{y}_k^{(j)} = \begin{pmatrix} y_{k,0}^{(j)} \\ y_{k,1}^{(j)} \\ \vdots \\ y_{k,N_r-1}^{(j)} \end{pmatrix}, \mathfrak{H}_{k,l}^{(j,i)} = \begin{pmatrix} h_{k,l,0}^{(j,i)} \\ h_{k,l,1}^{(j,i)} \\ \vdots \\ h_{k,l,N_r-1}^{(j,i)} \end{pmatrix}, \mathbf{n}_k^{(j)} = \begin{pmatrix} n_{k,0}^{(j)} \\ n_{k,1}^{(j)} \\ \vdots \\ n_{k,N_r-1}^{(j)} \end{pmatrix}. \tag{3.2}$$

From (3.1), we may then write the oversampled (vector) form of the signal in the k-th symbol interval as

$$\mathbf{y}_k^{(j)} = \sum_{i=1}^{T} \sum_{l=0}^{L-1} d_{k-l}^{(i)} \mathfrak{H}_{k,l}^{(j,i)} + \mathbf{n}_k^{(j)} \ ; \ j = 1, 2, \cdots, R. \tag{3.3}$$

The MIMO received signal of (3.3) may be then expressed in a compact matrix-vector form [32] as

$$\mathbf{y}_k = \sum_{l=0}^{L-1} \mathbf{H}_{k,l} \mathfrak{d}_{k-l} + \mathbf{n}_k \tag{3.4}$$

where

$$\mathbf{y}_k = \begin{pmatrix} \mathbf{y}_k^{(1)} \\ \mathbf{y}_k^{(2)} \\ \vdots \\ \mathbf{y}_k^{(R)} \end{pmatrix}, \mathfrak{d}_k = \begin{pmatrix} d_k^{(1)} \\ d_k^{(2)} \\ \vdots \\ d_k^{(T)} \end{pmatrix}, \mathbf{n}_k = \begin{pmatrix} \mathbf{n}_k^{(1)} \\ \mathbf{n}_k^{(2)} \\ \vdots \\ \mathbf{n}_k^{(R)} \end{pmatrix} \tag{3.5}$$

and where we define the RN_r x T channel matrix-taps

$$\mathbf{H}_{k,l} = \begin{pmatrix} \mathfrak{H}_{k,l}^{(1,1)} & \cdots & \mathfrak{H}_{k,l}^{(1,T)} \\ \vdots & \ddots & \vdots \\ \mathfrak{H}_{k,l}^{(R,1)} & \cdots & \mathfrak{H}_{k,l}^{(R,T)} \end{pmatrix}; \quad l = 0, 1, 2, \cdots, L-1. \tag{3.6}$$

To facilitate the description of the VGRLS estimator, we reformulate (3.4). First, we observe that there are L channel matrix-taps. Next, we represent each as a column vector using the operator $vec(\mathbf{H}_{k,l})$ [100] and stack the columns of $\mathbf{H}_k = [\mathbf{H}_{k,0}, \cdots, \mathbf{H}_{k,L-1}]$ into a length RN_rTL channel vector,

$$\begin{aligned}\mathfrak{h}_k &= vec(\mathbf{H}_k) \\ &= [h_{k,0,0}^{(1,1)} \cdots h_{k,0,N_r-1}^{(1,1)} \cdots h_{k,0,0}^{(R,1)} \cdots h_{k,0,N_r-1}^{(R,1)}, \cdots, \\ &\quad h_{k,0,0}^{(1,T)} \cdots h_{k,0,N_r-1}^{(1,T)} \cdots h_{k,0,0}^{(R,T)} \cdots h_{k,0,N_r-1}^{(R,T)}, \cdots, \\ &\quad h_{k,L-1,0}^{(1,1)} \cdots h_{k,L-1,N_r-1}^{(1,1)} \cdots h_{k,L-1,0}^{(R,1)} \cdots h_{k,L-1,N_r-1}^{(R,1)}, \cdots, \\ &\quad h_{k,L-1,0}^{(1,T)} \cdots h_{k,L-1,N_r-1}^{(1,T)} \cdots h_{k,L-1,0}^{(R,T)} \cdots h_{k,L-1,N_r-1}^{(R,T)}]^t\end{aligned} \tag{3.7}$$

where t denotes matrix transposition. To ensure dimensional compatibility, we also define a RN_r x RN_rTL transmitted data matrix $\mathfrak{D}_k$ as

$$\mathfrak{D}_k = \left[d_k^{(1)} \cdots d_k^{(T)}, d_{k-1}^{(1)} \cdots d_{k-1}^{(T)}, \cdots, d_{k-L+1}^{(1)} \cdots d_{k-L+1}^{(T)}\right] \otimes \mathbf{I}_{RN_r} \tag{3.8}$$

where $\mathbf{I}_{RN_r}$ is the RN_r x RN_r identity matrix and $\otimes$ is the Kronecker product. We may then write (3.4) in the compact form

$$\mathbf{y}_k = \mathfrak{D}_k \mathfrak{h}_k + \mathbf{n}_k. \tag{3.9}$$

3.3 CHANNEL ESTIMATION

Fundamental to the development of the VGRLS estimator is the polynomial series expansion of the underlying channel impulse responses using a Taylor's series in the time domain, i.e. a t-power series expansions [23]. With this model, a polynomial predictor is employed to predict the fading process. It is shown that known polynomial coefficients with different predictor lengths and polynomial or-

ders[4] can be derived *a priori* without requiring any channel statistics. Assuming small modelling error, the one-step polynomial prediction can be interpreted as an approximate channel state equation with *unforced dynamics*. This approximation then enables the prediction process to be done recursively using a form of the RLS-Kalman algorithm, i.e. GRLS, algorithm which is similar to a Kalman filter but without the process noise vector. In the following sections, we will describe these in more detail.

3.3.1 The Polynomial Series Model

Bello proposed in [23] that for a bandlimited fading channel, and given that the time variation of the channel is smooth, the randomly time-variant channel impulse response $h(t,\tau)$ can be approximated over a short interval by a polynomial series. $h(t,\tau)$ is indeed bandlimited in t where the bandwidth is bounded by the maximum Doppler spread, and if the channel fading is slow, then the time variation can be considered as smooth. We describe in the following how the channel taps can be approximated by polynomials.

We assume that $h(t,\tau)$ is a Gaussian process whose sample functions can be differentiated to any order in the mean squared sense [23]. As the fading process is bandlimited [38], it can be expanded as an N-th order polynomial in a small finite interval of interest $|t-\eta T_s| \leq \varphi T_s$ about $t = \eta T_s$ for $0 \leq \varphi \leq 1$ using Taylor's theorem [25],[23],[26] to obtain

$$h(t,\tau) = \sum_{n=0}^{N} a_{\eta}^{(n)}(\tau) b_{\eta}^{(n)}(t) + R_N(t,\tau) \tag{3.10}$$

where the coefficients $a_{\eta}^{(n)}(\tau)$ are given by

$$a_{\eta}^{(n)}(\tau) = \frac{T_s^n}{n!}\left[\frac{d^n h(t,\tau)}{dt^n}\right]_{t=\eta T_s} \tag{3.11}$$

with the elementary basis functions,

[4] We use a polynomial series expansion of order N, which when truncated results in polynomials of degree N.

$$b_\eta^{(n)}(t) = \left(\frac{t - \eta T_s}{T_s}\right)^n \tag{3.12}$$

and $R_N(t,\tau)$ is the remainder term, or residual (modelling) error, of the Taylor series, given by

$$R_N(t,\tau) = \frac{(t-\eta T_s)^{N+1}}{(N+1)!}\frac{d^{N+1}}{dt^{N+1}}h(t,\tau). \tag{3.13}$$

Provided that the sampling rate within the interval of interest is adequate in that a sufficient number of CIR samples is available, the mean squared value of $R_N(t,\tau) \to 0$ as the polynomial order $N \to \infty$, and the expansion (3.10) approaches

$$h(t,\tau) = \sum_n^\infty a_\eta^{(n)}(\tau) b_\eta^{(n)}(t). \tag{3.14}$$

Therefore with increasing polynomial order N, the polynomial approximation becomes more accurate. However, as N increases, the number of unknown parameters increases too, and for a fixed number of observations within the φ interval, these unknowns cannot be determined accurately if they are too numerous.

Given that the Doppler spread is normally much smaller than the sampling rate, only the first few terms in the expansions of (3.14) are significant. Any higher order terms can be ignored as long as the expression is used only in a small vicinity of ηT_s.

Using the Wiener-Khintchine theorem together with the fact that $\frac{d^{N+1}}{dt^{N+1}}(\bullet)$ is equivalent to multiplication by $(j2\pi f)^{N+1}$ in the frequency domain, and assuming finite support of the Doppler power spectral density over $|f| \leq f_D$, where f_D is the maximum Doppler frequency, we may write the mean squared value of the remainder $R_N(t,\tau)$,

$$\frac{1}{2}E\{R_N(t,\tau)R_N^*(t,\tau)\} = \frac{(t-\eta T_s)^{2(N+1)}}{(N+1)!^2}\int_{-f_D}^{f_D} |2\pi f|^{2(N+1)} S_{hh}(f)df \tag{3.15}$$

Table 3.1 Effect of Polynomial Orders on Φ_N with $f_D T_s = 0.002$

Order N	***Upper bound*** Φ_N
0	1.9739 x 10^{-5}
1	2.4352 x 10^{-11}
2	8.9018 x 10^{-18}

where $S_{hh}(f)$ is the Doppler power spectral density of the channel impulse response $h(t,\tau)$.

If we consider the polynomial series of (3.10) in an interval $(\eta-\varphi)T_s \leq t \leq (\eta+\varphi)T_s$, then an upper bound Φ_N, on the mean squared value of the remainder term is

$$\frac{1}{2}E\{R_N(t,\tau)R_N^*(t,\tau)\} \leq \frac{(2\pi\varphi T_s)^{2(N+1)}}{(N+1)!^2}\int_{-f_D}^{f_D}|f|^{2(N+1)}S_{hh}(f)df = \Phi_N. \quad (3.16)$$

Assuming that the fading process evolves according to Clarke's fading model [42] with a U-shaped Doppler spectrum [43] given by

$$S_{hh}(f) = \begin{cases} \frac{1}{\pi f_D\sqrt{1-\frac{f^2}{f_D^2}}} & |f| \leq f_D \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

the upper bound is approximately given by

$$\Phi_N \approx \frac{2(N+1)!(\pi\varphi f_D T_s)^{2(N+1)}}{N_r^{2(N+1)}\left((N+1)!\right)^4} \quad (3.18)$$

where N_r is the number of samples per symbol. As we can see, for a fixed N_r and φ, the approximate upper bound depends on the polynomial order N and the normalized fade rate, $f_D T_s$. We calculate in Table 3.1 the approximate values of the bound for various polynomial orders for $f_D T_s = 0.002$, $N_r = 1$ and $\varphi = 1$. As Φ_N is inversely proportional to $\frac{1}{((N+1)!)^3}$, for $f_D T_s \ll 1$, its value diminishes very quickly and hence a small polynomial order N is sufficient to represent the fading process.

3.3.2 Minimum Noise Gain Polynomial Predictor

With the channel impulse response modelled as a polynomial series in (3.10), it can be shown that *a priori* known one-step ahead polynomial predictor tap coefficients can be derived and used to predict the fading process [25],[26]. The first attempt to predict the process using a *polynomial predictor* appeared in [104]. The predictor used is actually equivalent to a least squares predictor where the optimization can also be achieved by using the Lagrange multiplier technique. The polynomial predictor used in [25],[26] can also be derived using Prony's method [105],[106]. Here we describe the derivation of the polynomial predictor tap coefficients.

Consider a general polynomial series, Ψ_m, of order N,

$$\Psi_m = \sum_{i=0}^{N} d_i m^i; \quad m = \cdots, -1, 0, 1, \cdots, \tag{3.19}$$

and we want to predict one-step ahead the terms in the polynomial series with a P-tap predictor, for $P \geq N+1$, where the one-step prediction can be expressed as

$$\sum_{i=0}^{N} d_i (m+1)^i = \sum_{p=1}^{P} a_p \sum_{i=0}^{N} d_i (m-p+1)^i = \sum_{i=0}^{N} d_i \sum_{p=1}^{P} a_p (m-p+1)^i. \tag{3.20}$$

For arbitrary polynomial coefficients $\{d_i\}$, we have

$$(m+1)^i = \sum_{p=1}^{P} a_p (m-p+1)^i. \tag{3.21}$$

Let $\mathbf{a} = [a_1, a_2, \cdots, a_P]$ be the predictor tap coefficient vector of the P-tap predictor which can be calculated in closed form with the following set of linear constraints on the coefficients [107][5]:

[5]Note that the constraints follow directly from (3.21)

$$g_0 = \sum_{p=1}^{P} a_p - 1 = 0 \tag{3.22}$$

$$g_i = \sum_{p=1}^{P} p^i a_p = 0; \quad i = 1, \cdots, N. \tag{3.23}$$

Observing (3.21) at $m = 0$, without loss of generality, we have

$$(1)^i = \sum_{p=1}^{P} a_p(-p+1)^i, \tag{3.24}$$

which after rearranging gives

$$\sum_{p=1}^{P} a_p(p-1)^i = (-1)^i; \quad i = 0, 1, \cdots, N. \tag{3.25}$$

Equation (3.25) gives a set of $N + 1$ equations which can be written compactly as

$$\mathbf{P}\mathbf{a}^t = \mathbf{b} \tag{3.26}$$

where $\mathbf{P}$ is the $(N + 1)$ x P matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & P-1 \\ 0 & 1^2 & 2^2 & \cdots & (P-1)^2 \\ & & \vdots & & \\ 0 & 1 & 2^N & \cdots & (P-1)^N \end{pmatrix} \tag{3.27}$$

and $\mathbf{b} = [1, -1, 1, \cdots, (-1)^N]^t$.

Now (3.26) is a system of $N + 1$ linear equations in P unknowns. Since $P \geq N + 1$, there may exist an infinite number of solutions. Different design methodologies for polynomial predictors can be found in [108],[109]. Here we consider a minimum noise gain[6] design [107] where the noise gain is defined by

[6]In the presence of AWGN, it is critical that the noise gain is minimized.

Table 3.2 Polynomial Predictor Tap Coefficients of Various Order and Length

Length P	***Order N***	***Polynomial Coefficients*** $\{a_1, a_2, \cdots, a_P\}$
2	0	{1/2, 1/2}
2	1	{2, -1}
3	0	{1/3, 1/3, 1/3}
3	1	{4/3, 1/3, -2/3}
3	2	{3, -3, 1}
4	0	{1/4, 1/4, 1/4, 1/4}
4	1	{1, -1/2, 0, 1/2}
4	2	{9/4, -3/4, -5/4, 3/4}
4	3	{4, -6, 4, -1}

$\mathbf{a}\mathbf{a}^t$. Hence, we want the solution of (3.26) with as small value of $\mathbf{a}\mathbf{a}^t$ as possible. This optimization can be achieved using a $(N+1)$ x 1 Lagrange multiplier vector $\mathbf{z}$, where

$$\mathfrak{L} = \mathbf{a}\mathbf{a}^t + \mathbf{z}^t(\mathbf{P}\mathbf{a}^t - \mathbf{b}). \tag{3.28}$$

Differentiating (3.28) with respect to $\mathbf{a}$ and equating the result to zero gives

$$\mathbf{a}^t = -\frac{1}{2}\mathbf{P}^t\mathbf{z}. \tag{3.29}$$

From (3.26) and (3.29), the Lagrange multiplier vector is given by $\mathbf{z} = -2(\mathbf{P}\mathbf{P}^t)^{-1}\mathbf{b}$. Substituting this into (3.29), we get

$$\mathbf{a}^t = \mathbf{P}^t(\mathbf{P}\mathbf{P}^t)^{-1}\mathbf{b}. \tag{3.30}$$

Polynomial predictor tap coefficients with different values of predictor length, P, and polynomial order, N, are calculated and given in Table 3.2. Note that with the constraint in (3.22), the sum of the coefficients for each value of P and N is 1.

The polynomial predictor will suppress or amplify noise depending on whether the norm[7] of the polynomial coefficient vector, $\sum_{p=1}^{P} |a_p^N|^2$, is less than or greater

[7] I.e. $\mathbf{a}\mathbf{a}^t$

than one. In the presence of AWGN with variance σ_n^2, it can be shown [49] that the variance of the prediction due to AWGN can be approximately given by

$$\sigma_{AWGN}^2 \approx \left(\sum_{p=1}^{P} |a_p^N|^2 \right) \sigma_n^2 \tag{3.31}$$

and that due to the approximate upper bound Φ_N on the mean squared value of the remainder term is approximately

$$\sigma_{res}^2 \approx \left(\sum_{p=1}^{P} |a_p^N|^2 \right) \Phi_N. \tag{3.32}$$

These affect the performance of the estimator as will be shown later.

3.3.3 Statistical State-space Model

Channel estimators based on the Kalman filter [100],[97] assume that the RN_rTL x 1 multipath fading channel response vector of (3.7) evolves according to an order P_a vector autoregressive (VAR) process [110]. With this assumption, a RN_rTLP_a x 1 channel state vector at time k can be written as

$$\mathbf{h}_k = \left[\mathfrak{h}_k^t, \mathfrak{h}_{k-1}^t, \cdots, \mathfrak{h}_{k-P_a+1}^t \right]^t \tag{3.33}$$

and its transition from time k to $k+1$ can be described by a state equation of the form,

$$\mathbf{h}_{k+1} = \mathbf{A}\mathbf{h}_k + \mathbf{v}_k \tag{3.34}$$

where $\mathbf{v}_k$ is a zero-mean process noise vector of dimension RN_rTLP_a x 1 such that

$$E\{\mathbf{v}_k \mathbf{v}_l^H\} = \begin{cases} \mathbf{R}_v & \text{for } k = l, \\ \mathbf{0}_{m,m} & \text{for } k \neq l. \end{cases} \tag{3.35}$$

with $\mathbf{0}_{m,m}$ being the (m x m) null matrix and $m = RN_rTLP_a$. The superscript H denotes Hermitian transposition and $\mathbf{A}$ is the RN_rTLP_a x RN_rTLP_a state transition matrix having the form

$$\mathbf{A} = \begin{pmatrix} \mathbb{A}_1 & \mathbb{A}_2 & \cdots & \mathbb{A}_{P_a-1} & \mathbb{A}_{P_a} \\ & \mathbf{I}_{RN_rTL(P_a-1)} & & & \mathbf{0}_{RN_rTL(P_a-1),RTL} \end{pmatrix} \tag{3.36}$$

where the matrices $\{\mathbb{A}_l = \phi_l \mathbf{I}_{RN_rTL}\}$, $l = 1, 2, \cdots, P_a$, are the RN_rTL x RN_rTL matrix coefficients of the VAR process. The AR coefficients $\{\phi_l\}$ and the process noise autocovariance matrix, $\mathbf{R}_v$, may be obtained by measuring the channel statistics and solving the resulting matrix-vector Yule-Walker equations [111], [112], [110], the derivation of which is included in Appendix A. The choice of the process order P_a is a trade-off between complexity and modelling accuracy [97]. When a high degree of accuracy is needed, a large P_a is selected such that the variances of the elements of $\mathbf{v}_k$ are small [29].

By defining a RN_r x RN_rTLP_a data matrix, $\mathbf{d}_k$, as

$$\mathbf{d}_k = \begin{bmatrix} \mathfrak{D}_k & | & \mathbf{0}_{RN_r,RN_rTL(P_a-1)} \end{bmatrix} \tag{3.37}$$

with $\mathfrak{D}_k$ given by (3.8), we may express the MIMO received signal of (3.9) as

$$\mathbf{y}_k = \mathbf{d}_k \mathbf{h}_k + \mathbf{n}_k \tag{3.38}$$

which provides an observation equation.

The state-space model used by the Kalman estimator [49],[100],[97] is then given by the state equation of (3.34) and the observation equation, (3.38). As the model is structured, it is restricted to Rayleigh fading channels. However it may be explicitly modified to model specular components [97] by reformulating the state transition matrix $\mathbf{A}$.

3.3.4 Polynomial-based State-space Model

Note that the state equation of (3.34) can be interpreted as a one-step length-P_a vector-matrix predictor of the channel state vector with the VAR matrix coefficients $\{\mathbb{A}_l\}$ for $l = 1, 2, \cdots, P_a$, being one-step prediction coefficients and $\mathbf{v}_k$ the associated prediction error.

Since the fading process varies smoothly, as shown previously we may model the time evolution of each of the samples as polynomial sequences of order N [29]. From the theory of polynomial prediction [104], a one-step predictor of length P with coefficients $\{a_p\}$ for $p = 1, 2, \cdots, P$, may be derived for each polynomial sequence. Following [104],[25],[113], for the μ-th scalar channel component of (3.7) for $\mu = 1, 2, \cdots, RN_rTL$, we may then write a one-step N-th order polynomial prediction equation at time k as

$$h_{k,\mu} = \sum_{p=1}^{P} a_p h_{k-p,\mu} + e_{k,\mu}(N, P) \tag{3.39}$$

where P is the length of the polynomial predictor, assuming that each channel response is modeled as a truncated t-power series [23] of order N and that the series converges over a window of size $P + 1$ [26] observations.

As shown in section 3.3.2, the polynomial predictor coefficients $\{a_p\}$ for $p = 1, 2, \cdots, P$, are dependent only on the values of N and P, and may be computed offline using a Lagrange multiplier technique [25], or a standard least square optimization approach [29]. Moreover the computation does not require any channel statistics. Polynomial coefficients for various orders, N, and lengths, P, are given in Table 3.2.

The prediction error arising from truncation of the series to the first N terms, $e_{k,\mu}(N, P)$, is dependent on the order of the polynomial series and the predictor length, where $e_{k,\mu}(N, P) \to 0$ as $N \to \infty$ [29]. It will be small if the window of expansion (i.e., the predictor length, P) is small, thereby allowing the use of a small value of N. Using (3.39) a VAR-like model of the channel vector of (3.7) may be written as

$$\mathfrak{h}_k = \sum_{p=1}^{P} \mathfrak{U}_p \mathfrak{h}_{k-p} + \mathbf{e}_k(N,P) \tag{3.40}$$

where the RN_rTL x RN_rTL polynomial predictor matrices are given by $\mathfrak{U}_p = a_p \mathbf{I}_{RN_rTL}$ for $p = 1, 2, \cdots, P$. The model is only VAR-like because the error vector

$$\begin{aligned} \mathbf{e}_k(N,P) = [&e_{k,0}^{(1,1)} \cdots e_{k,0}^{(R,1)} \cdots e_{k,0}^{(1,T)} \cdots e_{k,0}^{(R,T)} \cdots \\ &e_{k,L-1}^{(1,1)} \cdots e_{k,L-1}^{(R,1)} \cdots e_{k,L-1}^{(1,T)} \cdots e_{k,L-1}^{(R,T)}]^t \end{aligned} \tag{3.41}$$

is not necessarily zero-mean or white [29] as required by an actual VAR process. In general, the elements of the covariance of $\mathbf{e}_k(N,P)$ will be small over a suitably small window of expansion around each sampling instant [29]. As a result, if $\mathbf{e}_k(N,P)$ is assumed to be approximately zero, which holds if the polynomial model of (3.10) is used only in the vicinity of ηT, a state-space model similar in form to (3.34), but with unforced dynamics is obtained from (3.40) as

$$\mathbf{h}_{k+1} = \mathbf{U}\mathbf{h}_k \tag{3.42}$$

where

$$\mathbf{h}_k = \left[\mathfrak{h}_k^t, \mathfrak{h}_{k-1}^t, \cdots, \mathfrak{h}_{k-P+1}^t\right]^t \tag{3.43}$$

is the RN_rTLP x 1 channel state vector at time k and the associated state transition matrix is given by

$$\mathbf{U} = \begin{pmatrix} \mathfrak{U}_1 & \mathfrak{U}_2 & \cdots & \mathfrak{U}_{P-1} & \mathfrak{U}_P \\ & \mathbf{I}_{RN_rTL(P-1)} & & & \mathbf{0}_{RN_rTL(P-1),RN_rTL} \end{pmatrix}. \tag{3.44}$$

This is similar in form to (3.36), but with P_a replaced by P and the matrices $\mathbb{A}_l$ replaced by the matrices $\mathfrak{U}_P$. The observation equation associated with (3.42) is similar to (3.38), except that P_a in the data matrix of (3.37) becomes P.

Equations (3.38) and (3.42) define a polynomial-based state-space model with unforced dynamics. It does not require channel statistics in the derivation of the state transition matrix coefficients, and can be used with both Rayleigh and Rician fading channels with no explicit reconfiguration of the state transition matrix $\mathbf{U}$.

3.3.5 Derivation of the VGRLS Algorithm

With reference to [102],[26], we shall develop the algorithm from the first principles using a general time-indexed[8] state transition matrix, $\mathbf{U}_{k/k-1}$. The algorithm is recursive in the sense that, on the receipt of a current signal sample, it repeats a sequence of operations on a set of parameter values determined after the receipt of the previous signal sample. It is least squares in the sense that it minimizes the weighted squared error in the estimate of the signal sample.

In our work, the MIMO channel estimator operates with a RN_rTLP x 1 component vector, $\hat{\mathbf{h}}_k$, which is the k-th estimate of the channel vector $\mathbf{h}_k$ in equation (3.43), and is expressed as

$$\hat{\mathbf{h}}_k = \left[\hat{\mathfrak{h}}_k^t, \hat{\mathfrak{h}}_{k-1}^t, \cdots, \hat{\mathfrak{h}}_{k-P+1}^t\right]^t, \tag{3.45}$$

and the data vector $\mathbf{d}_k$, of size RN_r x RN_rTLP in equation (3.37) is also repeated here for convenience

$$\mathbf{d}_k = \left[\quad \mathfrak{D}_k \quad | \qquad \mathbf{0}_{RN_r, RN_rTL(P_a-1)} \quad \right]. \tag{3.46}$$

From (3.45) and (3.46), an estimate of the measurement vector (the received signal vector) formed by the estimator is

$$\hat{\mathbf{y}}_k = \mathbf{d}_k \hat{\mathbf{h}}_k. \tag{3.47}$$

In relation to the actual received vector $\mathbf{y}_k$, the error in the estimate is

[8]This is to show explicitly the temporal progression of the algorithm.

$$\bar{\mathbf{e}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k. \tag{3.48}$$

The vector $\hat{\mathbf{h}}_k$ determined by the channel estimator is such as to minimize the time-average weighted cost function

$$\begin{aligned} J(k) &= \sum_{l=0}^{k} \lambda^{k-l} |\bar{\mathbf{e}}_l|^2 \\ &= \sum_{l=0}^{k} \lambda^{k-l} \left(\mathbf{y}_l - \mathbf{d}_l \hat{\mathbf{h}}_l \right)^2. \end{aligned} \tag{3.49}$$

where λ is a real-valued constant known as the weighting or 'forget factor' in the range of $0 < \lambda < 1$. Assuming that the estimator starts operation on receipt of $\mathbf{y}_0$, the quantity $J(k)$ is the weighted squared error in $\{\hat{\mathbf{y}}_l\}$, starting with $\hat{\mathbf{y}}_0$, for $l = \{0, 1, \cdots, k\}$.

We define a general time-indexed state transition matrix, $\mathbf{U}_{k/k-1}$, which is the RN_rTLP x RN_rTLP state transition matrix relating the channel state vector at time k-1 to the one at time k and is similar in form to (3.44). With this, an estimate $\hat{\mathbf{h}}_k$ of the channel state vector can be expressed as

$$\hat{\mathbf{h}}_k = \mathbf{U}_{k/k-1} \hat{\mathbf{h}}_{k-1}. \tag{3.50}$$

As we want to determine the channel estimation vector $\hat{\mathbf{h}}_k$, which together with the state transition matrix, minimizes $J(k)$, then from (3.49), the cost function may be expressed as

$$J(k) = \sum_{l=0}^{k} \lambda^{k-l} \left(\mathbf{y}_l - \mathbf{d}_l \hat{\mathbf{h}}_{l/l} \right)^H \left(\mathbf{y}_l - \mathbf{d}_l \hat{\mathbf{h}}_{l/l} \right) \tag{3.51}$$

where $\hat{\mathbf{h}}_{k/k}$ is the k-th estimate of the state vector $\mathbf{h}_k$ given all information up to and including information at time k and H denotes Hermitian transposition. Minimizing $J(k)$ in fact minimizes the weighted sum of all squared errors of each

of the elements of the error vector $\left(\mathbf{y}_l - \mathbf{d}_l\hat{\mathbf{h}}_{l/l}\right)$, up to and including the k-th error.

We can express $\{\hat{\mathbf{h}}_{l/l}\}$ in (3.51) in term of $\hat{\mathbf{h}}_{k/k}$ and we note that $\hat{\mathbf{h}}_{k/k}$ can be obtained recursively from all its previous estimates $\{\hat{\mathbf{h}}_{l/l}\}$. From this, we have

$$\begin{aligned}
\hat{\mathbf{h}}_{k/k} &= \mathbf{U}_{k/k-1}\hat{\mathbf{h}}_{k-1/k-1} \\
&= \mathbf{U}_{k/k-1}\left(\mathbf{U}_{k-1/k-2}\hat{\mathbf{h}}_{k-2/k-2}\right) \\
&= \mathbf{U}_{k/k-1}\left(\mathbf{U}_{k-1/k-2}\left[\mathbf{U}_{k-2/k-3}\hat{\mathbf{h}}_{k-3/k-3}\right]\right) \\
&= \left(\mathbf{U}_{k/k-1}\mathbf{U}_{k-1/k-2}\cdots\mathbf{U}_{l+1/l}\right)\hat{\mathbf{h}}_{l/l} \\
&= \mathbf{U}_{k/l}\hat{\mathbf{h}}_{l/l}
\end{aligned} \tag{3.52}$$

where the last line of (3.52) is obtained by using the properties of the state transition matrix [49],

$$\mathbf{U}_{k/k-1}\mathbf{U}_{k-1/k-2}\cdots\mathbf{U}_{l+1/l} = \mathbf{U}_{k/l}. \tag{3.53}$$

Further recognizing [49] that

$$\mathbf{U}_{k-1/k} = \mathbf{U}_{k/k-1}^{-1} \tag{3.54}$$

where $\mathbf{U}_{k-1/k}$ is the backward state transition matrix from time k to k-1, we may rearrange the terms in equation (3.52) to get

$$\begin{aligned}
\hat{\mathbf{h}}_{l/l} &= \mathbf{U}_{k/l}^{-1}\hat{\mathbf{h}}_{k/k} \\
&= \mathbf{U}_{l/k}\hat{\mathbf{h}}_{k/k}.
\end{aligned} \tag{3.55}$$

We may then rewrite (3.51) in terms of the estimate $\hat{\mathbf{h}}_{k/k}$ as

$$J(k) = \sum_{l=0}^{k}\lambda^{k-l}\left(\mathbf{y}_l - \mathbf{d}_l\mathbf{U}_{l/k}\hat{\mathbf{h}}_{k/k}\right)^H\left(\mathbf{y}_l - \mathbf{d}_l\mathbf{U}_{l/k}\hat{\mathbf{h}}_{k/k}\right). \tag{3.56}$$

Note that (3.56) reduces to the cost function used in the derivation of the scalar observation GRLS algorithm [102],[26] for a communication system having one transmit and one receive antenna ($R = T = 1$).

The cost function (3.56) may be expanded as

$$J(k) = \sum_{l=0}^{k} \lambda^{k-l} \left(\mathbf{y}_l^H \mathbf{y}_l - \mathbf{y}_l^H \mathbf{d}_l \mathbf{U}_{l/k} \hat{\mathbf{h}}_{k/k} - \hat{\mathbf{h}}_{k/k}^H \mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{y}_l + \hat{\mathbf{h}}_{k/k}^H \mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{d}_l \mathbf{U}_{l/k} \hat{\mathbf{h}}_{k/k} \right), \tag{3.57}$$

and is a quadratic function in $\hat{\mathbf{h}}_{k/k}$. We assume that $J(k)$ (which is real and positive) is a unimodal function in the space spanned by $\hat{\mathbf{h}}_{k/k}$ with a global minimum at a particular value of $J(k)$. To minimize $J(k)$, its gradient is evaluated and equated to zero. The definition of the complex gradient operation for vectors and matrices is found in [49].

From (3.57) the gradient of $J(k)$ with respect to $\hat{\mathbf{h}}_{k/k}$ is

$$\nabla J(k) = \sum_{l=0}^{k} \lambda^{k-l} \left(-2\mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{y}_l + 2\mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{d}_l \mathbf{U}_{l/k} \hat{\mathbf{h}}_{k/k} \right). \tag{3.58}$$

Equating the gradient in equation (3.58) to zero and rearranging the terms yields

$$\left(\sum_{l=0}^{k} \lambda^{k-l} \mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{d}_l \mathbf{U}_{l/k} \right) \hat{\mathbf{h}}_{k/k} = \sum_{l=0}^{k} \lambda^{k-l} \mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{y}_l. \tag{3.59}$$

Equation (3.59) is in fact a form of the weighted time average normal equations. In Wiener filter theory [49], the tap coefficients giving the minimum mean squared error are obtained by solving the normal equations. In this case, not the statistical mean but the actual difference between the desired response vector and the estimate is minimized.

Let

$$\mathbf{R}_{k/k} = \sum_{l=0}^{k} \lambda^{k-l} \mathbf{U}_{l/k}^H \mathbf{d}_l^H \mathbf{d}_l \mathbf{U}_{l/k} \tag{3.60}$$

and

$$\mathbf{Q}_{k/k} = \sum_{l=0}^{k} \lambda^{k-l} \mathbf{U}_{l/k}^{H} \mathbf{d}_{l}^{H} \mathbf{y}_{l}. \tag{3.61}$$

The normal equations of (3.59) then become

$$\mathbf{R}_{k/k} \hat{\mathbf{h}}_{k/k} = \mathbf{Q}_{k/k}. \tag{3.62}$$

The vector $\hat{\mathbf{h}}_{k/k}$ is the required least squares estimates of the corresponding channel vector. However, to evaluate $\hat{\mathbf{h}}_{k/k}$ directly from the RN_rTLP equations of (3.62) would be inefficient and impractical for each new received signal vector. To avoid this, $\hat{\mathbf{h}}_{k/k}$ is instead determined recursively.

First, it is necessary to determine the corresponding recursive formulations for $\mathbf{R}_{k/k}$ and $\mathbf{Q}_{k/k}$. These can be obtained by using (3.53) and the fact that $\mathbf{U}_{k/k} = \mathbf{I}_{RN_rTLP}$. Isolating the term in (3.60) at time $l = k$ (which gives $\mathbf{d}_{k}^{H}\mathbf{d}_{k}$), and recognizing that $\mathbf{R}_{k-1/k-1} = \sum_{l=0}^{k-1} \lambda^{(k-1)-l} \mathbf{U}_{l/k-1}^{H} \mathbf{d}_{l}^{H} \mathbf{d}_{l} \mathbf{U}_{l/k-1}$, the recursive formulation of (3.60) is then given by

$$\mathbf{R}_{k/k} = \lambda \mathbf{U}_{k-1/k}^{H} \mathbf{R}_{k-1/k-1} \mathbf{U}_{k-1/k} + \mathbf{d}_{k}^{H} \mathbf{d}_{k} \tag{3.63}$$

and for (3.61)

$$\mathbf{Q}_{k/k} = \lambda \mathbf{U}_{k-1/k}^{H} \mathbf{Q}_{k-1/k-1} + \mathbf{d}_{k}^{H} \mathbf{y}_{k}. \tag{3.64}$$

Using (3.62) to replace $\mathbf{Q}_{k/k}$ by $\mathbf{R}_{k/k}\hat{\mathbf{h}}_{k/k}$ and $\mathbf{Q}_{k-1/k-1}$ by $\mathbf{R}_{k-1/k-1}\hat{\mathbf{h}}_{k-1/k-1}$ in equation (3.64), we get

$$\mathbf{R}_{k/k} \hat{\mathbf{h}}_{k/k} = \lambda \mathbf{U}_{k-1/k}^{H} \mathbf{R}_{k-1/k-1} \hat{\mathbf{h}}_{k-1/k-1} + \mathbf{d}_{k}^{H} \mathbf{y}_{k}. \tag{3.65}$$

Equation (3.65) gives a relationship between $\hat{\mathbf{h}}_{k/k}$ and $\hat{\mathbf{h}}_{k-1/k-1}$ which forms the basis of the required recursive algorithm to determine $\hat{\mathbf{h}}_{k/k}$. However, the equation involves the two matrices $\mathbf{R}_{k/k}$ and $\mathbf{R}_{k-1/k-1}$, and it is desirable to replace $\mathbf{R}_{k-1/k-1}$ to simplify the computation.

Using (3.54), the relationship in (3.63) is rewritten as

$$\lambda \mathbf{R}_{k-1/k-1} = \left(\mathbf{U}_{k/k-1}^H \mathbf{R}_{k/k} \mathbf{U}_{k/k-1}\right) - \left(\mathbf{U}_{k/k-1}^H \mathbf{d}_k^H \mathbf{d}_k \mathbf{U}_{k/k-1}\right). \tag{3.66}$$

Substituting (3.66) into equation (3.65), we obtain

$$\begin{aligned} \mathbf{R}_{k/k}\hat{\mathbf{h}}_{k/k} &= \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\mathbf{R}_{k/k}\mathbf{U}_{k/k-1}^H\mathbf{U}_{k-1/k}^H\right) \\ &\quad - \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\mathbf{d}_k^H\mathbf{d}_k\mathbf{U}_{k/k-1}^H\mathbf{U}_{k-1/k}^H\right) + \mathbf{d}_k^H\mathbf{y}_k \\ &= \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\mathbf{R}_{k/k}\right) - \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\mathbf{d}_k^H\mathbf{d}_k\right) + \mathbf{d}_k^H\mathbf{y}_k. \end{aligned} \tag{3.67}$$

We then rearrange the terms so that

$$\begin{aligned} \hat{\mathbf{h}}_{k/k} &= \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\right) - \left(\hat{\mathbf{h}}_{k-1/k-1}\mathbf{U}_{k/k-1}\mathbf{d}_k^H\mathbf{d}_k\mathbf{R}_{k/k}^{-1}\right) + \mathbf{d}_k^H\mathbf{y}_k\mathbf{R}_{k/k}^{-1} \\ &= \hat{\mathbf{h}}_{k/k-1} + \mathbf{P}_{k/k}\mathbf{d}_k^H\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right) \end{aligned} \tag{3.68}$$

where

$$\hat{\mathbf{h}}_{k/k-1} = \mathbf{U}_{k/k-1}\hat{\mathbf{h}}_{k-1/k-1} \tag{3.69}$$

and

$$\mathbf{P}_{k/k} = \mathbf{R}_{k/k}^{-1}. \tag{3.70}$$

Equation (3.68) is known as the recursive estimate update equation. The new estimate $\hat{\mathbf{h}}_{k/k}$ on the l.h.s. of (3.68) is updated by the weighted error vector, which is the second term on the r.h.s. of (3.68). Analogous equations are also found in other algorithms such as those for Kalman and LMS estimators.

Upon observation, however, equation (3.68) is still not fully recursive since it is dependent on the parameter $\mathbf{P}_{k/k}$ without showing how this can be evaluated from $\mathbf{P}_{k-1/k-1}$. The next step is therefore to develop a relationship between $\mathbf{P}_{k/k}$ and $\mathbf{P}_{k-1/k-1}$. From equation (3.63), by substituting $\mathbf{R}_{k/k} = \mathbf{P}_{k/k}^{-1}$ and

$\mathbf{R}_{k-1/k-1} = \mathbf{P}^{-1}_{k-1/k-1}$ we obtain

$$\begin{aligned}\mathbf{P}^{-1}_{k/k} &= \lambda \mathbf{U}^{H}_{k-1/k}\mathbf{P}^{-1}_{k-1/k-1}\mathbf{U}_{k-1/k} + \mathbf{d}^{H}_{k}\mathbf{d}_{k} \\ &= \lambda \left(\mathbf{U}_{k/k-1}\mathbf{P}_{k-1/k-1}\mathbf{U}^{H}_{k/k-1}\right)^{-1} + \mathbf{d}^{H}_{k}\mathbf{d}_{k}.\end{aligned} \tag{3.71}$$

It can be shown that the bracket term in equation (3.71) equals $\mathbf{P}_{k/k-1}$,

$$\begin{aligned}\mathbf{U}_{k/k-1}\mathbf{P}_{k-1/k-1}\mathbf{U}^{H}_{k/k-1} &= \mathbf{P}_{k/k}\mathbf{U}^{H}_{k/k-1} \\ &= \mathbf{U}_{k-1/k}\mathbf{P}_{k/k} \\ &= \mathbf{P}_{k/k-1}.\end{aligned} \tag{3.72}$$

Therefore, we obtain the relationship

$$\begin{aligned}\mathbf{P}^{-1}_{k/k-1} &= \lambda \left(\mathbf{U}_{k/k-1}\mathbf{P}_{k-1/k-1}\mathbf{U}^{H}_{k/k-1}\right)^{-1} \\ &= \lambda \mathbf{U}^{H}_{k-1/k}\mathbf{P}^{-1}_{k-1/k-1}\mathbf{U}_{k-1/k}\end{aligned} \tag{3.73}$$

and we observe that

$$\mathbf{P}_{k/k-1} = \lambda^{-1}\mathbf{U}_{k/k-1}\mathbf{P}_{k-1/k-1}\mathbf{U}^{H}_{k/k-1}. \tag{3.74}$$

Substituting equation (3.73) into (3.71) the expression becomes

$$\mathbf{P}^{-1}_{k/k} = \mathbf{P}^{-1}_{k/k-1} + \mathbf{d}^{H}_{k}\mathbf{d}_{k}. \tag{3.75}$$

To evaluate the inverse of $\mathbf{P}_{k/k}$, we invoke the *matrix inversion lemma*[9]. By defining $\mathcal{A} = \mathbf{P}_{k/k}$, $\mathcal{B}^{-1} = \mathbf{P}^{-1}_{k/k-1}$, $\mathcal{C} = \mathbf{d}^{H}_{k}$, $\mathcal{D}^{-1} = \mathbf{I}_{R}$, $\mathcal{C}^{H} = \mathbf{d}_{k}$, the recursion (3.75) is then expressed as

[9]Given $\mathcal{A} = \mathcal{B}^{-1} + \mathcal{C}\mathcal{D}^{-1}\mathcal{C}^{H}$, the inverse of $\mathcal{A}$ is given by $\mathcal{A}^{-1} = \mathcal{B} - \mathcal{B}\mathcal{C}(\mathcal{D} + \mathcal{C}^{H}\mathcal{B}\mathcal{C})^{-1}\mathcal{C}^{H}\mathcal{B}$. For more details see [49].

$$\mathbf{P}_{k/k} = \mathbf{P}_{k/k-1} - \mathbf{P}_{k/k-1}\mathbf{d}_k^H \left(\mathbf{I}_R + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H\right)^{-1} \mathbf{d}_k\mathbf{P}_{k/k-1}. \tag{3.76}$$

Now we define the Kalman gain matrix as

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{d}_k^H \left(\mathbf{I}_{RN_r} + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H\right)^{-1} \tag{3.77}$$

so that (3.76) becomes

$$\mathbf{P}_{k/k} = \left(\mathbf{I}_{RN_rTLP} - \mathbf{K}_k\mathbf{d}_k\right)\mathbf{P}_{k/k-1}. \tag{3.78}$$

Equations (3.74), (3.77) and (3.78) give a recursive relationship between $\mathbf{P}_{k/k}$ and $\mathbf{P}_{k-1/k-1}$, which can be used with equation (3.68) to give a recursive relationship between $\hat{\mathbf{h}}_{k/k}$ and $\hat{\mathbf{h}}_{k-1/k-1}$.

However, a further simplification of the algorithm can be achieved. We post-multiply (3.78) by $\mathbf{d}_k^H$ and obtain

$$\mathbf{P}_{k/k}\mathbf{d}_k^H = \mathbf{P}_{k/k-1}\mathbf{d}_k^H - \mathbf{K}_k\mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H. \tag{3.79}$$

Substituting (3.77) into (3.79) yields

$$\begin{aligned} \mathbf{P}_{k/k}\mathbf{d}_k^H &= \mathbf{K}_k\left(\mathbf{I}_M + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H\right) - \mathbf{K}_k\mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H \\ &= \mathbf{K}_k \end{aligned} \tag{3.80}$$

which is another expression for the Kalman gain vector.

Now equation (3.68) is given by

$$\hat{\mathbf{h}}_{k/k} = \hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right). \tag{3.81}$$

and the derivation of the VGRLS algorithm is completed.

Now we employ the polynomial predictor based model of (3.42) in the VGRLS algorithm to directly estimate the channel tap or state vector, $\mathbf{h}_k$. The coefficients of the state transition matrix $\mathbf{U}$ (3.44) are pre-determined (cf. Table 3.2) for a given predictor length, P, and polynomial order, N. Given that the state transition matrix is fixed, we drop the time index in $\mathbf{U}_{k/k-1}$ and use $\mathbf{U}$ in the VGRLS algorithm from here onwards.

In summary, assuming that $\hat{\mathbf{h}}_{k/k-1}$ and $\mathbf{P}_{k/k-1}$ are known, the update equations for the algorithm may be expressed as

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{d}_k^H \left(\mathbf{I}_{RN_r} + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H\right)^{-1} \tag{3.82}$$

$$\mathbf{P}_{k/k} = \left(\mathbf{I}_{RN_rTLP} - \mathbf{K}_k\mathbf{d}_k\right)\mathbf{P}_{k/k-1} \tag{3.83}$$

$$\hat{\mathbf{h}}_{k/k} = \hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right). \tag{3.84}$$

The prediction equations may then be written as

$$\hat{\mathbf{h}}_{k+1/k} = \mathbf{U}\hat{\mathbf{h}}_{k/k} \tag{3.85}$$

$$\mathbf{P}_{k+1/k} = \lambda^{-1}\mathbf{U}\mathbf{P}_{k/k}\mathbf{U}^H \tag{3.86}$$

where $\hat{\mathbf{h}}_{k/k-1}$ is the estimate of the channel state vector at time k based on all (k-1) prior received samples, λ is the RLS 'forget factor', $\mathbf{K}_k$ is analogous to the Kalman gain vector [49] and $\mathbf{P}_{k/k}$ is the so-called 'intermediate' matrix. In a conventional RLS algorithm, $\mathbf{P}_{k/k}$ is the inverse input autocorrelation matrix but that is not the case here, hence the term 'intermediate' matrix.

To initialize the algorithm, we set the estimated channel state vector $\hat{\mathbf{h}}_{1/0}$ to the null vector and let $\mathbf{P}_{1/0} = \delta^{-1}\mathbf{I}_{RN_rTLP}$, where δ is a small positive real constant. Also note that when $P = 1$ and $N = 0$, the VGRLS algorithm reduces

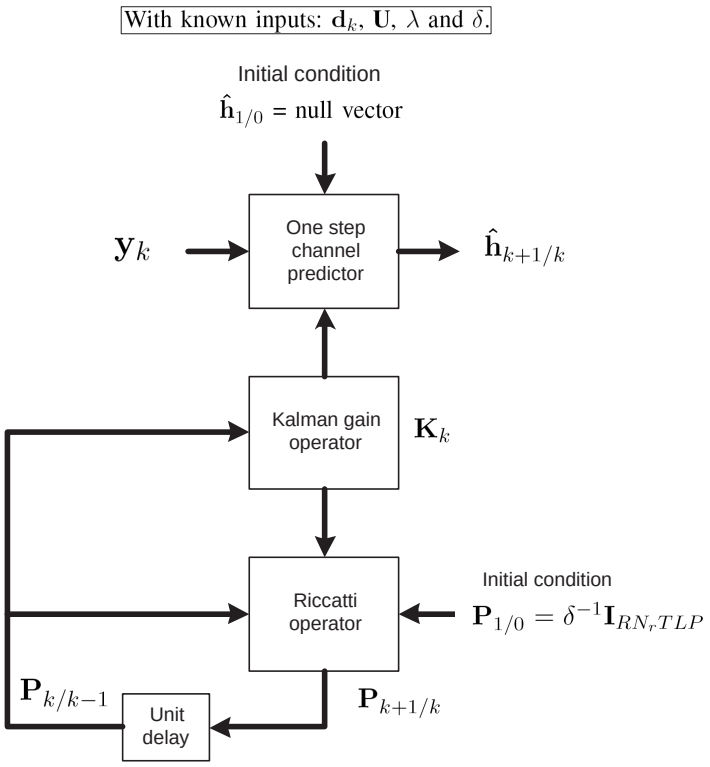

Figure 3.2 Block diagram of the VGRLS algorithm.

to a conventional vector RLS estimation algorithm [29].

The input to the algorithm consists of the vector received samples, $\mathbf{y}_k$ and the resulting output equals $\hat{\mathbf{h}}_{k+1/k}$, the one-step predicted channel vector. Besides these, the algorithm also requires the input of $\mathbf{U}$, $\mathbf{d}_k$, λ and δ, all assumed known quantities. When the VGRLS estimator is operated in isolation, $\mathbf{d}_k$ is the vector of known training symbols. When it is operated in tandem with an equalizer in decision-directed mode, $\mathbf{d}_k$ is the output vector of the equalizer.

We note from (3.82), (3.83) and (3.86) that the Kalman gain $\mathbf{K}_k$ and the recursive update of $\mathbf{P}_{k/k}$ are independent of the received vector $\mathbf{y}_k$. Consequently these quantities may be computed before the VGRLS algorithm is put into operation and this provides a basis for further complexity reduction as will be shown in the next chapter.

The VGRLS algorithm can be represented by the block diagram in Fig. 3.2 which is based on the three components:

1. Kalman gain operator which produces $\mathbf{K}_k$

2. Riccatti operator which produces $\mathbf{P}_{k+1/k}$

3. One-step channel predictor which produces $\hat{\mathbf{h}}_{k+1/k}$

The details of these three components are shown in Figs. 3.3, 3.4 and 3.5 respectively.

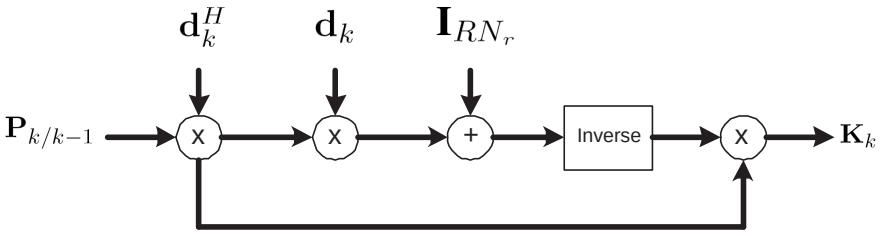

Figure 3.3 Signal flow diagram of the Kalman gain operator.

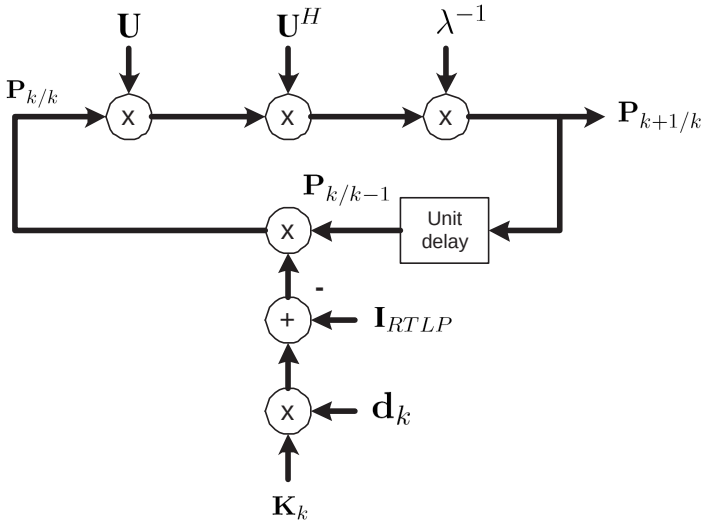

Figure 3.4 Signal flow diagram of the Riccatti operator.

The VGRLS estimator of (3.82) - (3.86) is similar in structure to a Kalman estimator as both consist of time-update and prediction equations. Due to the Riccatti recursion in (3.83) and (3.86), the complexity of the estimator in the highest term is $\mathcal{O}\big((RN_rTLP)^3\big)$, which is similar to that of the Kalman filter. Therefore the 'baseline' complexity of the two algorithms is similar. However,

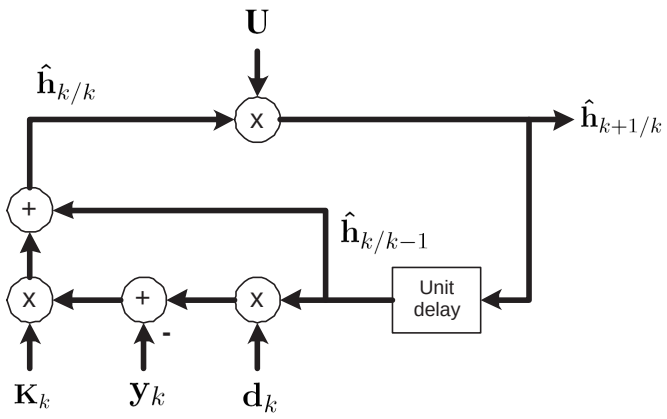

Figure 3.5 Signal flow diagram of the one-step channel predictor.

the VGRLS does not require channel statistics to compute the coefficients of the state transition matrix.

3.4 PARAMETERS THAT AFFECT THE PERFORMANCE OF THE ESTIMATOR

The performance of the VGRLS estimator is evaluated during steady state to which it settles after an initial transient period. The average squared norm difference, or error, between the original and the estimated responses, known as the 'mean squared deviation' (MSD), depends on several parameters. Some of these are inherent to the estimator itself, such as the

- predictor length, P,
- polynomial order, N,
- 'forget factor', λ,

while the rest are system parameters such as the

- signal to noise ratio, SNR,
- normalized channel fade rate, $f_D T$,

- number of the training symbols, L_t,

We will examine in the following how these parameters influence the performance of the estimator.

3.4.1 Effect of the Predictor Length and Polynomial Order

In deriving the estimator, we have assumed that the channel fading process may be expanded as an N-th order Taylor's series over a small window. Since the Doppler spread is much less than the sampling frequency, only the first few terms (i.e., N terms) of the expansion are significant. However, ignoring the remainder terms causes the model to have a residual modelling error. An approximate upper bound on its mean squared value is given as Φ_N in (3.18) which is reproduced here for convenience as

$$\Phi_N \approx \frac{2(N+1)!(\pi\varphi f_D T_s)^{2(N+1)}}{N_r^{2(N+1)}\left((N+1)!\right)^4}. \tag{3.87}$$

It was shown in [25] that the variance of the estimation error, which is termed the effective noise, of the polynomial predictor consists of two parts, namely $\sigma_{eff}^2 = \sigma_{AWGN}^2 + \sigma_{res}^2$, where σ_{AWGN}^2 is the effective variance component due to AWGN given in (3.31) as

$$\sigma_{AWGN}^2 \approx \left(\sum_{p=1}^{P} |a_p^N|^2\right) \sigma_n^2 \tag{3.88}$$

and σ_{res}^2 in (3.32) is the effective variance component due to Φ_N,

$$\sigma_{res}^2 \approx \left(\sum_{p=1}^{P} |a_p^N|^2\right) \Phi_N. \tag{3.89}$$

For a fixed predictor length P, the larger the polynomial order N, the better the fit to the actual fading process and hence the smaller the achievable Φ_N, and this is desirable. However, a higher polynomial order results in larger values of the predictor tap coefficients (cf. Table 3.2), thereby increasing the squared norm

Table 3.3 Norm of the Predictor Tap Vector $\sum_{p=1}^{P} |a_p^N|^2$ (refer Table 3.2)

	Order N			
Length P	**0**	**1**	**2**	**3**
2	0.5	5	-	-
3	0.33	2.33	19	-
4	0.25	1.5	7.75	69

of the coefficient vector $\sum_{p=1}^{P} |a_p^N|^2$, as shown in Table 3.3. This will result in a larger effective AWGN variance σ^2_{AWGN} especially at low SNR. However, as most wireless systems operate with relatively high SNR, the effect due to this variance becomes very small. Note that even though a larger norm will increase σ^2_{res}, this is compensated by a better model fit, resulting in a small residual modelling error, hence a negligibly small Φ_N (cf. Table 3.1).

For a fixed polynomial order N, it is possible to reduce the effective AWGN variance σ^2_{AWGN} by using more predictor taps, i.e., a longer predictor length, P. However, this directly increases the complexity of the estimator, which increases as $(RN_rTLP)^3$. From our simulation results in the next section, it is found that using a predictor length of $P = 3$ with a polynomial order of $N = 2$ is sufficient for most channel conditions.

3.4.2 Effect of the 'Forget Factor'

As with a conventional RLS algorithm, with a 'forget factor' of $\lambda < 1$, the estimates may become 'noisy' [49]. A smaller value of λ results in a noisier adaptive process and at low SNR the 'mean squared deviation' behavior tends to get worse due to this 'adaptation noise'. The value of λ also affects the effective memory of the algorithm according to $\frac{1}{1-\lambda}$ [49] where the memory is effectively shortened with a smaller λ. This means the algorithm uses a smaller number of significant previous samples in the adaptive process. In a fast fading environment where tracking becomes more challenging and the resulting estimates get noisier, this is beneficial as a smaller number of the noisy samples are used in the subsequent recursive updates.

3.4.3 Effect of SNR

The performance of the estimator at low SNR is influenced by AWGN more than by the residual modelling error. From Table 3.3 and equation (3.88), we note that, depending on the choice of P and N, the effective variance due to AWGN σ^2_{AWGN} can be suppressed or amplified. From the theory of least squares estimation [105], the effect of noise averaging increases as the ratio of P to $N+1$ increases. Hence the lower the polynomial order N in relation to the predictor length P, the less the noise enhancement. For $N = 0$, the noise is actually suppressed. Therefore for a fixed predictor length P at low SNR, a lower polynomial order N will perform better than a higher order one due to a smaller effective AWGN variance.

As SNR increases, the effect of AWGN decreases and after a certain SNR called the 'transition SNR', the residual modelling error (3.89) becomes dominant. A higher order model will then perform better than a lower order one due to a better model fit.

For a fixed polynomial order N, as the SNR increases, there will be a floor in the MSD behavior because as the AWGN decreases and the residual error become dominant, there is essentially no improvement in the MSD performance with further increase in SNR, except through using a higher polynomial order. In general, for a given predictor length P at low SNR, a lower polynomial order N performs better than a higher order but at high SNR the reverse is true.

3.4.4 Effect of the Normalized Fade Rate

The normalized fade rate, $f_D T_s$, affects the mean squared value of the residual error, Φ_N, in (3.16) through the Doppler spectrum. For a Jakes' fading model, an approximate upper bound on Φ_N is given in (3.87). We see that as $f_D T_s$ increases, so does Φ_N. Hence we see that the MSD behavior of the estimator will degrade as $f_D T_s$ increases due to a larger residual error. According to Table 3.3, the effective variance σ^2_{res} of (3.89) due to this increased residual error may be reduced by increasing the predictor length P (while keeping N fixed), as this

gives a smaller vector norm of the associated polynomial coefficients. However, this will increase the complexity of the estimator as $(RN_rTLP)^3$. At high fade rate, the effect of fading at low SNR is significant compared to the AWGN. Hence the MSD even when using a lower order estimator will be high when compared with the MSD at a low fade rate.

3.4.5 Effect of the Training Sequence Length

As mentioned at the beginning of this section, the MSD is a measure of the steady state behavior of the estimator, assuming that it has settled following an initial transient period. For a given training sequence length of L_t, if this assumption is not valid, a variance component due to this 'transient noise' will be added to the overall MSD, resulting in worse MSD behavior. It is expected that in general as L_t increases the MSD behavior will improve as the estimator would have converged closer to its steady state. We will show this effect in the next section.

3.4.6 Estimator Variances

Summarizing from the above, the MSD behavior of the estimator is influenced by several types of errors, or variances, associated with different values of the parameters, as follows:

1. Inherent to the modelling of the fading process using finite Taylor's series expansion, there is a *residual modelling error* due to the truncated number of terms in the series. This is determined by the polynomial order N as the residual error decreases with increasing N (i.e. increasing number of terms retained).

2. For a fixed predictor length P, the norm of the polynomial coefficients increases as the polynomial order N increases. It is shown that this has a direct effect on the effective variances due to AWGN and residual error.

3. A variance component due to 'adaptation noise', which is dependant on SNR and the 'forget factor', λ, also affects the MSD.

4. A fast fading channel where the channel condition changes significantly over the interval of polynomial expansion is more difficult to estimate and track. Thus, an error is introduced depending on the normalized channel fade rate, $f_D T_s$.

5. As with any other communication system, the received signal at the input of the estimator is corrupted by AWGN. Hence, the estimate will be noisy and this is directly related to the system SNR.

6. A variance component due to transient noise will be introduced if an insufficient number of training symbols is used so that the estimator has not converged to steady state when the MSD is calculated.

These effects will be illustrated in the following section dealing with the performance evaluation of the estimator.

3.5 PERFORMANCE EVALUATION

We evaluate the performance of the VGRLS estimator in terms of the 'mean square deviation' (MSD), which is the time-averaged squared norm difference, or error, between the actual and estimated channel impulse responses. The estimator is operated using knowledge of the transmitted training symbols and it constantly updates the estimated channel responses. It is assumed to operate in a transient mode during the L_t training symbols, after which it is assumed to operate in steady state mode. The MSD measures this steady state performance, and therefore the first L_t symbols of each frame are not included in the MSD calculation. At the beginning of each new frame the estimator re-initializes, and starts channel acquisition again. The MSD performance versus SNR of the estimator with a fixed predictor length of $P = 3$ and 4, and polynomial orders, $N =$ 0, 1, 2, 3, is evaluated. The steady state MSD in the α-th frame is estimated as

$$\sigma^2_{MSD}(\alpha) = \langle ||\mathfrak{h}_k - \hat{\mathfrak{h}}_{k/k-1}||^2 \rangle \tag{3.90}$$

where $\mathfrak{h}_k$ is the channel vector at time k, $\hat{\mathfrak{h}}_{k/k-1}$ is the one-step ahead estimated channel vector and $\langle ||.|| \rangle$ denotes the time average of the Euclidean norm operator. The MSD for each sub-channel is accumulated and averaged for 10,000 frames. The overall MSD is then averaged across the RT sub-channels of the MIMO system.

The SNR is defined per received antenna as

$$SNR = \left(\frac{\sigma^2_{\mathfrak{d}} \sigma^2_c(\tau) \int_{-\infty}^{\infty} |p(t)|^2 dt}{\sigma^2_n} \right) \tag{3.91}$$

where $\sigma^2_{\mathfrak{d}}$ is the total average energy per data symbol, $\sigma^2_c(\tau) = \sum_{k=1}^{3} \sigma^2_c(\tau_k)$ is the average subchannel power with 3 multipath rays each, $p(t)$ is the transmit pulse shape and σ^2_n is the AWGN variance at the input of each receiver. We use QPSK modulation with unit power and normalize the overall effective channel tap power to unit power. Unless stated otherwise, we assume that the total transmitted power is unity and allocated equally among the T antennas.

3.5.1 Predictor Length, Polynomial Order, SNR and Training Sequence Length

Fig. 3.6 shows the MSD behaviors of a VGRLS estimator and a Kalman filter [10] at a normalized fade rate of $f_D T_s = 0.002$, where f_D is the maximum Doppler frequency. We observe how the polynomial order, N, system SNR and training sequence length, L_t, affect the MSD. At low SNR where noise dominates, an estimator with order $N = 0$ has a slightly better MSD than the others, since the algorithm then acts primarily as a noise averaging filter which tends to suppresses the noise [29] (cf. Table 3.3). This is also attributed to a smaller norm of the polynomial coefficients for $N = 0$ because a larger norm amplifies the noise

[10] We assume a VAR order $P_a = 3$ for the Kalman filter, the same as the predictor length, P, in the simulations.

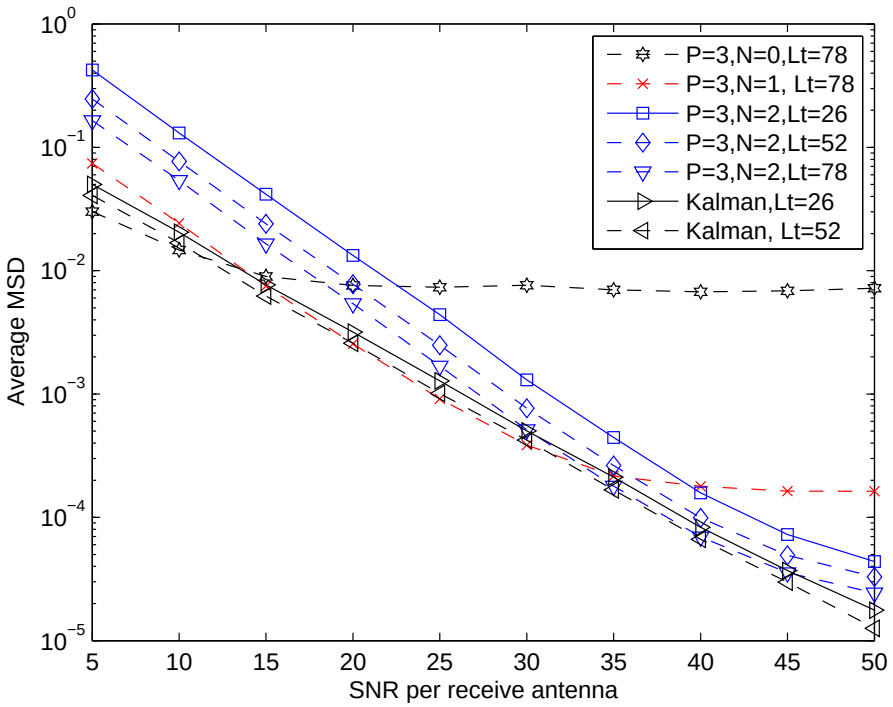

Figure 3.6 MSD of the VGRLS estimator and that of a Kalman filter for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002. VGRLS with $N = 0$ corresponds to a conventional vector RLS algorithm. With sufficient training sequence length, the MSD of VGRLS with $N = 2$ approaches that of a Kalman filter's.

according to (3.88). At moderate SNR, an order of $N = 1$ not only performs linear interpolation but also noise averaging, and has the lowest MSD. At high SNR where the effect of modelling error dominates, an order of $N = 2$ has the lowest MSD. These behaviors are consistent with those reported in [29].

It is interesting to note that the estimator MSD exhibits a floor at sufficiently high SNR regardless of the value of N. This was initially thought to be because of non-convergence of the estimator due to an insufficient number of training symbols. However, when the number of training symbols is increased[11] to $L_t = 78$, the floor for $N = 2$ is still visible at very high SNR although the effect is slight within the observed SNR range. This is unlike the scalar case of [29] where the use of $L_t = 52$ effectively removes the error floor for $N = 2$ at high SNR. As will be shown later, this effect is attributable to error introduced by the channel fading which cannot be reduced by increasing the SNR or the polynomial order,

[11]Data length L_d is still 116 symbols.

N. We also note that increasing the length of the training sequence, e.g. using L_t = 130, improves the steady state MSD performance of the estimator across the SNR region for all estimator orders.

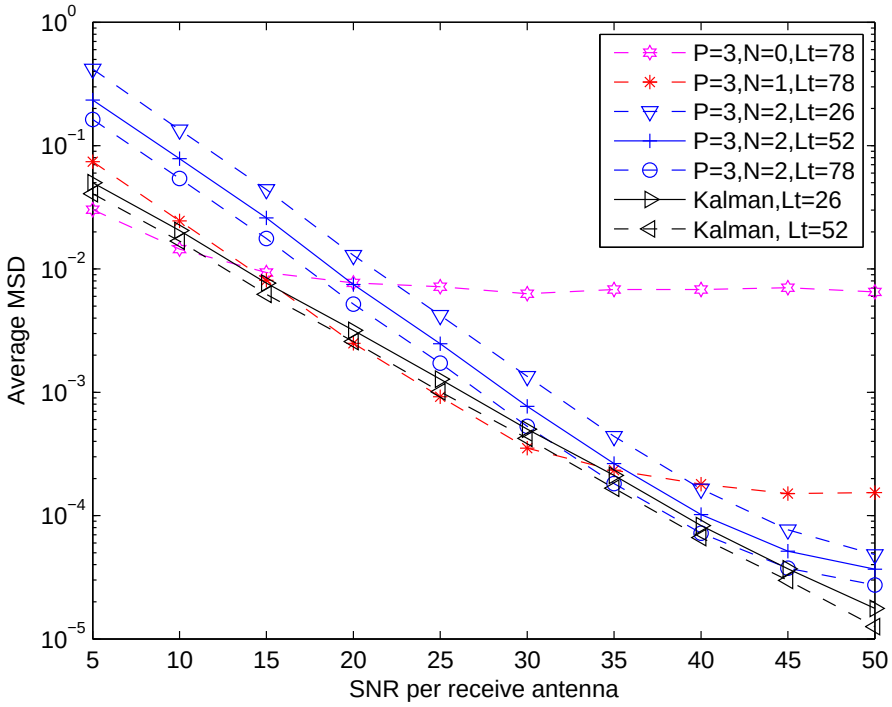

Figure 3.7 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002. The VRGLS has a predictor length of P = 3 and various N with various training symbol length as shown. The power delay profile is SUI-4 [114] which is non-uniform with a power profile of (0dB, -4dB and -8dB).

We have assumed a uniform power delay profile in our simulations for simplicity. However, in reality the power delay profile may not be uniform. We have investigated the MSD performance of the VGRLS estimator at a normalized fade rate of 0.002 for a (2,2) Rayleigh fading, with a non-uniform power delay profile modeled according to the SUI-4 channel model [114]. This is a 3 ray model with a power profile of (0dB, -4dB and -8dB). The result is shown in Fig. 3.7. Compared to Fig. 5.12 for a uniform power delay profile, we note that there is negligible difference in performance. A uniform profile is considered as one of the more severe profiles as all the multipath rays have equal power. It is also used as a test profile for the purpose of GSM's equalizer testing [115]. In the following, a uniform delay profile is used unless otherwise stated.

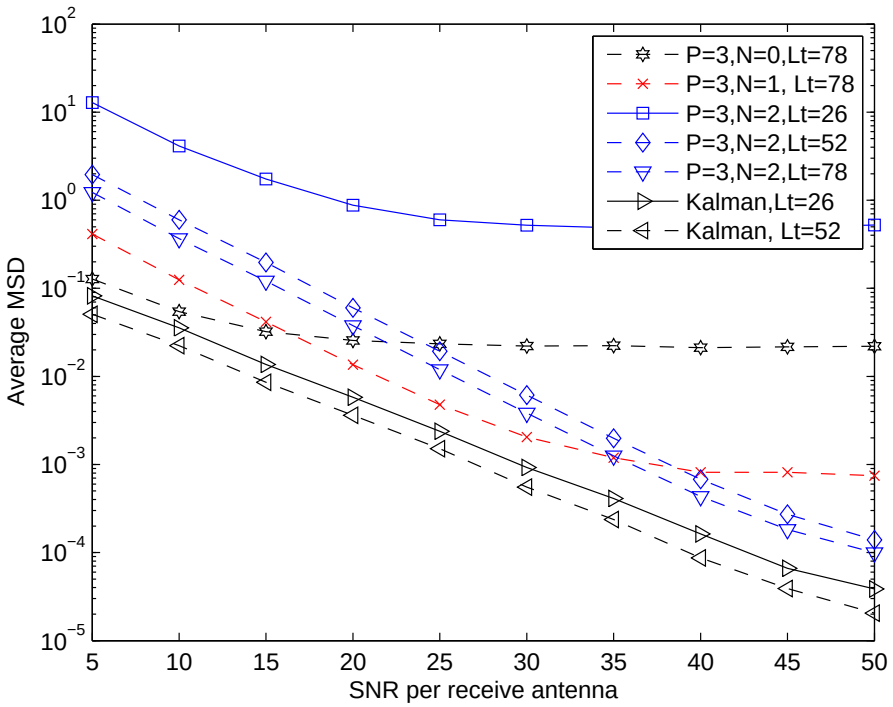

Figure 3.8 MSD of the VGRLS estimator and a Kalman filter for a (4,4) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002.

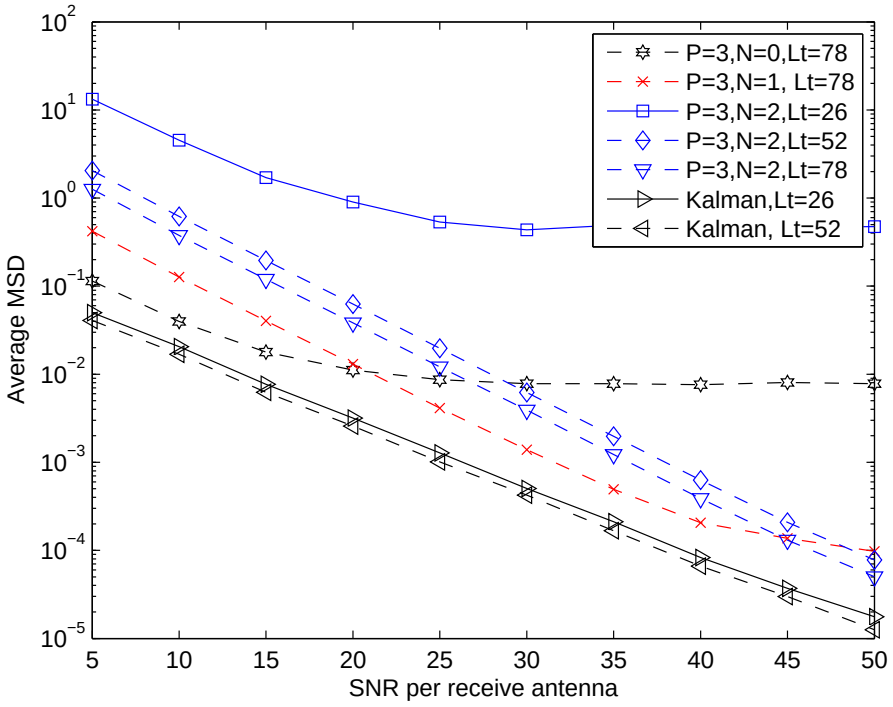

Figure 3.9 MSD of the VGRLS estimator for a (4,4) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.0001.

For simplicity[12], most of the evaluations in this thesis have been limited to a (2,2) MIMO system. We have, however, also included some key results for a (4,4) system. Figs. 3.8 and 3.9 shows the MSD performance of the estimator for a (4,4) MIMO system at normalized fade rates of 0.002 and 0.0001 respectively. In general the MSD performance is worse than that of a (2,2) system although it is improved with a longer training sequence length, L_t. We note the very high irreducible MSD floor for $N = 2$ with $L_t = 26$. This appears to be due to the failure of the estimator to converge within 26 symbol periods.

3.5.2 'Forget Factor'

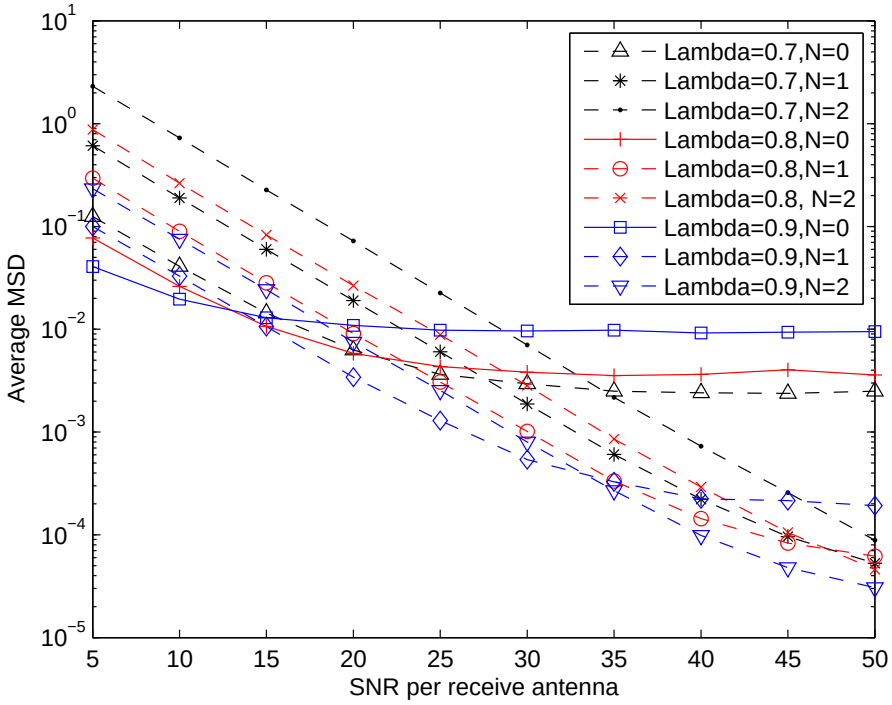

Figure 3.10 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with different values of lambda at $f_D T_s = 0.002$.

Fig. 3.10 shows the effect of different values of 'forget factor' λ on the performance of the estimator. At low SNR, an estimator with a smaller value of λ has worse MSD behavior because of the noisier adaptive process associated with a smaller λ. As SNR increases, this noise effect becomes less dominant and the

[12]The complexity of VGRLS increases as $(RN_rTLP)^3$

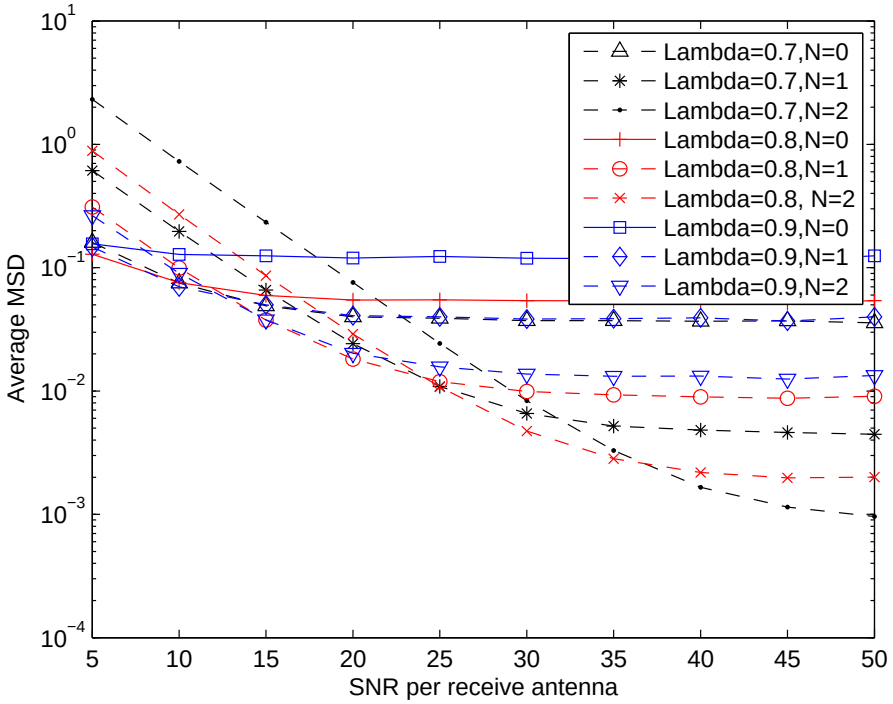

Figure 3.11 MSD of the VGRLS estimator with different values of lambda at $f_D T_s = 0.01$.

estimates start to have better MSD behavior. It is also because the estimator with a smaller memory effectively utilizes a smaller number of previous CIR samples in updating the estimates, and hence the updated estimates are less noisy. This is more evident in fast fading when tracking becomes more challenging and therefore the estimates get noisier. As shown in Fig. 3.11, for $P = 3$ and $N = 2$ at $f_D T_s = 0.01$, the MSD can be reduced significantly at high SNR by reducing λ from 0.9 to 0.7. However this reduction is achieved at the expense of a significantly increased MSD at low SNR due to the noisier adaptive process.

3.5.3 Normalized Fade Rate

Fig. 3.12 shows the MSD of the VGRLS at a slower normalized fade rate of 0.0001 and Fig. 3.13 shows the MSD at a faster normalized fade rate of 0.01. The results show a similar trend in MSD performance, i.e., at low SNR there is not much difference between the various orders but at high SNR $N = 2$ offers significantly lower MSD. The results show that the VGRLS estimator is able to

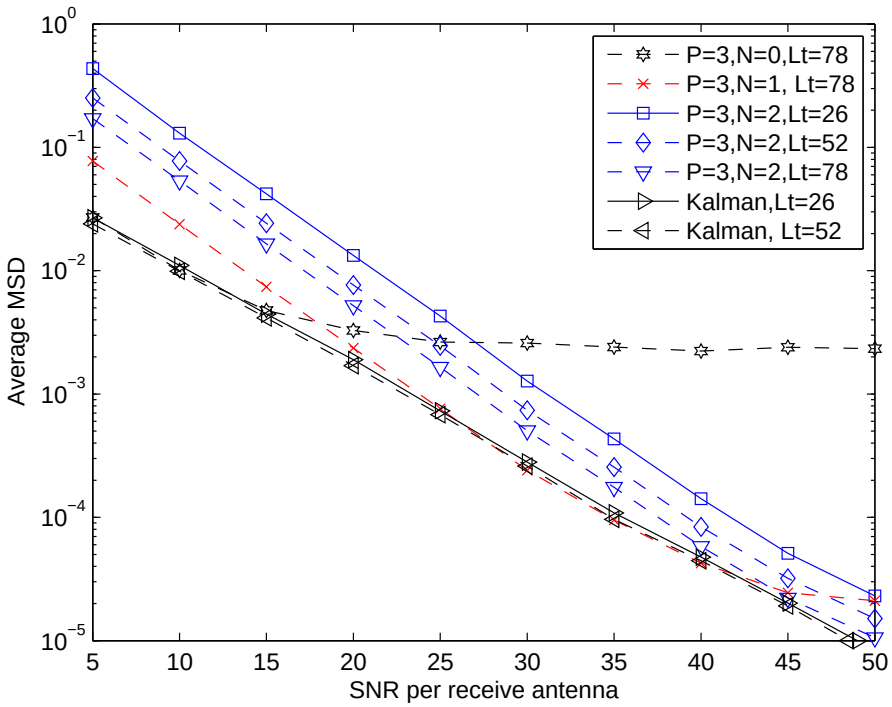

Figure 3.12 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.0001.

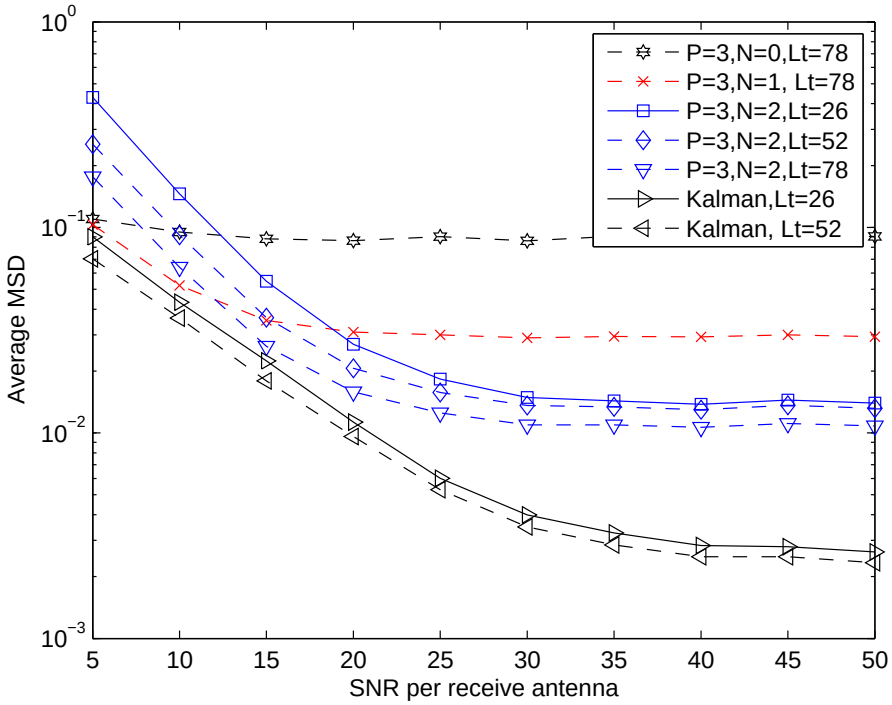

Figure 3.13 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.01.

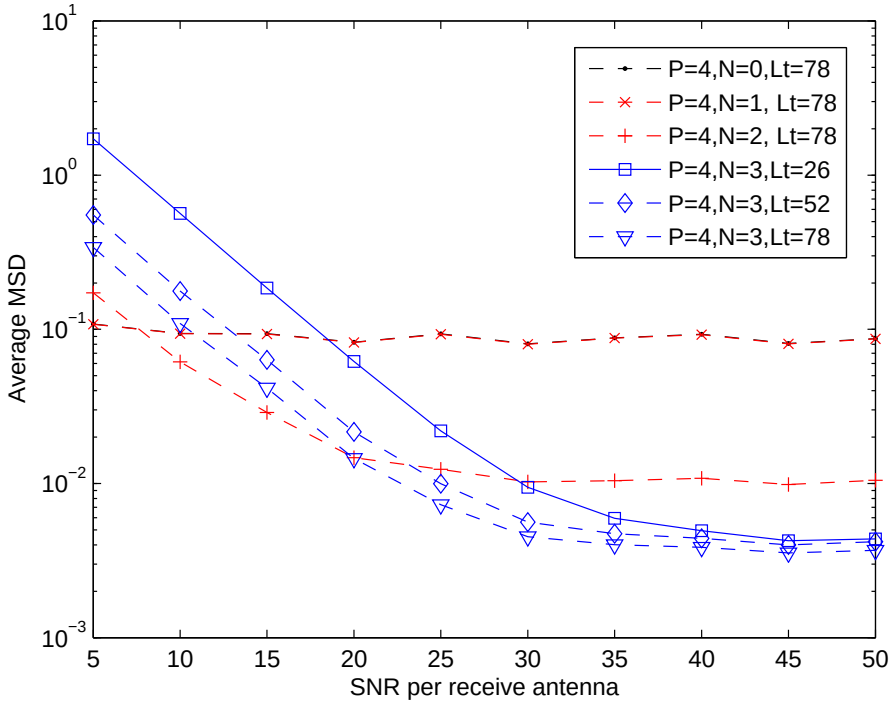

Figure 3.14 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.01. The VRGLS has a predictor length of $P = 4$ and various order N as shown. (Note: the curves for $N = 0$ and $N = 1$ overlap each other)

operate in both slow and fast fading environments because it converges in both scenarios.

These figures also reveal the effect of fade rate on the estimator. A faster fading channel is more difficult to track and hence it introduces a fade-rate-related error. Furthermore, due to the truncation effect of the Taylor series expansion, more terms in the series, hence a higher polynomial order and predictor length[13], are required to support a higher fade rate [113]. This is shown in Fig. 3.14 where a VGRLS estimator with $P = 4$ and $N = 3$ produces a lower MSD at a fade rate of 0.01 compared to the results of Fig. 3.13. However, we note that in slower fading a larger polynomial order and predictor length does not offer any substantial advantage. As shown in Fig. 3.15 for a fade rate of 0.002, the VGRLS estimator with $P = 4$ and $N = 3$ has the same MSD at high SNR as that with $P = 3$ and $N = 2$. We deduce from these results that the 'floor' at

[13]Alternatively the sampling rate can be increased to provide more samples, or the interval of expansions is reduced to preserve the 'smoothness' in the samples.

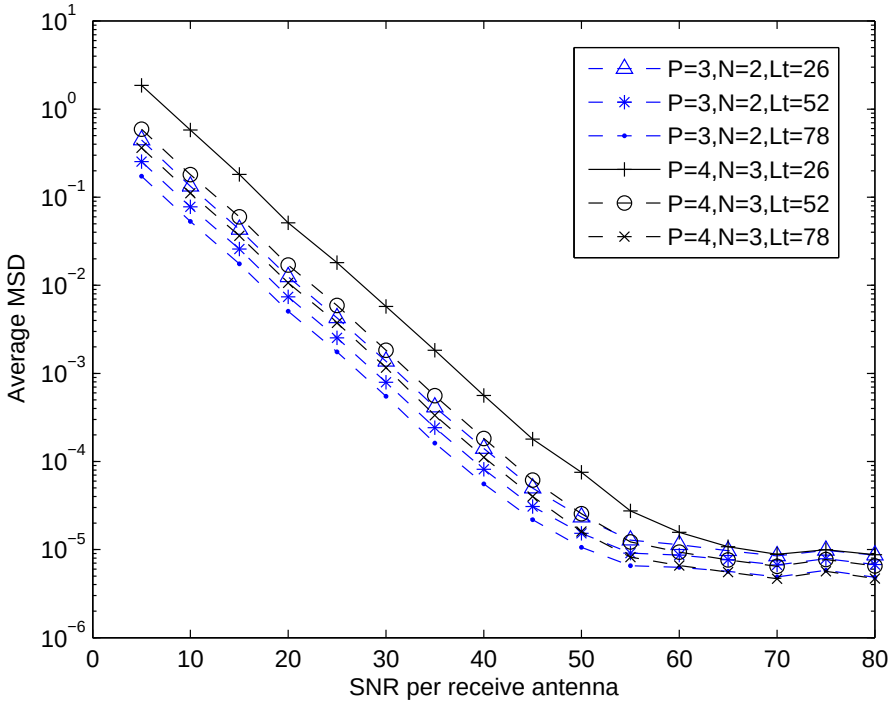

Figure 3.15 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002. The VRGLS has a predictor length of $P = 3$, $N = 2$ and $P = 4$, $N = 3$, with various training sequence length.

high SNR is attributed to fade-rate-related error and cannot be lowered by using higher values of P and N.

3.5.4 Rician Fading Channel

So far all the above simulation results pertain to a Rayeligh fading channel. Here we show that the VGRLS estimator can also operate in a Rician channel readily. We assume that each channel coefficient consists of a non-random (specular) component and a random (diffuse) component as $h_{k,l}^{(j,i)} = h_l^{(nr),(j,i)} + h_{k,l}^{(r),(j,i)}$. The power ratio between the specular and the diffuse components is given by the Rice K-factor,

$$K = \frac{\left|h^{(nr)}\right|^2}{E\left\{\left|h^{(r)}\right|^2\right\}} \tag{3.92}$$

where a K value of 0 corresponds to Rayleigh fading and a large K corresponds to Rician fading. In reality a specular component can be present in any or all of the paths and the value of the K-factor can be the same or different for each path. For simplicity, we assume here that all multipath components contain a specular component with the same value of K.

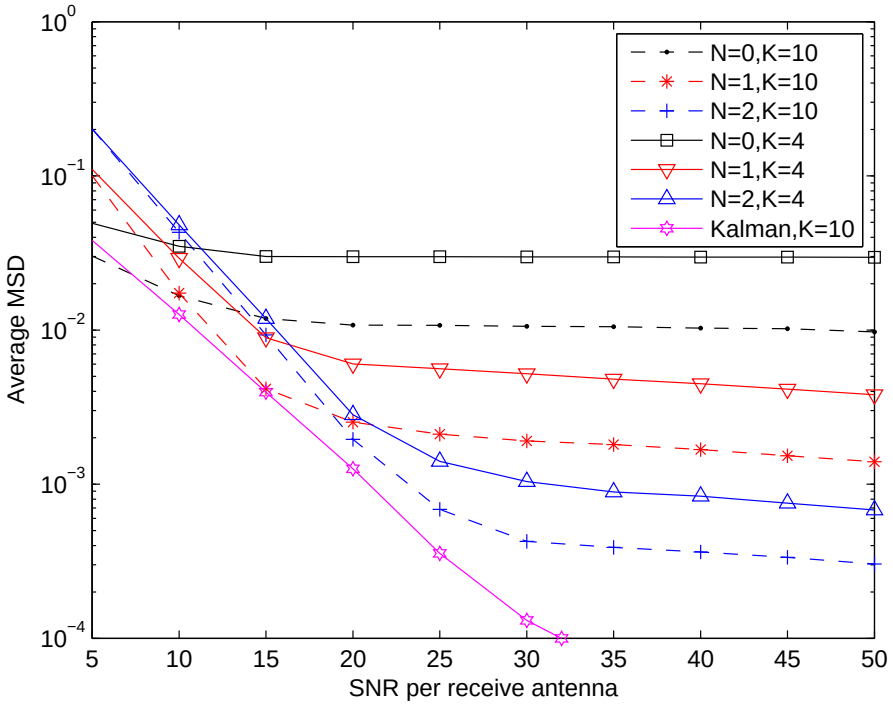

Figure 3.16 MSD of a VGRLS estimator for a (2,2) MIMO system in a Rician fading channel with a normalized fade rate $f_D T_s$ of 0.01 with K-factors of 4 and 10. The VRGLS has a predictor length of $P = 3$ and various polynomial orders N.

Fig. 3.16 and 3.17 show the MSD performances of a VGRLS estimator with a predictor length $P = 3$ with various polynomial orders N and a Kalman estimator. The normalized fade rate is 0.01 in Fig. 3.16 with K-factors of 4 and 10. In obtaining the result for the Kalman estimator, for simplicity, the non-random components of the fading channel are assumed known. In practice, the state transition matrix of the Kalman estimator needs to be reconfigured to suit a Rician fading channel. This is not necessary for the VGRLS estimator as it does not require channel statistics to derive the coefficients of the state transition matrix. When comparing Fig. 3.16 with Fig. 3.13, we can see that the MSD performances of the VGRLS estimator are greatly improved. In a Rician

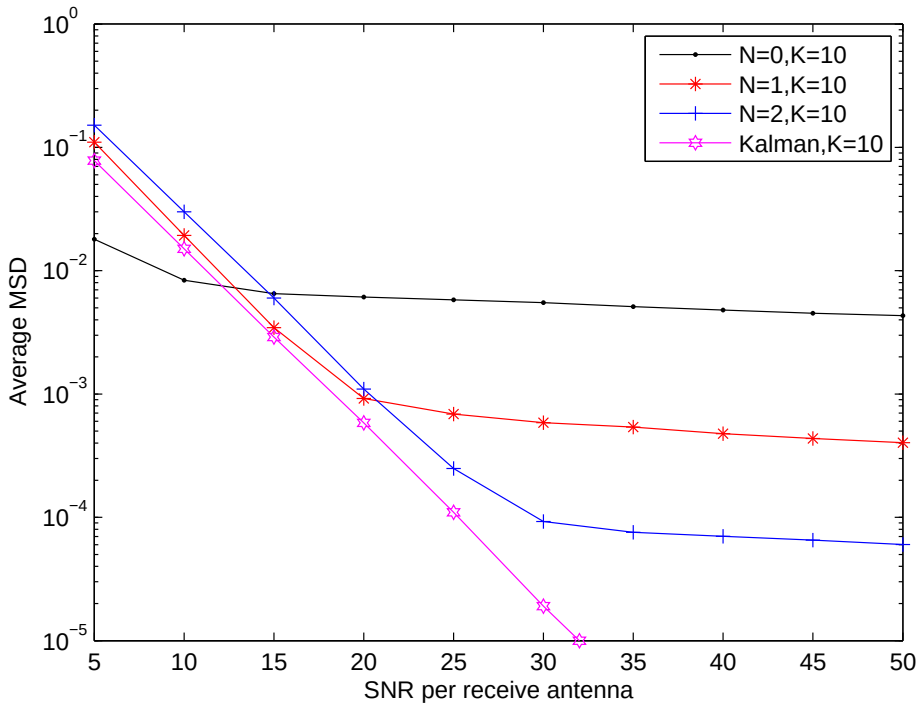

Figure 3.17 MSD of the VGRLS estimator for a (2,2) MIMO system in a Rician fading channel with a normalized fade rate $f_D T_s$ of 0.007 with a K-factor of 10.

environment, the channel is easier to track and hence the cross-over point where a larger N has a better MSD performance occurs at higher SNR. For example, for a K-factor of 4, the cross-over point for N = 0 and 2 occurs at SNR = 12 dB, while for a K-factor of 10, this occurs at about 15 dB. The same is observed for the cross-over point for $N = 0$ and 1. In Fig. 3.17, the MSD performance is for a normalized fade rates of 0.007 and K-factor of 10. These parameters are used because in the subsequent chapter we will compare the performance of an integrated receiver with the results of [100] which use the same parameter values.

3.6 SUMMARY

We could interpret the VGRLS algorithm as a special case of a Kalman filter where there is no process noise vector and the state transition matrix is fixed. The algorithm therefore operates as an approximation to a Kalman filter. As such its performance is not expected to be superior than that of a Kalman filter. This explains why the performance of the VGRLS estimator presented in the

previous section never exceeds that of the Kalman estimator.

However, with appropriately chosen parameters, the VGRLS estimator is able to offer comparable MSD performance to a Kalman estimator, with a similar level of complexity. Furthermore, it has a major advantage over the Kalman estimator in that it does not require any channel statistics, i.e., second order channel statistics and SNR, in order to operate. This is because it employs a fixed state transition matrix that can be derived *a priori* without knowing the statistics. This derivation is based on the theory of polynomial prediction and a Taylor's series expansions of the channel impulse response in the time domain, i.e., t-powers series expansions. The estimator arrives at an approximate channel state model with unforced dynamics which does not require knowledge of the process noise autocovariance, and hence the SNR. This makes the VGRLS algorithm an attractive alternative to the Kalman estimator.

The MSD performance of the VGRLS estimator is affected by several parameters, some of which are inherent to the algorithm, such as the predictor length, P, polynomial order, N, and forget factor, λ, while the rest are determined by the conditions under which it is operated, such as the normalized channel fade rate, $f_D T_s$, system SNR and the length of the training sequence, L_t.

It is noted that the algorithm's parameters tend to have opposing effects across the range of SNR studied. For example, for a fixed predictor length P, a higher polynomial order N offers a better fit to the fading process, hence reducing the residual error and offers a better MSD at high SNR; however, with a higher polynomial order the norm of the polynomial coefficients is larger thereby increasing the effective variance component due to the AWGN, resulting in worse MSD especially at low SNR. Similarly, adjusting the 'forget factor' λ will also result in this contradicting effect on the MSD across the SNR range. Therefore a chosen set of parameters may not be suitable for every condition and adaptive adjustment of some of the parameters, e.g. λ, may be necessary.

Nevertheless, the simulation results under a broad range of channel conditions have indicated that with sufficient training in the mid to high SNR range of 25

to 50 dB and for the range of fade rates studied, the VGRLS estimator with a predictor length $P = 3$, polynomial order $N = 2$, and forget factor $\lambda = 0.9$ offers excellent MSD performance.

Chapter 4

REDUCED COMPLEXITY CHANNEL ESTIMATION

4.1 INTRODUCTION

We have shown in Chapter 3 that the VGRLS estimator is able to offer comparable 'mean square deviation' (MSD) performance to an optimum Kalman estimator and at a similar level of complexity, namely $\mathcal{O}\big((\mathcal{N})^3\big)$ real operations per iteration, where $\mathcal{N}$ is the dimension of the channel state vector of the VGRLS algorithm. The primary computational load of the VGRLS estimator (also the Kalman estimator) is the recursive computation of the Riccatti update equation for $\mathbf{P}_{k/k}$. As noted, this computation does not depend on the received samples $\mathbf{y}_k$. Referring to Fig. 3.4, with known inputs, i.e. $\mathbf{d}_k$, $\mathbf{U}$, λ and δ, the Riccatti equation can actually be computed *before* the algorithm is initiated. This provides a basis for computing the equation *off-line* which offers complexity reduction in the implementation of the algorithm.

The potential for this simplification is recognized by [30] which replaces the on-line computation of the Riccatti equation with an off-line recursive computation. This results in a reduced complexity estimator known as the polynomial-based *generalized least mean squares* (GLMS) algorithm. With appropriately chosen predictor lengths and polynomial orders, the GLMS estimator offers substantially better performance than that of a conventional LMS algorithm. However, in general its performance is not as good as the GRLS and Kalman estimators. The degradation in performance represents a trade-off for the reduced

computational load.

The GLMS algorithm is interesting in that it represents a *state-space* approach to a conventional LMS algorithm. It achieves this by being a *model dependant* algorithm which is a departure from the traditional *model independent* approach of the LMS algorithm [91]. As such, it offers better tracking in a fading environment, but at an increased level of complexity. With $\mathcal{O}\big((\mathcal{N})^2\big)$ real operations per iteration, it is an order of magnitude more complex than the LMS algorithm at $\mathcal{O}\big((\mathcal{N})\big)$ real operations per iteration. Since the GLMS algorithm is a simplification of the GRLS algorithm, it provides an approach to reducing the complexity of a Kalman-like algorithm.

We note that there have been several approaches to reducing the complexity of an actual Kalman estimator. In [116], an LMS-like channel estimation algorithm is proposed by replacing the online computation of Riccatti equation of a Kalman filter with an equivalent algebraic equation that is pre-computed. This equation is dependant on a model of the channel dynamics and is different for various models. Examples are shown using an autoregressive (AR) second order model and an integrated random walk model. Furthermore, similar to the Kalman filter, it requires channel statistics in order to derive the AR parameters and the process noise autocovariance.

In [117], a predictive LMS-type channel estimator known as the Wiener LMS (WLMS) algorithm is proposed. It employs Wiener filters and also requires the channel covariance. Another predictive LMS-type estimator, known as the modified LMS estimator, is proposed in [118] by simplifying a Kalman filter. It is only considered for a second order Markov model of the channel, and is equivalent to a special case of the GLMS algorithm with a predictor length $P = 2$ and polynomial order $N = 1$. On the other hand, the GLMS algorithm of [30] may be used for higher order polynomial models of the channel. In contrast to these algorithms, the GLMS algorithm, like its predecessor GRLS algorithm [29], also does not require any *a priori* knowledge of the channel statistics.

Motivated by [30], we investigate the reduced complexity form of the VGRLS

algorithm of Chapter 3. As the resulting algorithm is a vector form of the GLMS algorithm, we call it a polynomial predictor based *Vector GLMS* (VGLMS) algorithm. In the following sections, we derive the algorithm and compare its performance to that of the VGRLS algorithm.

4.2 COMPLEXITY REDUCTION OF THE VGRLS ALGORITHM

Assuming that the channel state vector $\hat{\mathbf{h}}_{k/k-1}$ and the intermediate matrix $\mathbf{P}_{k/k-1}$ are known, the update equations for the VGRLS algorithm may be expressed as

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{d}_k^H \left(\mathbf{I}_{RN_r} + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H\right)^{-1} \tag{4.1}$$

$$\mathbf{P}_{k/k} = \left(\mathbf{I}_{RN_rTLP} - \mathbf{K}_k\mathbf{d}_k\right)\mathbf{P}_{k/k-1} \tag{4.2}$$

$$\hat{\mathbf{h}}_{k/k} = \hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right) \tag{4.3}$$

and the prediction equations may then be written as

$$\hat{\mathbf{h}}_{k+1/k} = \mathbf{U}\hat{\mathbf{h}}_{k/k} \tag{4.4}$$

$$\mathbf{P}_{k+1/k} = \lambda^{-1}\mathbf{U}\mathbf{P}_{k/k}\mathbf{U}^H. \tag{4.5}$$

In order to reduce the complexity of the VGRLS algorithm, it is necessary to replace the on-line recursive computation of $\mathbf{P}_{k/k}$. We follow the approach of [92] where in analyzing the steady state tracking performance of the RLS algorithm, it is assumed that the elements of $\mathbf{P}_{k/k}$ converge to some fixed values represented by the elements of a matrix $\hat{\mathbf{P}}$. This is also assumed in [30] where an approximation to $\hat{\mathbf{P}}$ is achieved by inverting $\hat{\mathbf{P}}^{-1}$ which can be approximated by $lim_{k\to\infty}E\left[\mathbf{P}_{k/k}^{-1}\right]$. In the following we will show the derivation of $\hat{\mathbf{P}}$.

First, we note that in steady state the inverse of $\mathbf{P}_{k/k}$, i.e. $\mathbf{P}_{k/k}^{-1}$, can be modelled as [92]

$$\mathbf{P}_{k/k}^{-1} = E\left[\mathbf{P}_{k/k}^{-1}\right] + \vartheta_k \tag{4.6}$$

where ϑ_k is a zero mean complex Gaussian perturbation matrix. For $\lambda \rightarrow 1$ [92],[26], the variance of the elements of the perturbation matrix will be small, and hence $\hat{\mathbf{P}}^{-1} = lim_{k\rightarrow\infty} E\left[\mathbf{P}_{k/k}^{-1}\right]$.

Next, we recall equations (3.73) and (3.75) of Chapter 3, namely

$$\begin{aligned} \mathbf{P}_{k/k-1}^{-1} &= \lambda\left(\mathbf{U}_{k/k-1}\mathbf{P}_{k-1/k-1}\mathbf{U}_{k/k-1}^{H}\right)^{-1} \\ &= \lambda\mathbf{U}_{k-1/k}^{H}\mathbf{P}_{k-1/k-1}^{-1}\mathbf{U}_{k-1/k}, \end{aligned} \tag{4.7}$$

and

$$\mathbf{P}_{k/k}^{-1} = \mathbf{P}_{k/k-1}^{-1} + \mathbf{d}_k^H\mathbf{d}_k. \tag{4.8}$$

We then use them to evaluate $\hat{\mathbf{P}}^{-1}$ as

$$\hat{\mathbf{P}}^{-1} = lim_{k\rightarrow\infty} E\left[\mathbf{P}_{k/k}^{-1}\right] = lim_{k\rightarrow\infty} E\left[\lambda\mathbf{G}^H\mathbf{P}_{k-1/k-1}^{-1}\mathbf{G} + \mathbf{d}_k^H\mathbf{d}_k\right] \tag{4.9}$$

where $\mathbf{G}$ is the backward transition matrix defined as $\mathbf{G} = \mathbf{U}^{-1}$ [49]. Expanded backward recursively using (4.7) and (4.8), we may write $E\left[\mathbf{P}_{k/k}^{-1}\right]$ as

$$\begin{aligned} E\left[\mathbf{P}_{k/k}^{-1}\right] &= \lambda\mathbf{G}^H E\left[\lambda\mathbf{G}^H\mathbf{P}_{k-2/k-2}^{-1}\mathbf{G} + \mathbf{d}_k^H\mathbf{d}_k\right]\mathbf{G} + E\left[\mathbf{d}_k^H\mathbf{d}_k\right] \\ &= \lambda^k\left(\mathbf{G}^H\right)^k\mathbf{P}_{0/-1}^{-1}\left(\mathbf{G}\right)^k + \sum_{l=1}^{k}\lambda^l\mathbf{G}^H\mathbf{R}_\mathbf{d}\mathbf{G} + \mathbf{R}_\mathbf{d} \\ &= \lambda^k\left(\mathbf{G}^H\right)^k\mathbf{P}_{0/-1}^{-1}\left(\mathbf{G}\right)^k + \left(\frac{\lambda^{k+1}-\lambda}{\lambda-1}\right)\mathbf{G}^H\mathbf{R}_\mathbf{d}\mathbf{G} + \mathbf{R}_\mathbf{d} \end{aligned} \tag{4.10}$$

where $E\left[\mathbf{d}_k^H \mathbf{d}_k\right] = \mathbf{R_d}$ is the autocorrelation matrix of $\mathbf{d}_k$.

It is generally intractable [26] to evaluate the asymptotic value of the first term on the r.h.s. of (4.10) as $k \to \infty$. As an alternative, we may run (4.10) recursively for some suitably chosen large value of k by assuming that the system reaches steady state for that value of k. An approximation to $\hat{\mathbf{P}}$ is then obtained by inverting the resulting matrix. However, there are instances for which the matrix is ill-conditioned and direct inversion leads to inaccurate results.

To circumvent this problem, we resort to evaluating (4.10) recursively *without* involving matrix inversion. We assume that the autocorrelation matrix of $\mathbf{d}_k$ may be factorized as $\mathbf{R_d} = \mathcal{D}^H\mathcal{D}$, where $\mathcal{D}$ is a square matrix of size RN_rTLP x RN_rTLP. Then, (4.10) may be expressed as

$$\begin{aligned} E\left[\mathbf{P}_{k/k}^{-1}\right] &= \lambda \mathbf{G}^H E\left[\mathbf{P}_{k-1/k-1}^{-1}\right]\mathbf{G} + \mathcal{D}^H\mathcal{D} \\ &= E\left[\mathbf{P}_{k/k-1}^{-1}\right] + \mathcal{D}^H\mathcal{D} \end{aligned} \tag{4.11}$$

where

$$E\left[\mathbf{P}_{k/k-1}^{-1}\right] = \lambda \mathbf{G}^H E\left[\mathbf{P}_{k-1/k-1}^{-1}\right]\mathbf{G}. \tag{4.12}$$

Using the *matrix inversion lemma*[1], and by defining $A = E\left[\mathbf{P}_{k/k}^{-1}\right]$, $B^{-1} = E\left[\mathbf{P}_{k/k-1}^{-1}\right]$, $C = \mathcal{D}^H$, $D^{-1} = \mathbf{I}_{RN_rTLP}$, $C^H = \mathcal{D}$, the inverse of equation (4.11) may be expressed as

$$\begin{aligned} E\left[\mathbf{P}_{k/k}^{-1}\right]^{-1} &= E\left[\mathbf{P}_{k/k-1}^{-1}\right]^{-1} \\ &\quad - E\left[\mathbf{P}_{k/k-1}^{-1}\right]^{-1}\mathcal{D}^H\left(\mathbf{I}_{RN_rTLP} + \mathcal{D}E\left[\mathbf{P}_{k/k-1}^{-1}\right]^{-1}\mathcal{D}^H\right)^{-1} E\left[\mathbf{P}_{k/k-1}^{-1}\right]^{-1}. \end{aligned} \tag{4.13}$$

[1]Given $A = B^{-1} + CD^{-1}C^H$, the inverse of A is given by $A^{-1} = B - BC(D + C^HBC)^{-1}C^HB$. For more details see [49].

By further defining $\hat{\mathbf{P}}_{k/k} = E\left[\mathbf{P}_{k/k}^{-1}\right]^{-1}$ and $\hat{\mathbf{P}}_{k/k-1} = E\left[\mathbf{P}_{k/k-1}^{-1}\right]^{-1}$, this can be reduced to a simplified form as

$$\hat{\mathbf{P}}_{k/k} = \hat{\mathbf{P}}_{k/k-1} - \hat{\mathbf{P}}_{k/k-1}\mathcal{D}^H\Psi_{k/k-1}\mathcal{D}\hat{\mathbf{P}}_{k/k-1} \tag{4.14}$$

where

$$\Psi_{k/k-1} = \left(\mathbf{I}_{RN_rTLP} + \mathcal{D}\hat{\mathbf{P}}_{k/k-1}\mathcal{D}^H\right)^{-1}. \tag{4.15}$$

Using similar definitions (4.12) may also be expressed as

$$\hat{\mathbf{P}}_{k/k-1} = \lambda^{-1}\mathbf{U}\hat{\mathbf{P}}_{k-1/k-1}\mathbf{U}^H. \tag{4.16}$$

In steady state, the 'intermediate' matrix $\hat{\mathbf{P}}$ is well approximated by $\hat{\mathbf{P}}_{k/k}$ for large values of k and $\hat{\mathbf{P}}_{0/-1} = \mathbf{P}_{0/-1}$. In this manner, the matrix $\hat{\mathbf{P}}$ may be obtained by computing $\hat{\mathbf{P}}_{k/k}$ recursively off-line following (4.16), (4.15) and (4.14) and without any of the numerical problems associated with matrix inversion. It has been observed that as $\lambda \to 1$, a large value of k is needed for $\hat{\mathbf{P}}_{k/k}$ to reach steady state. The number of recursions required to obtain a good approximation was not mentioned in [30]. However, no more than 1000 iterations were empirically found to be sufficient for the channels employed in this thesis.

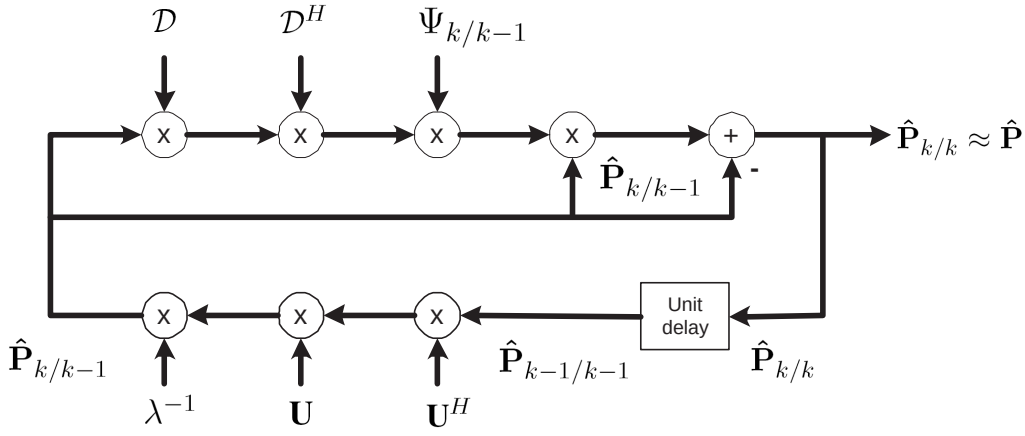

Figure 4.1 Signal flow diagram of the time-invariant intermediate matrix $\hat{\mathbf{P}}$.

To evaluate the time-invariant approximation $\hat{\mathbf{P}}$ requires knowledge of the

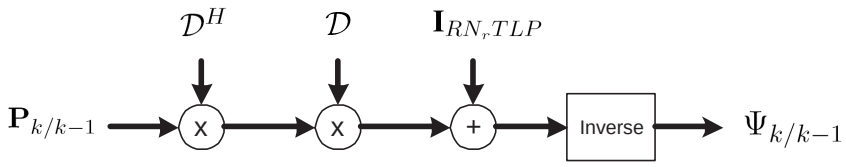

Figure 4.2 Signal flow diagram of $\Psi_{k/k-1}$.

autocorrrelation matrix of the data vector $\mathbf{R_d} = E\left[\mathbf{d}_k^H \mathbf{d}_k\right]$, the state transition matrix $\mathbf{U}$ and the 'forget factor' λ, and in general these parameters are known *a priori*. A signal flow diagram for $\hat{\mathbf{P}}$ and $\Psi_{k/k-1}$ is given in Fig. 4.1 and 4.2.

From (3.80), we recognize that the Kalman gain $\mathbf{K}_k = \mathbf{P}_{k/k}\mathbf{d}_k^H$ and by replacing $\mathbf{P}_{k/k}$ with $\hat{\mathbf{P}}$, we obtain the approximation

$$\mathbf{K}_k = \hat{\mathbf{P}}\mathbf{d}_k^H \tag{4.17}$$

and given that $\mathbf{d}_k$ is the vector of known training symbols, it can also be derived *a priori*.

Substituting $\mathbf{P}_{k/k} = \hat{\mathbf{P}}$ in the VGRLS algorithm of equations (4.1) to (4.5), we may write the real time one-step prediction equation of the simplified estimation algorithm as

$$\begin{aligned}
\hat{\mathbf{h}}_{k+1/k} &= \mathbf{U}\hat{\mathbf{h}}_{k/k} \\
&= \mathbf{U}\left[\hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right)\right] \\
&= \mathbf{U}\left[\hat{\mathbf{h}}_{k/k-1} + \hat{\mathbf{P}}\mathbf{d}_k^H\left(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}\right)\right].
\end{aligned} \tag{4.18}$$

The corresponding signal flow diagram is given in Fig. 4.3.

We call the reduced complexity algorithm of (4.18) a *Vector GLMS* (VGLMS) algorithm. We note that in the special case where $P = 1$ and $N = 0$, the algorithm reduces to a conventional vector LMS algorithm with a step size $\mu = 1 - \lambda$.

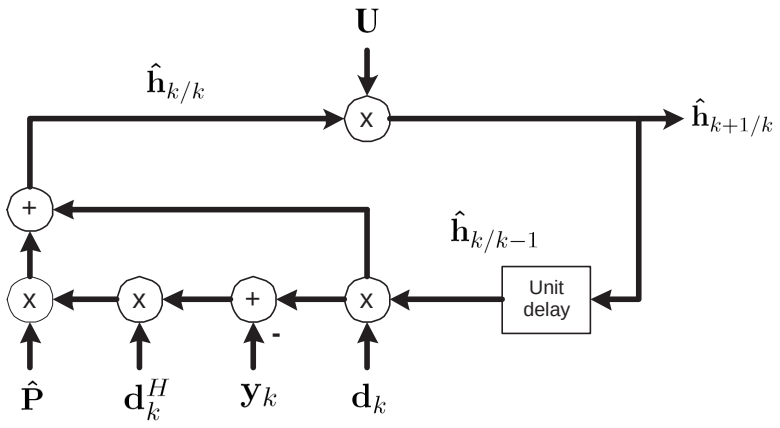

Figure 4.3 Signal flow diagram of the simplified one step prediction.

Table 4.1 Numbers of real operation per iteration of the VGRLS and VGLMS algorithms for a (2, 2) MIMO system with a delay spread, $L = 3$, $N_r = 1$ sample per symbol and a predictor length, $P = 3$.

Algorithm Type	$\times$	$+$	$\div$
VGRLS	400608	387288	4
VGLMS	14400	11448	4

4.3 A COMPLEXITY COMPARISON OF THE VGRLS AND VGLMS ALGORITHMS

We compare the complexity of the VGRLS and VGLMS algorithms for the following scenario: a (2,2) MIMO system with $L = 3$ multipath rays in each of the subchannels and a predictor length of $P = 3$ and $N_r = 1$ sample per symbol. Table 4.1 gives the numbers of real operation[2] needed per iteration of the algorithms.

It is clear from Table 4.1 that substantial computational savings are achieved by the VGLMS algorithm. This is because for VGRLS the on-line recursion of (4.2) for computing $\mathbf{P}_{k/k}$ requires $\mathcal{O}\left((RN_rTLP)^3\right)$ real operations per iteration in the highest term of calculations. In fact this constitutes the bulk of the computational complexity and by getting rid of this online recursion, we reduce the complexity of the algorithm to just $\mathcal{O}\left((RN_rTLP)^2\right)$ real operations. The sav-

[2]We assume that one complex multiplication requires 4 real multiplications and 2 real additions; while one complex addition requires 2 real additions.

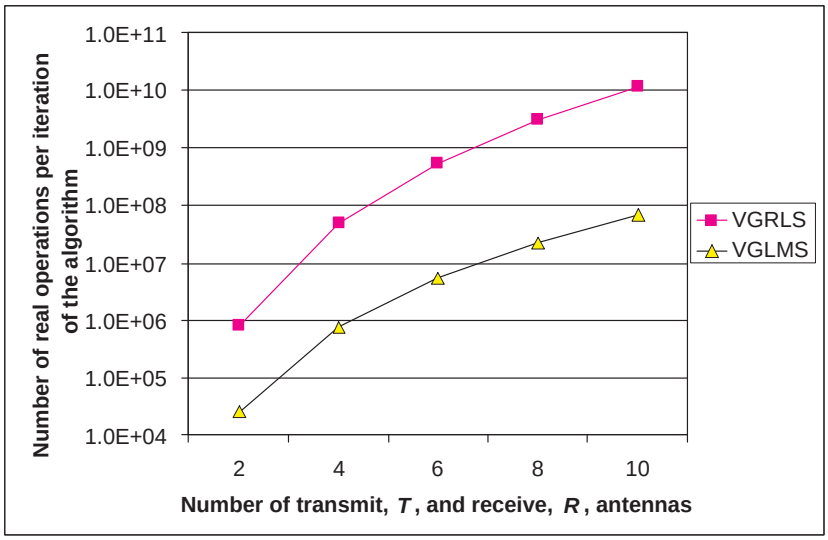

Figure 4.4 Comparison of the required numbers of real operation per iteration of the VGRLS and VGLMS algorithms with a fixed predictor length $P = 3$, $N_r = 1$ sample per symbol and delay spread of $L = 3$ for various (T, R) MIMO system for $T = R$.

ings become more significant as the dimension of the MIMO system increases, as shown in Fig. 4.4.

4.4 PERFORMANCE EVALUATION

The 'mean square deviation' (MSD) performance of the VGLMS estimator is presented in this section. The MSD estimates the average squared norm, or error, between the original and the estimated channel impulse responses in steady state. The estimator is operated alone with known transmitted frames. Various lengths, $L_t = 52$, 104 and 208, of training sequence followed by 116 data symbols, all using QPSK modulation, are randomly generated. We assume that the estimator is in transient mode during the training period, after which it settles into steady state mode. Hence for MSD evaluation the first L_t symbols of each frame are not included. At the beginning of each frame the estimator re-initializes and starts acquisition again. Simulation results for the estimator with predictor lengths $P = 3$, 4 and polynomial orders $N = 0$, 1, 2, 3 are presented.

Fig. 4.5 shows the average MSD performance of the VGLMS estimator with a predictor length $P = 3$ in a (2,2) MIMO Rayleigh fading channel with a nor-

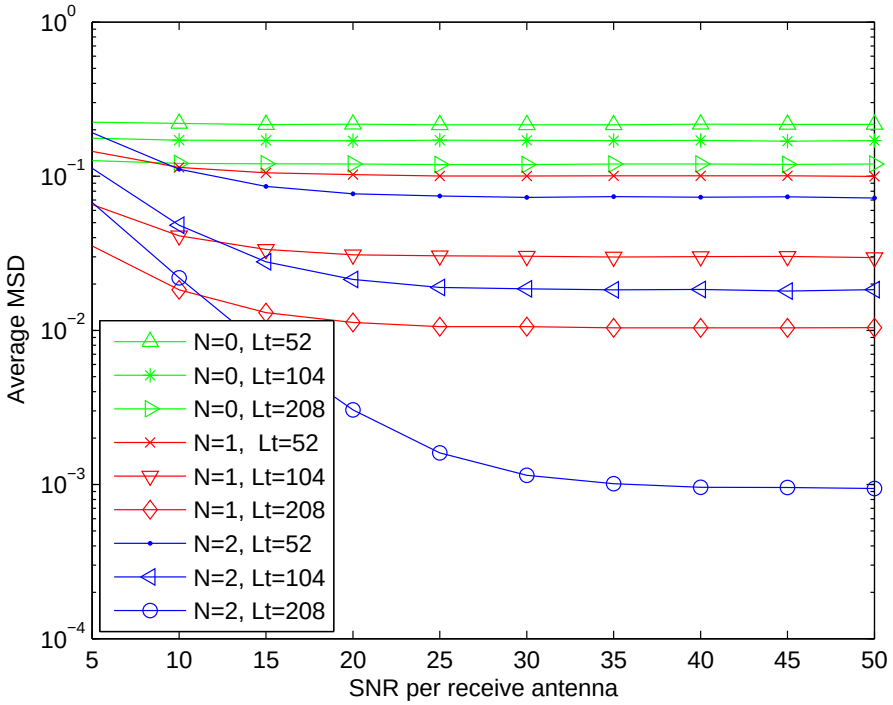

Figure 4.5 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.002$ and $\lambda = 0.95$.

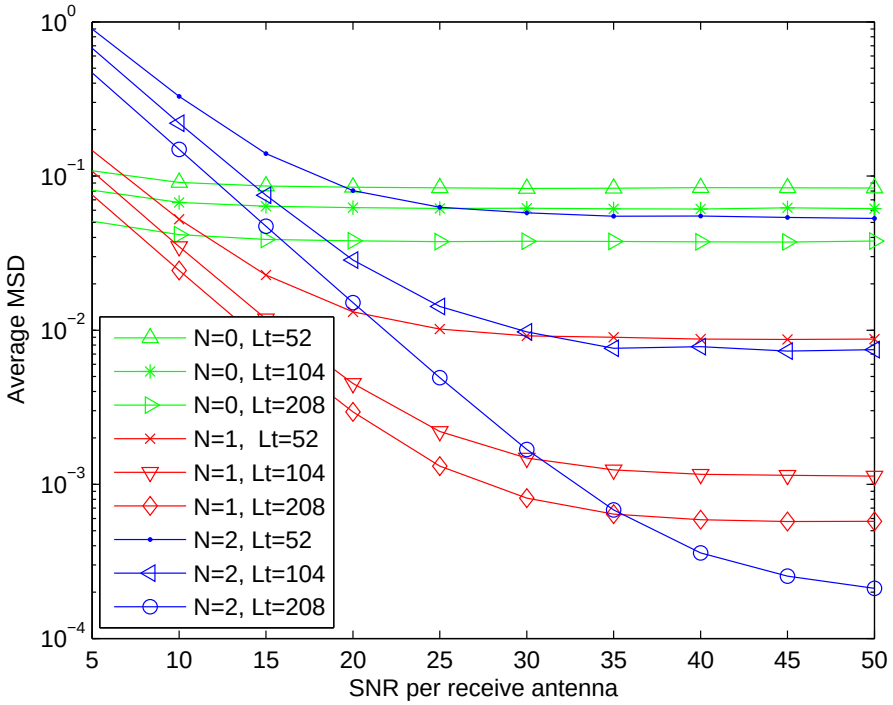

Figure 4.6 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.002$ and $\lambda = 0.90$.

malized fade rate $f_D T_s = 0.002$ and a forget factor $\lambda = 0.95$. The performance for $N = 0$, which is equivalent to the vector LMS algorithm, is relatively very poor at 10^{-1} and that of $N = 1$ is in between at approximately 10^{-2}. Similar to the VGRLS estimator, the VGLMS estimator with $N = 2$ offers the best performance. By changing the forget factor λ to 0.90, as shown in Fig. 4.6, the performance for $N = 1$ and 2 is improved at high SNR although for $N = 2$, the performance tends to get worse below 35dB. In general, the MSD performance of the VGLMS estimator is worse than that of the VGRLS estimator. For example comparing Figs. 4.6 and 3.6 at high SNR the MSD for VGLMS estimator with $N = 2$ is close to 10^{-3} whereas for that of the VGRLS estimator it gets very close to 10^{-5}.

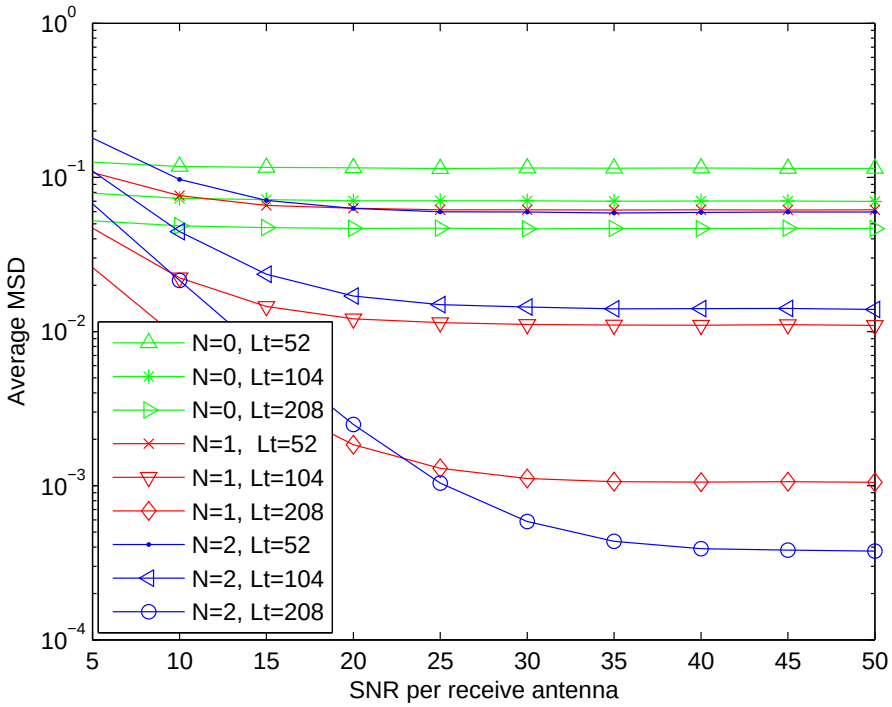

Figure 4.7 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.0001$ and $\lambda = 0.95$.

Fig. 4.7 and 4.8 show the MSD performance of the VGLMS estimator with similar settings as above but with a normalized fade rate of 0.0001. The results show that the performance of the VGLMS estimator is better in a slower fading environment. With $\lambda = 0.95$, $N = 2$ offers the best performance whereas with λ

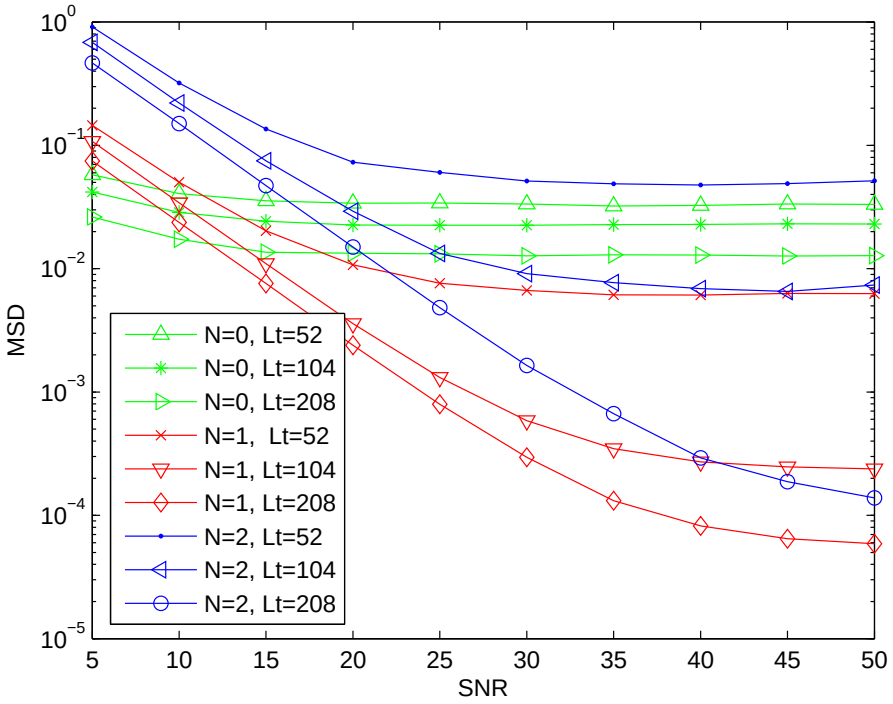

Figure 4.8 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.0001$ and $\lambda = 0.90$.

$= 0.90$, $N = 1$ offers the best performance. We note that in general the VGLMS estimator requires a longer training sequence, about twice that required by the VGRLS estimator, to converge. The MSD performance for faster fading with a normalized fade rate of 0.01 is shown in Fig. 4.9 and 4.10. At this fade rate, the estimator does not really converge with $\lambda = 0.95$, and with $\lambda = 0.90$ the MSD performance is still very poor even though it converges to about 10^{-1}. This implies that the VGLMS estimator does not track a very fast fading environment well.

Now we show in Fig. 4.11 and 4.12 the MSD performance of the VGLMS estimator with a predictor length $P = 4$, forget factor $\lambda = 0.95$ and 0.90, at a normalized fade rate of 0.002. The performance is slightly worse than that of $P = 3$. For $\lambda = 0.90$, $N = 3$ does not converge. This also occurs at a normalized fade rate of 0.0001. The results show that using a predictor length $P = 3$ is sufficient for most situations.

We show the MSD performance of the VGLMS estimator in a Rician fading

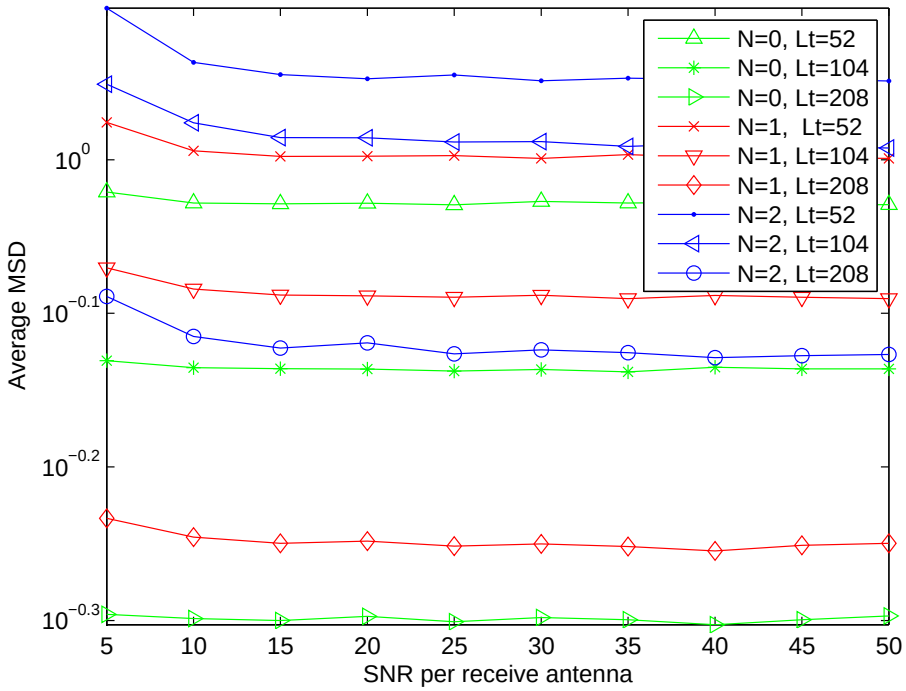

Figure 4.9 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.01$ and $\lambda = 0.95$.

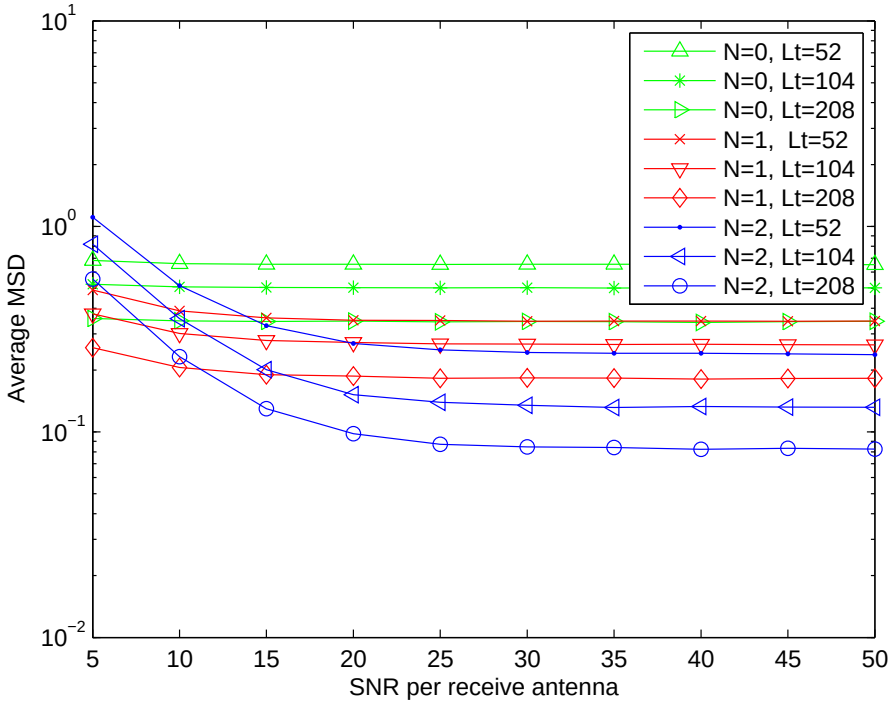

Figure 4.10 MSD of the VGLMS estimator for $P = 3$ and various N in a Rayleigh fading with $f_D T_s = 0.01$ and $\lambda = 0.90$.

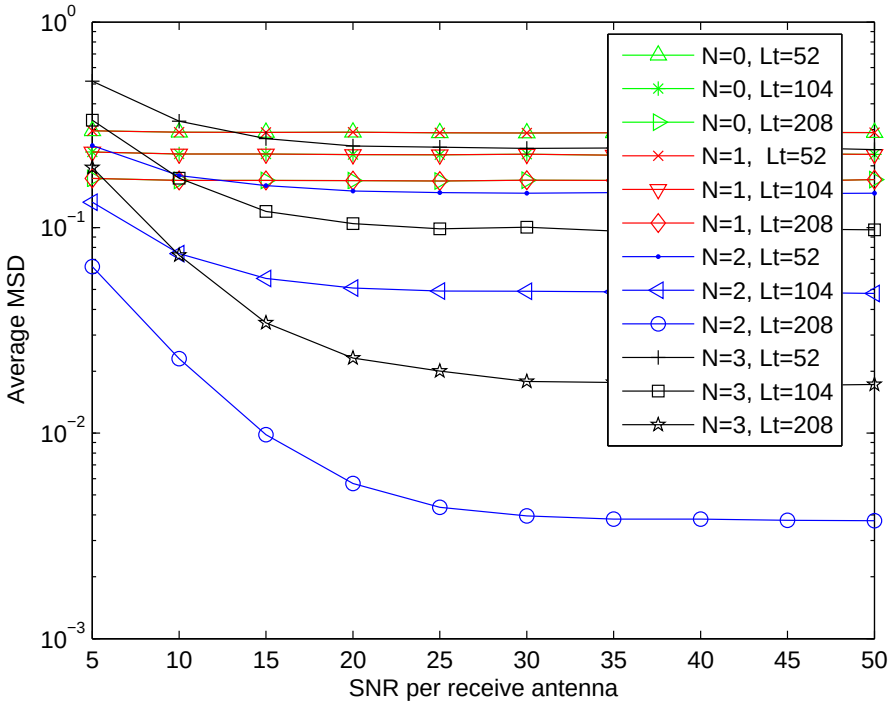

Figure 4.11 MSD of the VGLMS estimator for $P = 4$ and various N in a Rayleigh fading with $f_D T_s = 0.002$ and $\lambda = 0.95$.

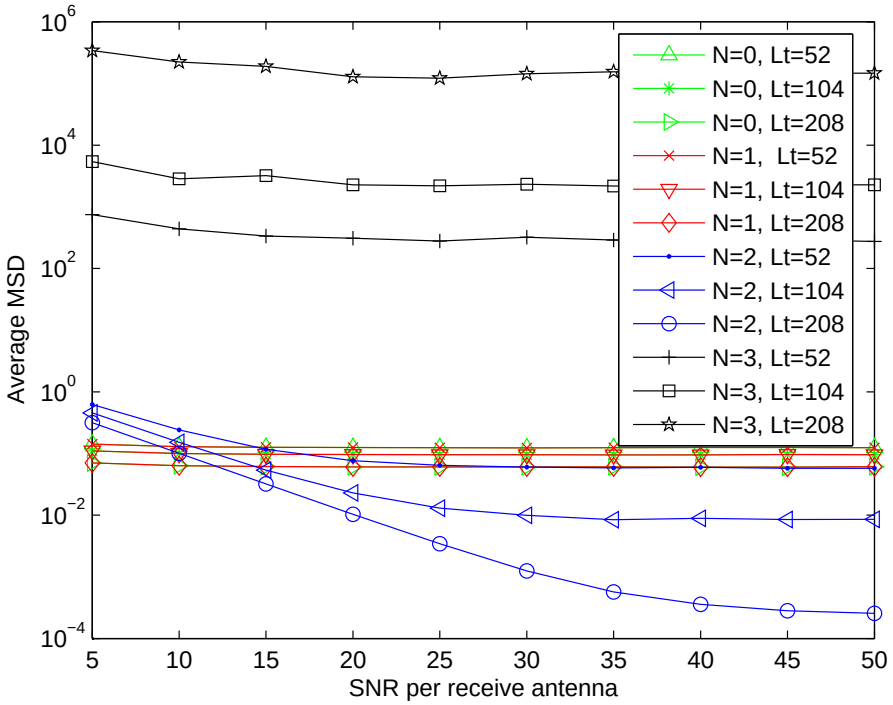

Figure 4.12 MSD of the VGLMS estimator for $P = 4$ and various N in a Rayleigh fading with $f_D T_s = 0.002$ and $\lambda = 0.90$.

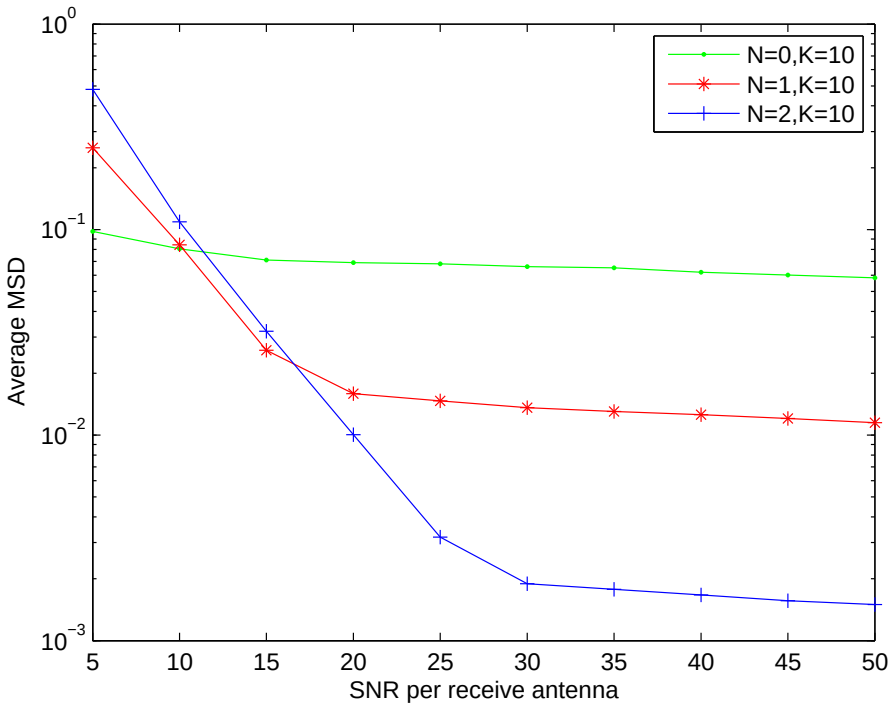

Figure 4.13 MSD of the VGLMS estimator for $P = 3$ and various N in a Rician fading with $f_D T_s = 0.007$, $K = 10$ and $\lambda = 0.90$.

channel with a normalized fade rates of 0.007 and K-factor of 10 in Fig. 4.13. The result shows that the VGLMS estimator, like the VGRLS estimator, can readily operate in a Rician fading environment without requiring reconfiguration of the state transition matrix, and the MSD performance is also greatly improved in such a fading channel.

4.5 SUMMARY

A reduced complexity channel estimator based on the previously derived VGRLS estimator, known as the VGLMS estimator, has been developed. This is achieved by replacing the online recursive computation of the Riccatti equation by a offline approximation that can be computed *a priori*. This reduces the computational load of the algorithm by an order of magnitude. However, as a result, the MSD performance of the VGLMS estimator is not as good as the VGRLS estimator. It is shown that the VGLMS estimator can still offer sufficiently low MSD in slow fading although it is not suitable for fast fading.

Chapter 5

INTEGRATED SYMBOL-BY-SYMBOL BASED RECEIVER

5.1 INTRODUCTION

A polynomial predictor based channel estimator for MIMO fading dispersive channels known as the VGRLS estimator was developed in Chapter 3. Simulation results show that the VGRLS estimator is able to offer comparable MSD performance to an optimum Kalman filter based estimator without requiring knowledge of channel statistics. Its structure and complexity is similar to the Kalman estimator. A simplified version of the estimator known as the VGLMS estimator was developed in Chapter 4. Even though the performance of the VGLMS estimator is not as good as its predecessor the VGRLS estimator, it offers significant reduction in computational complexity.

As an application of the channel estimators, we now integrate them into a receiver employing a vector decision feedback equalizer (DFE) structure similar to those of [100] and [31]. Channel estimates from either the VGRLS or the VGLMS channel estimator are used to calculate the filter tap coefficients of the DFE.

Intersymbol interference (ISI) is a significant impairment in digital communication systems operating in a frequency selective fading environment. System performance, in terms of average symbol error probability, may be severely degraded by ISI. For ISI-corrupted received signals, maximum likelihood sequence estimation (MLSE) is an optimum equalization method [36],[35]. However, for

a given modulation format, its complexity increases exponentially with both the channel delay spread, L, and number of transmit antennas, T. Furthermore, its decision delay is significant (typically about $5L$) [36]. We have, therefore, employed a vector minimum mean square error (MMSE) DFE [31] structure here.

The principle of a DFE is to use previous decisions, combined with a knowledge of the channel response, to form an estimate of that portion of the ISI at the decision instant due to previously transmitted symbols. This estimate is then subtracted from the received signal, thereby reducing the effect of ISI. If the previous decisions are all correct and the channel response is known perfectly, the ISI due to previously detected symbols is eliminated entirely.

A DFE implementation consists of a linear feedforward filter, which attempts to equalize the ISI due to symbols transmitted in the future, followed by a decision device and a feedback filter. The output of the decision device is input to the feedback filter to form an estimate of the ISI due to previously transmitted symbols. This estimate is subtracted from the signal at the input to the decision device. Both filters are usually implemented as tapped delay lines. The number of taps in the feedback filter determines the number of previous decisions which are assumed to affect the current decision.

The decisions at the output of the DFE are used as reference signals for the channel estimator in updating the channel estimates. The current decisions are needed to obtain up-to-date channel estimates which are required when recalculating the filter tap coefficients of the DFE. However, the decision delay of the DFE, which arises as the received signal passes through the feedforward filter, causes the output at the decision device to be delayed. Therefore the channel estimates are also delayed, and a 'time-lag' is created. In order to bridge this gap, channel prediction is employed. Here a simple polynomial-based channel prediction module which exploits the fixed *a priori* derived polynomial coefficients employed in the VGRLS and VGLMS estimators, is employed for predicting the channel response.

In the following sections, we describe the overall receiver structure which in-

cludes the development of the vector DFE, the integration with the VGRLS/VGLMS channel estimator and the polynomial-based channel prediction module.

5.2 INPUT-OUTPUT SIGNAL MODEL

The input-output signal model used in this chapter is defined here. Following the definitions used in [31], we treat the transmitted data symbol as the *input* and the received signal as the *output* of the channel. For convenience we reproduce some of the signal model equations used in Chapter 3 here.

At time k, the j-th symbol-rate channel output may be written as

$$y_k^{(j)} = \sum_{i=1}^{T} \sum_{l=0}^{L-1} d_{k-l}^{(i)} h_{k,l}^{(j,i)} + n_k^{(j)} \;\; ; \;\; j = 1, 2, \cdots, R \tag{5.1}$$

where $d_k^{(i)}$ is the k-th transmitted complex baseband M-ary input data symbol from the i-th antenna, $\{h_{k,l}^{(j,i)}\}_{l=0}^{l=L-1}$ is the sampled fading dispersive composite[1] channel impulse response between the i-th transmit and j-th receive antennas at time k with delay spread of L symbol periods, and $n_k^{(j)}$ is sampled additive white Gaussian noise (AWGN) with variance, σ_n^2.

Assuming an oversampling factor of $N_r \geq 1$ so that sampling occurs every T_s/N_r seconds T_s being the symbol period, we define vectors of N_r samples in the k-th symbol period as

$$\mathbf{y}_k^{(j)} = \begin{pmatrix} y_{k,0}^{(j)} \\ y_{k,1}^{(j)} \\ \vdots \\ y_{k,N_r-1}^{(j)} \end{pmatrix}, \mathfrak{H}_{k,l}^{(j,i)} = \begin{pmatrix} h_{k,l,0}^{(j,i)} \\ h_{k,l,1}^{(j,i)} \\ \vdots \\ h_{k,l,N_r-1}^{(j,i)} \end{pmatrix}, \mathbf{n}_k^{(j)} = \begin{pmatrix} n_{k,0}^{(j)} \\ n_{k,1}^{(j)} \\ \vdots \\ n_{k,N_r-1}^{(j)} \end{pmatrix}. \tag{5.2}$$

From (5.1), we may then write the oversampled (vector) form of the signal in the k-th symbol interval as

[1] Assumed to be the convolution of the transmit pulse shape and physical channel response.

$$\mathbf{y}_k^{(j)} = \sum_{i=1}^{T} \sum_{l=0}^{L-1} d_{k-l}^{(i)} \mathfrak{H}_{k,l}^{(j,i)} + \mathbf{n}_k^{(j)} \;\; ; \;\; j = 1, 2, \cdots, R. \tag{5.3}$$

This signal may then be expressed in a compact matrix-vector form [32] as

$$\mathbf{y}_k = \sum_{l=0}^{L-1} \mathbf{H}_{k,l} \mathfrak{d}_{k-l} + \mathbf{n}_k \tag{5.4}$$

where

$$\mathbf{y}_k = \begin{pmatrix} \mathbf{y}_k^{(1)} \\ \mathbf{y}_k^{(2)} \\ \vdots \\ \mathbf{y}_k^{(R)} \end{pmatrix}, \mathfrak{d}_k = \begin{pmatrix} d_k^{(1)} \\ d_k^{(2)} \\ \vdots \\ d_k^{(T)} \end{pmatrix}, \mathbf{n}_k = \begin{pmatrix} \mathbf{n}_k^{(1)} \\ \mathbf{n}_k^{(2)} \\ \vdots \\ \mathbf{n}_k^{(R)} \end{pmatrix} \tag{5.5}$$

and where we define the RN_r x T channel matrix-taps

$$\mathbf{H}_{k,l} = \begin{pmatrix} \mathfrak{H}_{k,l}^{(1,1)} & \cdots & \mathfrak{H}_{k,l}^{(1,T)} \\ \vdots & \ddots & \vdots \\ \mathfrak{H}_{k,l}^{(R,1)} & \cdots & \mathfrak{H}_{k,l}^{(R,T)} \end{pmatrix}; \quad l = 0, 1, 2, \cdots, L-1. \tag{5.6}$$

Over a block of N_f symbol periods (corresponding to the length of the DFE feedforward filter), the received signal vectors $\mathbf{y}_k$ of (5.4) can be written in matrix form as

$$\begin{pmatrix} \mathbf{y}_{k+N_f-1} \\ \mathbf{y}_{k+N_f-2} \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{k+N_f-1,0} & \cdots & \mathbf{H}_{k+N_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+N_f-2,0} & \cdots & \mathbf{H}_{k+N_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix} \begin{pmatrix} \mathfrak{d}_{k+N_f-1} \\ \mathfrak{d}_{k+N_f-2} \\ \vdots \\ \mathfrak{d}_{k-L+1} \end{pmatrix} + \begin{pmatrix} \mathbf{n}_{k+N_f-1} \\ \mathbf{n}_{k+N_f-2} \\ \vdots \\ \mathbf{n}_k \end{pmatrix} \tag{5.7}$$

This may be expressed in the compact form,

$$\mathbf{y}_{k+N_f-1:k} = \mathbf{C}\mathfrak{d}_{k+N_f-1:k-L+1} + \mathbf{n}_{k+N_f-1:k} \tag{5.8}$$

where $\mathbf{C}$ is the convolution matrix,

$$\mathbf{C} = \begin{pmatrix} \mathbf{H}_{k+N_f-1,0} & \cdots & \mathbf{H}_{k+N_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+N_f-2,0} & \cdots & \mathbf{H}_{k+N_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix} \tag{5.9}$$

Corresponding to this is the block of $(N_f + L - 1)$ input symbol vectors $\mathfrak{d}_{k+N_f-1:k-L+1}$ consisting of $(L$ - 1) past symbol vectors $\mathfrak{d}_{k-1:k-L+1}$ and $(N_f - 1)$ future symbol vectors $\mathfrak{d}_{k+N_f-1:k+1}$ that are yet to be detected. The feedback filter utilizes a subset $\mathfrak{d}_{k-1:k-N_b}$ of previously detected symbol vectors (assumed correct) to cancel their interfering effect on the current symbol vector $\mathfrak{d}_k$.

We define the $T(N_f + L - 1)$ x $T(N_f + L - 1)$ input auto-correlation matrix

$$\begin{aligned}\mathbf{R}_{\mathfrak{d}\mathfrak{d}} &= E\{\mathfrak{d}_{k+N_f-1:k-L+1}\mathfrak{d}^H_{k+N_f-1:k-L+1}\} \\ &= \sigma_d^2\mathbf{I}_{T(N_f+L-1)}\end{aligned} \tag{5.10}$$

where $\mathbf{I}_{T(N_f+L-1)}$ is an identity matrix of size $T(N_f+L-1)$ and σ_d^2 is the variance of the transmitted data symbol. Similarly we define the (RN_f) x (RN_f) noise auto-correlation matrix

$$\begin{aligned}\mathbf{R}_{nn} &= E\{\mathbf{n}_{k+N_f-1:k}\mathbf{n}^H_{k+N_f-1:k}\} \\ &= \sigma_n^2\mathbf{I}_{RN_f}\end{aligned} \tag{5.11}$$

where $\mathbf{I}_{RN_f}$ is an identity matrix of size RN_f and σ_n^2 is the AWGN noise variance.

The input-output cross-correlation and output auto-correlation matrices needed for the calculation of the DFE tap coefficients can then be written in terms of $\mathbf{R}_{\mathfrak{d}\mathfrak{d}}$, $\mathbf{R}_{nn}$ and $\mathbf{C}$ as

$$\mathbf{R}_{\mathfrak{d}y} = E\{\mathfrak{d}_{k+N_f-1:k-L+1}\mathbf{y}^H_{k+N_f-1:k}\} = \mathbf{R}_{\mathfrak{d}\mathfrak{d}}\mathbf{C}^H \tag{5.12}$$

and

$$\begin{aligned}\mathbf{R}_{yy} &= E\{\mathbf{y}_{k+N_f-1:k}\mathbf{y}^H_{k+N_f-1:k}\} \\ &= \mathbf{C}\mathbf{R}_{\mathfrak{d}\mathfrak{d}}\mathbf{C}^H + \mathbf{R}_{nn}.\end{aligned} \tag{5.13}$$

5.3 THE VECTOR DFE

We assume that the DFE contains N_f feedforward filter matrix taps, $\mathbf{F}_k$, and N_b feedback filter matrix taps, $\mathbf{B}_k$, as shown in Fig. 5.1. Each of the feedforward taps consists of $\frac{T_s}{N_r}$-spaced taps while each of the feedback taps is T_s-spaced. These matrix tap coefficients are optimized jointly based on the MMSE performance

criterion. Following [31], we describe the design of the optimum MMSE vector DFE in detail.

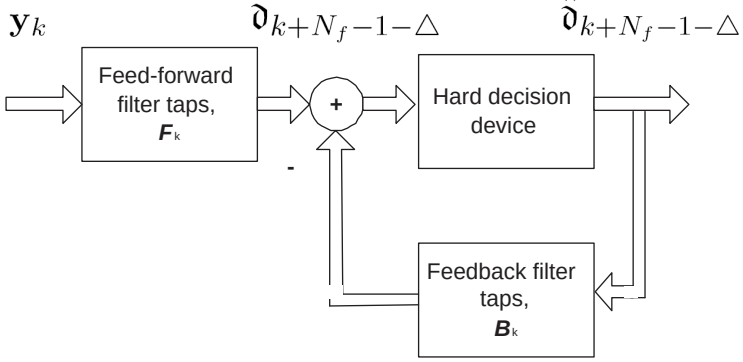

Figure 5.1 A vector MMSE DFE.

The vector DFE consists of a feedforward filter matrix

$$\mathbf{F}_k^H = \left[\begin{array}{cccc} \mathbf{F}_{k,0}^H & \mathbf{F}_{k,1}^H & \cdots & \mathbf{F}_{k,N_f-1}^H \end{array} \right] \tag{5.14}$$

with N_f matrix taps each of size (RN_r x T) and a feedback filter matrix

$$\mathbf{B}_k^H = \left[\begin{array}{ccc} \mathbf{B}_{k,1}^H & \cdots & \mathbf{B}_{k,N_b}^H \end{array} \right] \tag{5.15}$$

with N_b matrix taps each of size (T x T).

For analytical convenience, we define an *extended* T x $T(N_f+L-1)$ feedback filter matrix

$$\tilde{\mathbf{B}}_k^H = \left[\begin{array}{ccc} \mathbf{0}_{T,T\triangle} & \mathbf{I}_T & \mathbf{B}_k^H \end{array} \right], \tag{5.16}$$

that corresponds to the symbol vectors $\mathfrak{d}_{k+N_f-1:k-L+1}$ of (5.8). Note that $\triangle$ is the decision delay in a causal realization of the DFE that satisfies the condition $(\triangle+N_b+1) = (N_f+L-1)$. In general, for ISI cancellation, we require $N_b \geq L-1$. For the purpose of modelling, we assume here that $N_b = L-1$ so that $\triangle = N_f-1$ [31].

Referring to Fig. 5.1, the vector DFE's error vector at time k, assuming

correct past decisions, is given by

$$\begin{aligned}
\mathbf{E}_k &= \mathfrak{d}_{k+N_f-1-\triangle} - \tilde{\mathfrak{d}}_{k+N_f-1-\triangle} \\
&= \mathfrak{d}_{k+N_f-1-\triangle} - \textstyle\sum_{f=0}^{N_f-1} \mathbf{F}_{k,f}^H \mathbf{y}_{k+N_f-1-f} + \sum_{b=1}^{N_b} \mathbf{B}_{k,b}^H \mathfrak{d}_{k+N_f-1-\triangle-b} \\
&= \begin{bmatrix} \mathbf{0}_{T,T\triangle} & \mathbf{I}_{T,T} & \mathbf{B}_{k,1}^H & \cdots & \mathbf{B}_{k,N_b}^H \end{bmatrix} \mathfrak{d}_{k+N_f-1:k-L+1} \\
&\quad - \begin{bmatrix} \mathbf{F}_{k,0}^H & \cdots & \mathbf{F}_{k,N_f-1}^H \end{bmatrix} \mathbf{y}_{k+N_f-1:k} \\
&= \tilde{\mathbf{B}}_k^H \mathfrak{d}_{k+N_f-1:k-L+1} - \mathbf{F}_k^H \mathbf{y}_{k+N_f-1:k}.
\end{aligned} \tag{5.17}$$

The corresponding error auto-correlation matrix may then be written as

$$\begin{aligned}
\mathbf{R}_{ee} &= E[\mathbf{E}_k^H \mathbf{E}_k] \\
&= \tilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}\mathfrak{d}} \tilde{\mathbf{B}}_k - \tilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}y} \mathbf{F}_k - \mathbf{F}_k^H \mathbf{R}_{y\mathfrak{d}} \tilde{\mathbf{B}}_k + \mathbf{F}_k^H \mathbf{R}_{yy} \mathbf{F}_k \\
&= \tilde{\mathbf{B}}_k^H \left(\mathbf{R}_{\mathfrak{d}\mathfrak{d}} - \mathbf{R}_{\mathfrak{d}y} \mathbf{R}_{yy}^{-1} \mathbf{R}_{y\mathfrak{d}} \right) \tilde{\mathbf{B}}_k + \left(\mathbf{F}_k^H - \tilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}y} \mathbf{R}_{yy}^{-1} \right) \mathbf{R}_{yy} \left(\mathbf{F}_k^H - \tilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}y} \mathbf{R}_{yy}^{-1} \right)^H \\
&= \tilde{\mathbf{B}}_k^H \mathbf{R}^{\perp} \tilde{\mathbf{B}}_k + \mathbf{W}_k^H \mathbf{R}_{yy} \mathbf{W}_k
\end{aligned} \tag{5.18}$$

From (5.17) and applying the orthogonality principle using the least squares approach, which states that $E[\mathbf{E}_k \mathbf{y}_{k+N_f-1:k}^H] = 0$, the feedforward matrix filter that achieves optimal performance for a given feedback matrix filter $\tilde{\mathbf{B}}_k$ is given by

$$\mathbf{F}_k^H = \widetilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}y} \mathbf{R}_{yy}^{-1}. \tag{5.19}$$

Using this optimum feedforward matrix filter of (5.19) reduces the second term on the r.h.s of (5.18) to zero which then simplifies the equation to $\mathbf{R}_{ee} = \tilde{\mathbf{B}}_k^H \mathbf{R}^{\perp} \tilde{\mathbf{B}}_k$.

Employing equations (5.10) to (5.13) together with matrix inversion lemma[2],

[2]Given $A = B^{-1} + CD^{-1}C^H$, the inverse of A is given by $A^{-1} = B - BC(D + C^H BC)^{-1} C^H B$. For more details see [49].

the error auto-correlation matrix in (5.18) may then be written as

$$\begin{aligned}
\mathbf{R}_{ee} &= \tilde{\mathbf{B}}_k^H \mathbf{R}^{\perp} \tilde{\mathbf{B}}_k \\
&= \tilde{\mathbf{B}}_k^H (\mathbf{R}_{\mathfrak{d}\mathfrak{d}} - \mathbf{R}_{\mathfrak{d}y} \mathbf{R}_{yy}^{-1} \mathbf{R}_{y\mathfrak{d}}) \tilde{\mathbf{B}}_k \\
&= \tilde{\mathbf{B}}_k^H (\mathbf{R}_{\mathfrak{d}\mathfrak{d}} - \mathbf{R}_{\mathfrak{d}\mathfrak{d}} \mathbf{C}^H (\mathbf{C}\mathbf{R}_{\mathfrak{d}\mathfrak{d}}\mathbf{C}^H + \mathbf{R}_{nn})^{-1} \mathbf{C}\mathbf{R}_{\mathfrak{d}\mathfrak{d}}) \tilde{\mathbf{B}}_k \\
&= \tilde{\mathbf{B}}_k^H (\mathbf{R}_{\mathfrak{d}\mathfrak{d}}^{-1} + \mathbf{C}^H \mathbf{R}_{nn}^{-1} \mathbf{C})^{-1} \tilde{\mathbf{B}}_k \\
&= \tilde{\mathbf{B}}_k^H \mathbf{R}^{-1} \tilde{\mathbf{B}}_k
\end{aligned} \tag{5.20}$$

where

$$\mathbf{R} = \mathbf{R}_{\mathfrak{d}\mathfrak{d}}^{-1} + \mathbf{C}^H \mathbf{R}_{nn}^{-1} \mathbf{C}. \tag{5.21}$$

To calculate the optimal feedback matrix taps, we partition $\mathbf{R}$ into the sub-matrix form [31],

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12}^H & \mathbf{R}_{22} \end{pmatrix} \tag{5.22}$$

where $\mathbf{R}_{11}$ is the $T(\triangle + 1)$ x $T(\triangle + 1)$ upper left sub-matrix. We further define a matrix

$$\mathbf{G}^t = \begin{bmatrix} \mathbf{0}_{T,T\triangle} & \mathbf{I}_T \end{bmatrix} \tag{5.23}$$

and from (5.22) and (5.23), we obtain

$$\widetilde{\mathbf{B}}_{\mathbf{k}} = \begin{bmatrix} \mathbf{I}_{T(\triangle+1)} \\ \mathbf{R}_{12}^H \mathbf{R}_{11}^{-1} \end{bmatrix} \mathbf{G} = \begin{bmatrix} \mathbf{0}_{T,T\triangle} \\ \mathbf{I}_T \\ \mathbf{B}_k \end{bmatrix} \tag{5.24}$$

where $\widetilde{\mathbf{B}}_{\mathbf{k}}$ is the *extended* feedback matrix containing $\mathbf{B}_k$ as the optimal feedback matrix tap coefficients. The error auto-correlation matrix of (5.20) can then be calculated and the optimal $\triangle$ determined such that the trace of $\mathbf{R}_{ee}$ is minimized

[31]. There is no closed-form solution [55] to the $\triangle$ that minimizes the trace of $\mathbf{R}_{ee}$, and it must be found by numerical search. However, as we have assumed that the number of feedback matrix-taps is $N_b = L - 1$, the delay $\triangle$ is fixed at $N_f - 1$ which is found to be optimal for most practical channels [31]. The MMSE feedforward matrix tap coefficients are calculated as in (5.19) using $\widetilde{\mathbf{B}}_k^H$.

5.4 CHANNEL ESTIMATION

The above development of the DFE is based on the assumption that the channel convolution matrix $\mathbf{C}$ is available. In reality this matrix is obtained using channel estimates. Here we employ the VGRLS/VGLMS channel estimator to provide an estimate $\hat{\mathbf{C}}$ of the channel matrix $\mathbf{C}$.

With channel estimation, at each time instant k, we formulate the RN_rN_f x $T(N_f + L - 1)$ block pre-windowed channel convolution matrix $\hat{\mathbf{C}}$, the estimate of (5.9), where $\hat{\mathbf{H}}_{b,m}$ for $b = \{k + N_f - 1, \cdots, k\}$ and $m = \{0, 1, \cdots, L - 1\}$, are the estimates of the (RN_r x T) channel matrices, $\mathbf{H}_{k,l}$ of (5.6). Using the estimate $\hat{\mathbf{C}}$ in place of $\mathbf{C}$, the feedback matrix tap coefficients of the DFE are estimated following the steps described in (5.21) to (5.24), and the feedforward matrix tap as in (5.19).

5.5 THE INTEGRATED RECEIVER

The overall receiver structure is shown in Fig. 5.2 where the channel estimator provides vector estimates, $\{\hat{h}_{k,l,n}^{(j,i)}\}_{l=0}^{l=L-1}$, of the channel responses, $\{h_{k,l,n}^{(j,i)}\}_{l=0}^{l=L-1}$, for the adaptive equalization of the received signal streams, $\{y_{k,n}^{(j)}\}$, for $j = 1, 2, \cdots, R$, $i = 1, 2, \cdots, T$ and $n = 0, 1, \cdots, N_r - 1$. The transmitted frame consists of L_t training symbols and L_d data symbols. Initially the receiver operates in training mode where only the estimator is operating and the training sequence is used to obtain an initial channel estimate. Following this, the receiver operates in decision-directed mode during the L_d-symbol data transmission period, where the estimator and equalizer work in tandem.

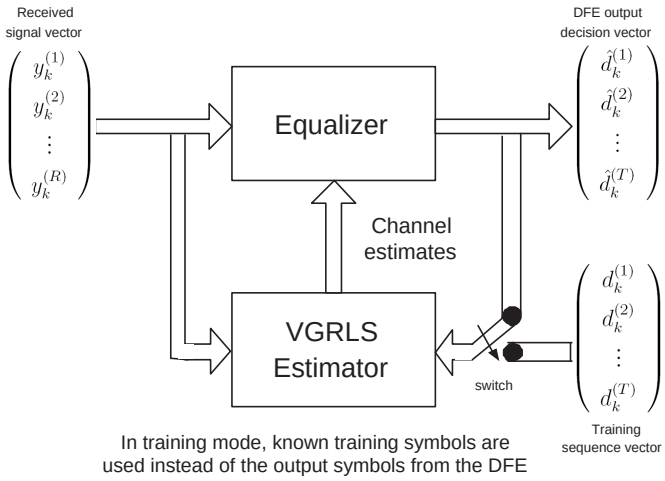

Figure 5.2 The overall receiver structure in decision-directed mode where the vector DFE and channel estimator work in tandem. Note initially when in training mode, the estimator operates alone using the known training symbols instead of the output symbols from the DFE.

5.5.1 Training Mode

In this mode, only the channel estimator is operated using a training sequence of length L_t, according to the following:

VGRLS Algorithm

Step 1: Initiate the VGRLS algorithm with an all-zero estimated channel vector, $\hat{\mathbf{h}}_{1/0}$, and 'intermediate' matrix, $\mathbf{P}_{1/0} = \delta^{-1}\mathbf{I}_{RN_rTLP}$, where δ is a small positive real constant, $\mathbf{I}_{RN_rTLP}$ is an identity matrix with dimension of RN_rTLP x RN_rTLP. Using the observation vector $\mathbf{y}_k$, compute the Kalman gain (3.82), update the 'intermediate' matrix (3.83), and update the estimated channel vector (3.84).

Step 2: Compute the one-step predicted channel vector (3.85) and one-step predicted 'intermediate' matrix (3.86).

Step 3: With every subsequent received observation vector, $\mathbf{y}_k$, $k \geq 2$, repeat steps 2 and 3 until the end of the training sequence.

VGLMS Algorithm

Step 1: Compute the offline recursion for $\hat{\mathbf{P}}_{k/k}$ according to (4.16), (4.15) and (4.14).

Step 2: Compute the one-step predicted channel vector (4.18).

Step 3: With every subsequent received observation vector, $\mathbf{y}_k$, $k \geq 2$, repeat steps 2 and 3 until the end of the training sequence.

5.5.2 Decision-directed Mode

In this mode, the channel estimator and the vector DFE are operated together during data transmission. Initially, channel estimates obtained from the training mode are used to calculate the DFE filter coefficients and equalize the received signals. The outputs from the DFE are then used by the channel estimator to provide the next channel estimates.

However, due to the equalizer decision delay of $\triangle = N_f - 1$ symbols with reference to the equalizer input, a time-lag is introduced where at time k-1 the output symbols from the DFE are delayed by $\triangle$ symbol periods. Thus, the output of the decision device is the estimated symbol vector $\hat{\mathfrak{d}}_{k-\triangle-1} = \left\{\hat{d}^{(1)}_{k-\triangle-1}, \hat{d}^{(2)}_{k-\triangle-1}, \cdots, \hat{d}^{(T)}_{k-\triangle-1}\right\}^t$ [3]. This is fed to the channel estimator in place of the training symbols to provide the next channel estimate vector at time k. Using the DFE decision vectors $\hat{\mathfrak{d}}_{k-\triangle-1}, \cdots, \hat{\mathfrak{d}}_{k-\triangle-L}$, the received vector $\mathbf{y}_{k-\triangle-1}$ and P previously estimated channel vectors, the channel estimator produces $\hat{\mathbf{h}}_{k-\triangle}$ [4]. To calculate the vector DFE at time k, the N_f most recent estimated channel vectors are needed. Up to time $k-\triangle$, the channel estimates are available from the channel estimator and the last $\triangle$ channel vectors need to be predicted. A simple method is to assume that the channel remains constant over $\triangle$ time symbols so that $\hat{\mathbf{h}}_k = \hat{\mathbf{h}}_{k-1} = \cdots = \hat{\mathbf{h}}_{k-\triangle}$ where $\hat{\mathbf{h}}_{k-\triangle}$ is available from the estimator. However, this strictly speaking applies only to a very slowly fading channel.

[3] For convenience of illustration we shift the time index of the DFE in this section from $k + N_f - 1 : k$ to $k : k - N_f + 1$ so that the output of the decision device at time k is indexed as $\hat{\mathfrak{d}}_{k-\triangle}$ instead of $\hat{\mathfrak{d}}_{k+N_f-1-\triangle}$.

[4] For brevity we simplify the notation $\hat{\mathbf{h}}_{k-\triangle/k-\triangle-1}$ to $\hat{\mathbf{h}}_{k-\triangle}$.

As an alternative, we employ a polynomial prediction module similar to that of [25] for predicting the $\triangle$ channel vectors. Since the underlying structure of the channel estimator uses a t-power series expansion [23] for modeling the channel fading process as an N-th order polynomial series, we already have the polynomial-based state transition matrix

$$\mathbf{U} = \begin{pmatrix} \mathfrak{U}_1 & \mathfrak{U}_2 & \cdots & \mathfrak{U}_{P-1} & \mathfrak{U}_P \\ \mathbf{I}_{RN_rTL(P-1)} & & & & \mathbf{0}_{RN_rTL(P-1),RN_rTL} \end{pmatrix} \tag{5.25}$$

where $\mathfrak{U}_p = a_p \mathbf{I}_{RN_rTL}$ and a_p for $p = 1, 2, \cdots, P$ are the polynomial coefficients. It is, therefore, straight forward to compute the predicted channel vectors as

$$\begin{aligned} \hat{\mathbf{h}}_{k-\triangle+1} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle} \\ \hat{\mathbf{h}}_{k-\triangle+2} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle+1} \\ \vdots &= \vdots \\ \hat{\mathbf{h}}_{k} &= \mathbf{U}\hat{\mathbf{h}}_{k-1}. \end{aligned} \tag{5.26}$$

The channel estimates provided by the channel estimator and the channel prediction module are used to compute the feedforward and feedback tap coefficients of the DFE. The received signal vectors are equalized by the DFE and the detected symbol vector is produced at the output of the decision device. This is used at the input of the channel estimator in decision-directed operation. The operations during decision-directed mode may be summarized as:

Step 1: With $\hat{\mathbf{h}}_{k-\triangle-1}$ available at time $k-1$, operate the channel estimator to produce $\hat{\mathbf{h}}_{k-\triangle}$ at time k using the DFE decisions $\hat{\mathfrak{d}}_{k-\triangle-1}, \cdots, \hat{\mathfrak{d}}_{k-\triangle-L}$, the received vector $\mathbf{y}_{k-\triangle-1}$ and P previously estimated channel vectors.

Step 2: Predict the next $\triangle$ channel vectors as in (5.26).

Step 3: Formulate $\hat{\mathbf{C}}$, the estimated convolution matrix of (5.9).

Step 4: Calculate the optimum coefficients of the DFE matrix-tap vectors, $\mathbf{B}_k$, and $\mathbf{F}_k$ of (5.24) and (5.19).

Step 5: Equalize the received vectors and obtain $\hat{\mathfrak{d}}_{k-\triangle}$.

Step 6: At the next time instance, repeat steps 1 to 6.

5.6 PERFORMANCE EVALUATION

We consider the error rate performance of the integrated receiver consisting of the VGRLS/VGLMS channel estimator and the vector DFE [31] operating in decision-directed mode. We assume throughout an uncoded, VBLAST-type [16] of MIMO system. Independent quadrature phase shift keying (QPSK) signal streams are transmitted from each transmit antenna. Each transmitted frame consists of L_t training symbols and L_d data symbols. We assume independent, WSSUS subchannels each with similar fading conditions. The fading processes are assumed to follow Clarke's model [42] and are simulated according to [44]. Each subchannel is assumed to have a uniform power delay profile with $L = 3$ multipath rays. Each of the multipath rays may contain both a non-random (specular) component and a random (diffuse) component as $h_{k,l}^{(j,i)} = h_l^{(nr),(j,i)} + h_{k,l}^{(r),(j,i)}$. The Rice K-factor, reproduced here from Chapter 3,

$$K = \frac{\left|h^{(nr)}\right|^2}{E\left\{\left|h^{(r)}\right|^2\right\}} \tag{5.27}$$

defines the power ratio between the specular and random components.

The estimated channel responses from the VGRLS/VGLMS channel estimator are used to calculate the tap coefficients of the DFE and the outputs of the DFE are used by the estimator to update the estimated channel responses. The simulation at each SNR point is carried out until 200 symbol errors are encountered in each of the streams, and the symbol error rate (SER) is averaged across the T transmitted signal streams. The simulations for the Kalman-filter-based receiver follow the same approach.

The SNR is defined per received antenna. Given that σ_n^2 is the AWGN variance at the input of each receiver, and with both the QPSK signals and the

overall random components of the multipath rays normalized to unit energy, we have

$$SNR = 10log\left(\frac{(1+K)}{\sigma_n^2}\right). \tag{5.28}$$

Unless stated otherwise, we assume that the total transmitted power is restricted to unity and allocated equally between the T antennas.

5.6.1 VGRLS Estimator with DFE

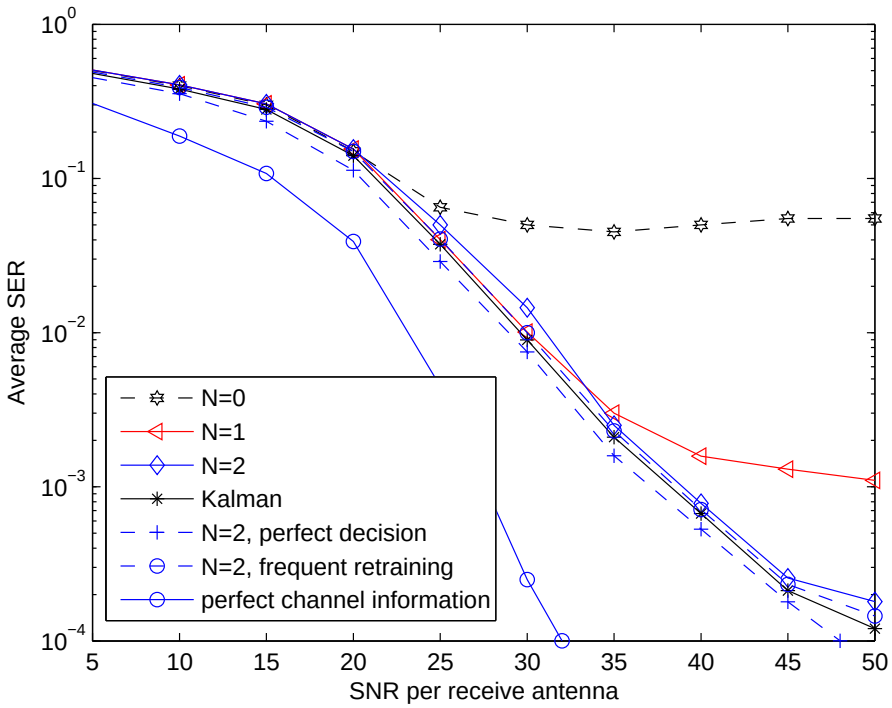

Figure 5.3 Average SER performance of the VGRLS estimator and a Kalman filter for a (2,2) MIMO VBLAST-type system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002, using a MIMO MMSE DFE.

Figs. 5.3 to 5.5 show the average SER performance of independently transmitted QPSK signal streams in (2,2) Rayleigh fading, (4,4) Rayleigh fading and (2,2) Rician fading MIMO systems. A MIMO MMSE-DFE with $N_f = 4$ feedforward filter taps, $N_b = 2$ feedback filter taps and a decision delay of $\triangle = 3$ is used together with a VGRLS channel estimator having a predictor length $P = 3$ and polynomial orders $N = 0, 1, 2$. Transmitted frames with a training

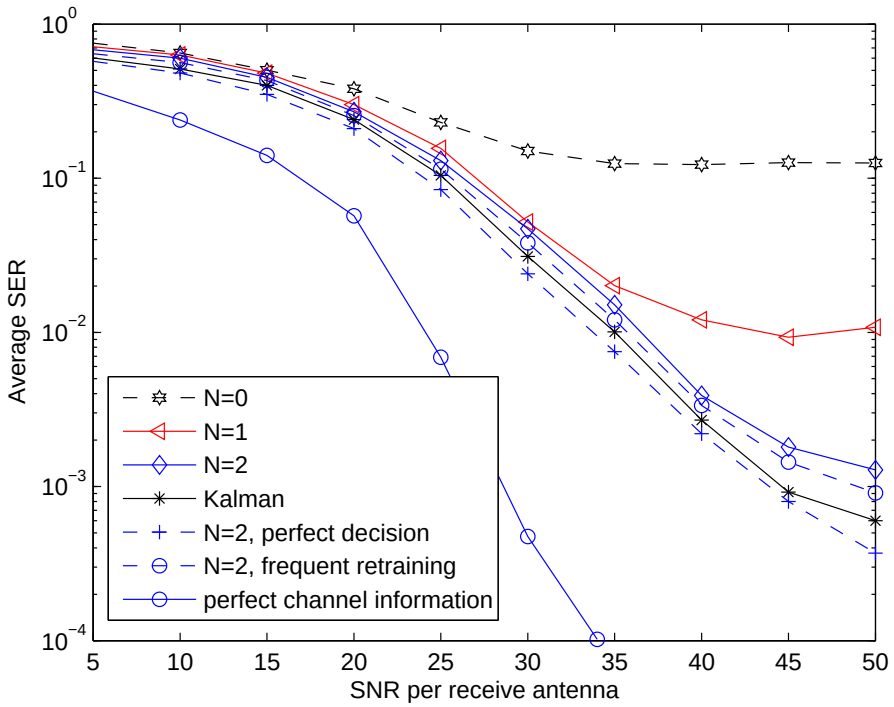

Figure 5.4 Average SER performance of the VGRLS estimator and a Kalman filter for a (4,4) MIMO VBLAST-type system in a Rayleigh fading channel with a normalized fade rate $f_D T_s$ of 0.002, using a MIMO MMSE DFE.

sequence of $L_t = 78$ symbols and a data length of $L_d = 1160$ symbols are used. The result shows that the receiver is able to track the channel over a reasonably long data frame before the next training phase in the subsequent frame. For $N = 2$, we have also simulated a more frequent periodic retraining using $L_t = 78$ symbols in the first frame and $L_t = 26$ in all subsequent frames with $L_d = 116$ in all frames. The results show that more frequent periodic retraining offers only marginal improvement in the error rate performance for the scenarios considered although the improvement appears to be slightly greater for the (4,4) case.

The Rayleigh simulations each have unity transmit power shared equally among the transmitters for a normalized fade rate of 0.002. From Fig. 5.3 and 5.4, we observe that at low to moderate SNR all the channel estimators lead to comparable SER performance regardless of the polynomial order used. However, at high SNR, $N = 2$ performs better and is comparable to that of the Kalman-filter-based receiver, which also starts to exhibit an error floor that is not much different from that seen when using the VGRLS-based receiver. We note that at a

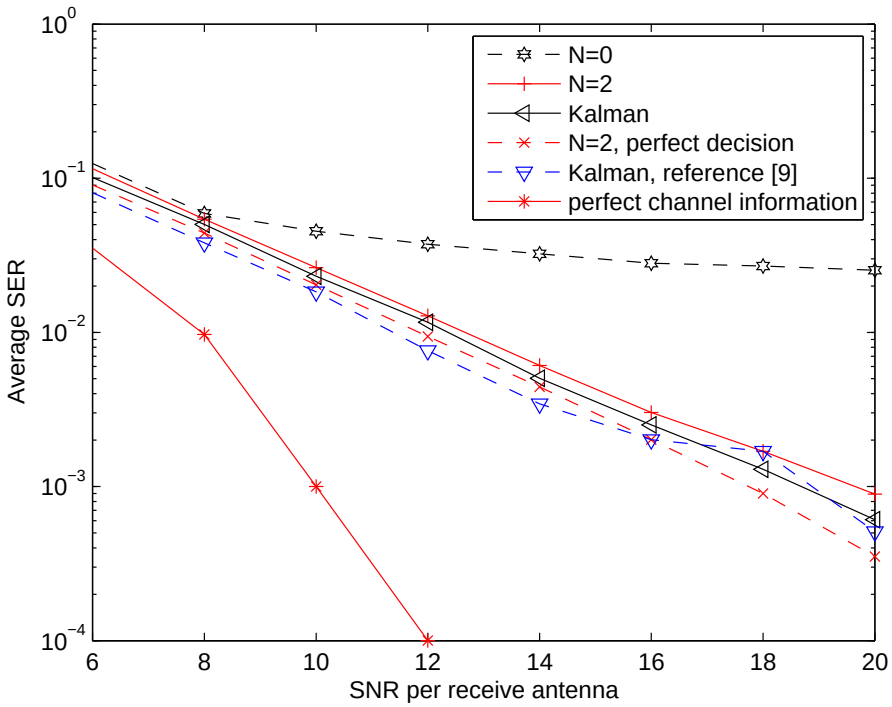

Figure 5.5 Average SER performance of the VGRLS estimator and a Kalman filter for a (2,2) MIMO VBLAST-type system in a Rician fading channel with a normalized fade rate $f_D T_s$ of 0.007 and K=10. The Kalman filter result of [100] is also plotted for reference.

SER of 10^{-3} there is a 5 dB difference in the Kalman filter's performance between a (2,2) and (4,4) system. Results with perfect decision feedback (i.e., using known transmitted symbols) to the estimator and perfect channel information for the DFE tap calculation when using the VGRLS estimator are also included, and show that the $N = 2$ case suffers only modest losses.

We follow the approach of [100] for the simulation of the (2,2) Rician fading. We allocate unit transmit power to each of the transmitters, so the resulting graph has a $log_{10}(T) = 3$dB increase in the SNR per antenna compared to when the total transmit power is limited to unit energy. A Rician K-factor of 10 and a normalized fade rate of 0.007 are used. We also assume the specular components of the fading channel responses to be known when simulating the Kalman filter. This simplifies the simulation by not requiring the state transition matrix to be restructured [97]. However, we have used 3 instead of 2 multipath rays in each sub-channel and this affects the vector DFE's design. We note that the resulting Kalman filter's curve is reasonably close to that of [100].

For the VGRLS estimator, the specular components of the fading channel responses are not known and are estimated together with the diffuse components. From Fig. 5.5, we note that at a SER of 10^{-3} the VGRLS/DFE combination with $N = 2$ is only 1 dB from the Kalman estimator. Results using perfect decision feedback to the estimators and perfect channel information for the DFE tap calculations are also included. They indicate almost a 9 dB loss with respect to the perfect channel information case at a SER of 10^{-3}, but only very modest losses with respect to a Kalman estimator using perfect decision feedback.

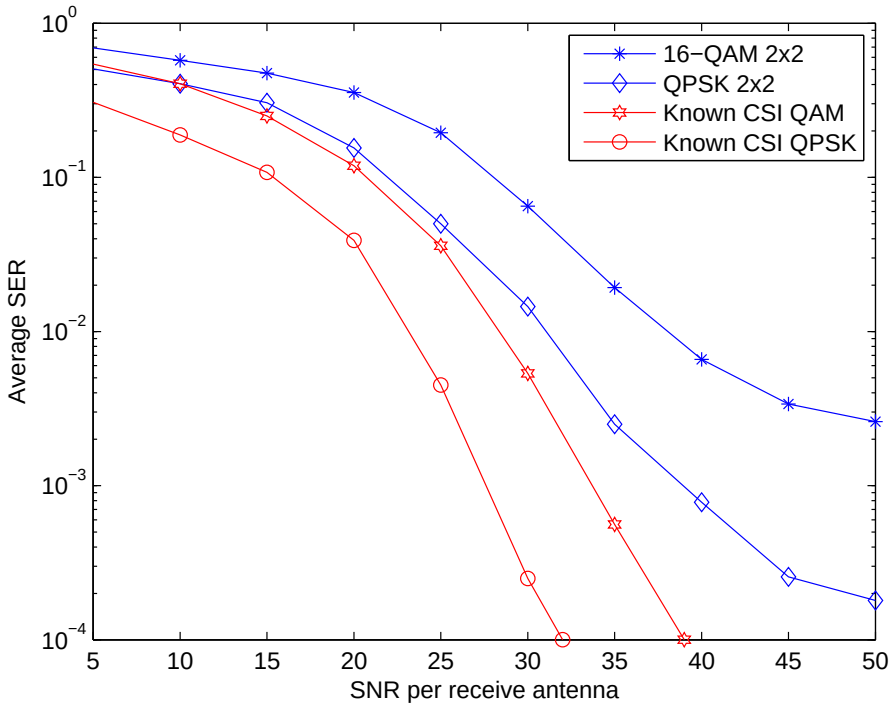

Figure 5.6 Average SER performance of the receiver with QPSK and 16-QAM for a (2,2) Rayleigh fading with a normalized fade rate of 0.002 using a VGRLS estimator with $P = 3$ and $N = 2$.

Besides QPSK, we have also simulated the performance of the receiver using 16-QAM. Figs. 5.6 and 5.7 show the average SER performance of the receiver for a (2,2) and (4,4) Rayleigh fading. The VGRLS estimator has a predictor length $P = 3$ and $N = 2$. We note that the performance of the receiver with a higher order modulation is relatively poor. This is partly due to the fact the higher order modulation is more susceptible to noise and channel estimation error.

In trying to improve the performance of the receiver, we investigate the effect

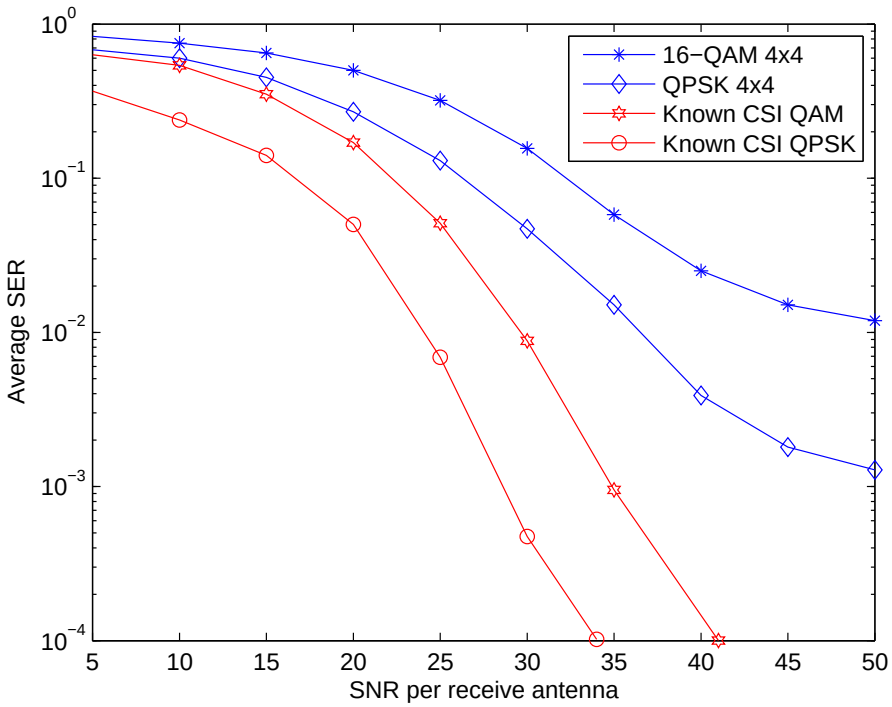

Figure 5.7 Average SER performance of the receiver with QPSK and 16-QAM for a (4,4) Rayleigh fading with a normalized fade rate of 0.002 using a VGRLS estimator with $P = 3$ and $N = 2$.

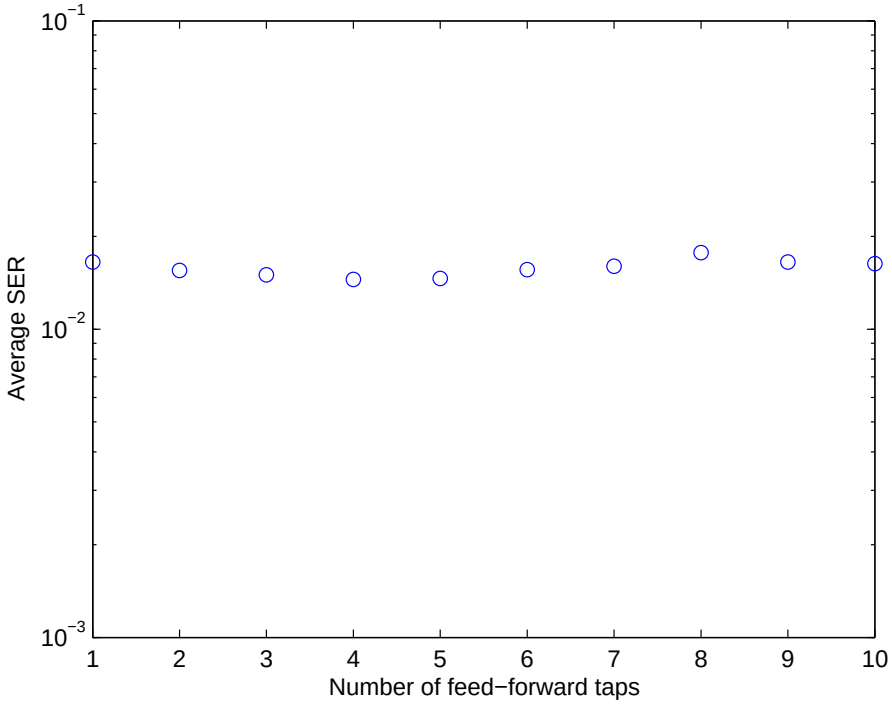

Figure 5.8 Average SER for different number of feedforward DFE taps.

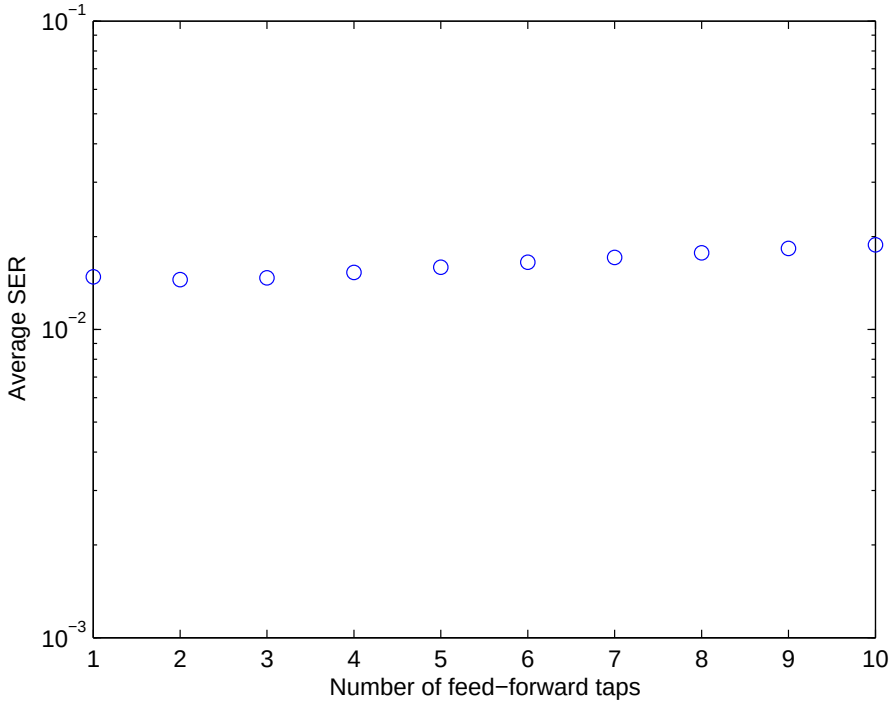

Figure 5.9 Average SER for different number of feedback DFE taps.

of using different numbers of feedforward and feedback taps in the DFE. Fig. 5.8 shows the performance of the DFE with 2 feedback taps and a varying number of feedforward taps while Fig. 5.9 is obtained by varying the number of feedback taps while keeping the feedforward taps at 4. Both plots are obtained at 30dB of SNR and with QPSK modulation. We can observe that in general a larger number of feedforward and/or feedback taps does not necessarily improve the error rate performance. In fact the performance drops (albeit slightly) as the length of either the feedforward or feedback taps increases. This is due to the fact that the equalizer time span is now a lot longer than the actual channel delay spread and when this happens the extra DFE coefficients actually create unwanted interference. Furthermore for a larger number of feedback taps, the effect of error propagation will last longer [36], causing a degradation in performance.

As mentioned previously, a polynomial-based channel prediction module is used to compensate the decision delay, $\triangle$, created by the DFE. In our work, $\triangle = 3$ symbols and during this period, a sufficiently slow fading channel may not vary much so that we can assume the channel coefficients to stay constant.

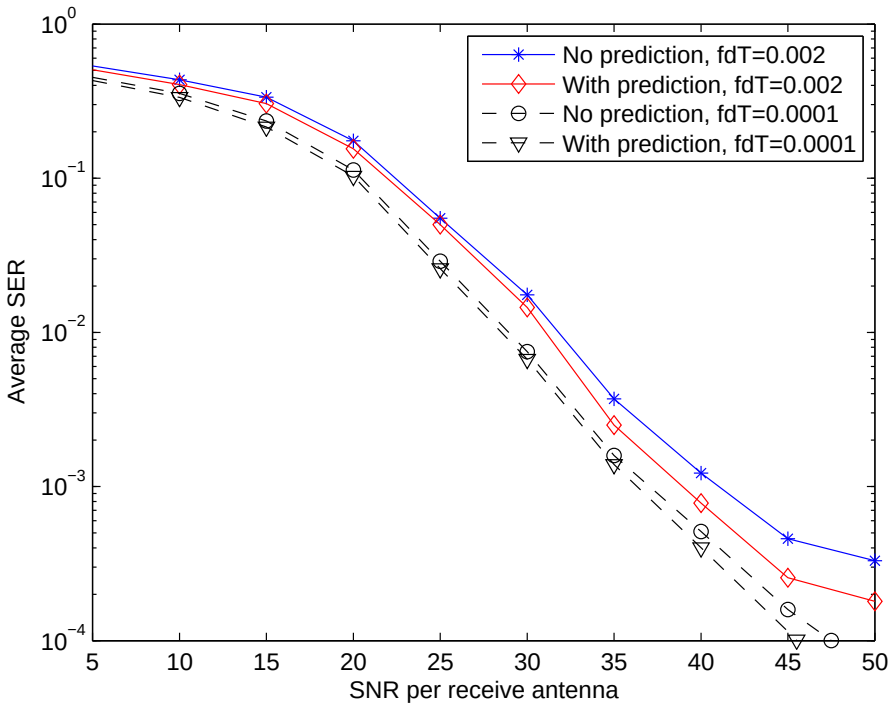

Figure 5.10 The effect of channel prediction module on the performance of the receiver at different normalized fade rates for a (2,2) MIMO system in a Rayleigh fading environment with QPSK modulation. The VGRLS estimator has $P = 3$ and $N = 2$.

The filter coefficients of the DFE are then calculated using the channel estimates obtained $\triangle$ symbols earlier, i.e., the channel estimates are 'out-dated' by $\triangle$ symbols. We investigate the effect of not using updated channel estimates on the performance of the receiver by not using the channel prediction module. Fig. 5.10 shows the performance of the receiver for a (2,2) MIMO system using QPSK modulation in a Rayleigh fading environment with normalized fade rates of 0.002 and 0.0001. The VGRLS estimator has a predictor length $P = 3$ and polynomial order $N = 2$. At a fade rate of 0.002, the performance of the receiver is degraded by approximately 2dB at an SER of 10^{-3}, while at a fade rate of 0.0001, the degradation is smaller at less than 1dB. Hence for a slowly fading channel, the use of the channel prediction module may not be necessary.

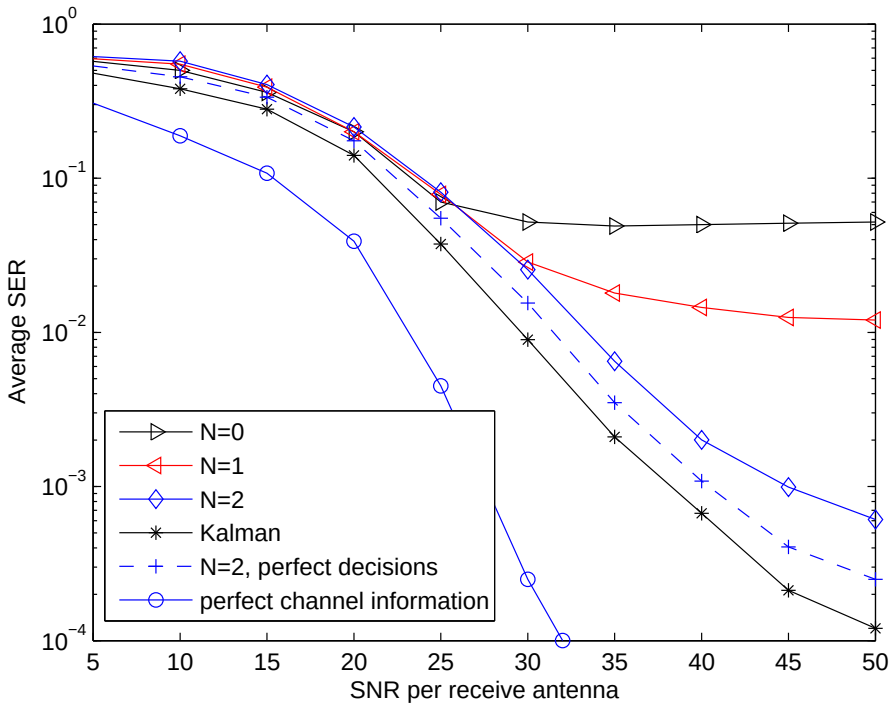

Figure 5.11 Average SER performance of a (2,2) MIMO system using a VGLMS estimator and a Kalman estimator, each operating with a vector DFE in a Rayleigh fading with a normalized fade rate of 0.002. The VGLMS estimator has a predictor length of $P = 3$ and various orders as shown. With perfect decisions for $N = 2$, the transmitted signals instead of the outputs from the DFE, are used by the VGLMS estimator.

5.6.2 VGLMS Estimator with DFE

With similar channel settings, we have also evaluated the error rate performance of a receiver consisting of a VGLMS estimator and a vector DFE [31] operating in a decision-directed mode using the immediate previously detected symbols as feedback. Similar to the previous setting, the DFE consists of 4 forward and 2 feedback matrix-taps with a decision delay of 3 symbols. Each frame consists of $L_t = 200$ training and $L_d = 1160$ data symbols. The simulation at each SNR point is carried out until 200 symbol errors are encountered per streams, and the symbol error rate (SER) is averaged across the T transmitted signal streams.

The average SER performance of a (2,2) system in a Rayleigh fading with a normalized fade rate of 0.002 is shown in Fig. 5.11. We include the performance of a vector DFE using a Kalman estimator, and a vector DFE having perfect channel information for reference. In general the system performance using the

VGLMS estimator is worse than that when using the optimum Kalman estimator. However, it is much better than that using a conventional vector LMS estimator (curve for $N = 0$). For $N = 2$, the estimator's performance is about 7 dB away from that of the Kalman estimator at an SER of 10^{-3}. When perfect decisions instead of the output decisions from the DFE are used by the estimator, the difference is reduced to about 2 - 3 dB at the same SER.

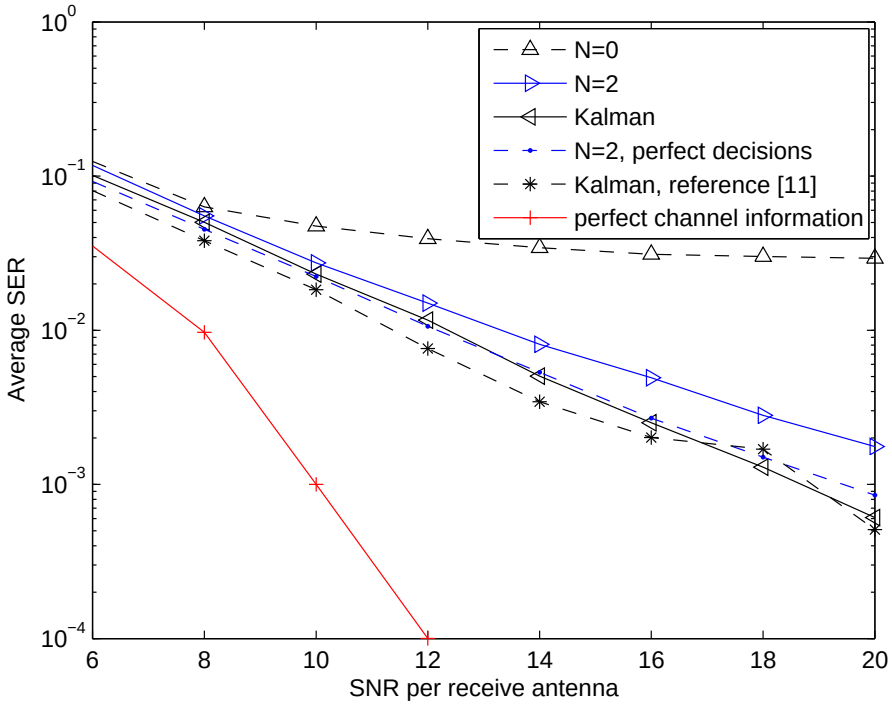

Figure 5.12 A (2,2) MIMO average SER performance of VGLMS and Kalman estimators operating with a vector DFE in a Rician fading with a normalized fade rate of 0.007 and a K-factor of 10. The VGLMS estimator has a predictor length of $P = 3$. The results of [100] is also plotted for reference.

Fig. 5.12 shows the average SER of the above receivers in a Rician fading environment. In order to compare our results with [100] we follow their approach. We allocate unit transmit power to each of the transmitters, so the resulting graph has a $log_{10}(T) = 3$dB increase in the SNR per antenna. A Rician K-factor of 10 and a normalized fade rate of 0.007 is used. We also assume the specular components of the fading channel responses to be known when simulating the Kalman filter. This simplifies the simulation by not requiring the state transition matrix to be restructured [97]. However, we have used 3 instead of 2 multipath rays in each sub-channel and this affects the vector DFE's design. We note that

the resulting Kalman filter's curve is reasonably close to that of [100].

We note that at an SER of 10^{-3} the system using a VGLMS estimator with $N = 2$ is 2 - 3 dB worse than the Kalman estimator-based system. This margin is worse when compared to the 1 dB difference between a VGRLS estimator and a Kalman estimator. However, the degradation is compensated by the reduced complexity in the channel estimation process. The results with perfect decision feedback to the estimator and perfect channel information are also included. They indicate almost a 10 dB loss with respect to perfect channel information at a SER of 10^{-3}, but modest losses with respect to a Kalman estimator.

5.6.3 A Comparison Between the Estimators

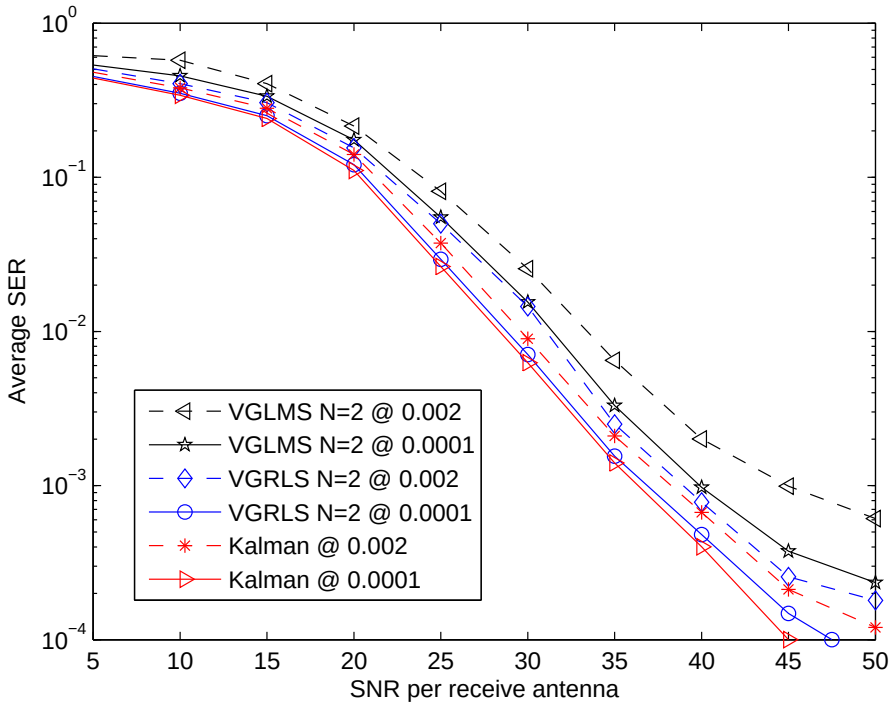

Figure 5.13 Average SER performance of a (2,2) MIMO system using VGLMS, VGRLS and Kalman estimators operating with a vector DFE in a Rayleigh fading with normalized fade rates of 0.002 (dashed lines) and 0.0001 (solid lines) as indicated. The VGLMS and VGRLS estimators each has a predictor length of $P = 3$.

For comparison, we have plotted in Fig. 5.13 the error performance curves of the VGLMS, VGRLS and Kalman estimator operating with the above vector DFE in a (2,2) MIMO system in Rayleigh fading with normalized fade rates of

0.002 and 0.0001. At a fade rate of 0.002, the VGLMS based system is degraded between 6 and 7 dB at an SER of 10^{-3} compared to systems using the VGRLS and Kalman estimators respectively. However, in slower fading, i.e. at a normalized fade rate of 0.0001, the VGLMS estimator performs much better since the channel is easier to track. The SER at 10^{-3} is improved by about 5 dB compared to that at a fade rate of 0.002. This shows that the VGLMS-based receiver offers reasonably good performance in a slowly fading environment.

5.7 SUMMARY

We have developed a symbol-by-symbol based MIMO receiver structure with integrated channel estimation and tracking. A vector MMSE DFE, whose matrix tap coefficients are derived using the channel estimates from the VGRLS/VGLMS estimator, is used as an equalizer. A simple polynomial-based channel prediction module is used to compensate the time-lag created by the decision delay of the equalizer. Using a VGRLS estimator, the resulting symbol error rate performance in Rician and Rayleigh fading channels is shown to be within 1 - 3 dB of that obtained using an optimum Kalman-based estimator. For a slowly fading channel, the performance penalty for not using updated channel estimates is found to be negligible and therefore the channel prediction module may in practice be omitted. We have also demonstrated that the VGLMS estimator is able to offer sufficiently good performance in slow Rayleigh fading or Rician fading with a strong mean component, where the performance is only 2 to 4 dB worse than that of a Kalman-estimator-based system.

Chapter 6

INTEGRATED SEQUENCE-BASED RECEIVER

6.1 INTRODUCTION

Maximum likelihood sequence estimation (MLSE) using the Viterbi algorithm (VA) has been shown to be the optimum equalization method [33],[34] for compensating the effect of ISI in a frequency selective fading environment. However, its complexity increases exponentially with the length of the channel delay spread. In a multiple-input multiple-output (MIMO) system, MLSE requires the use of the vector Viterbi algorithm (VVA) [35] and this adds significant complexity to the receiver.

The complexity of MLSE using the VVA depends on the number of states in the ISI-induced trellis. In a MIMO context, this depends on the modulation constellation size, M, the number of transmit antennas, T, and the channel delay spread, L, according to M^{TL}. For a given value of M, the complexity of MLSE increases exponentially with both T and L. For example, a 4 x 4 MIMO system ($T = 4$) transmitting QPSK signals ($M = 4$) in a fading environment assuming a channel delay spread of 2 symbols ($L = 2$) requires 65536 states. For larger constellations and more transmit antennas, the number of states quickly grows out of hand. Moreover, in some channels the delay spread is significantly longer which further increases complexity. Channel shortening filters that reduce the length of the effective channel impulse response (CIR) [119], reduced state sequence estimation (RSSE) [67] and delayed decision feedback sequence estimation (DDFSE) [66] can be used to reduce the number of states. However, there

is still an exponential dependance on both T and L.

As a result, reduced complexity sequence estimation techniques that offer a linear increase in complexity with T and/or L are therefore highly desired. In [120], a decision feedback MLSE (DFMLSE) scheme is proposed where complexity increases linearly with L, according to LM^T. For the above example, the number of states is then 512. Alternatively, a partitioned Viterbi algorithm (PVA) with a linear increase in complexity with T as TM^L can be employed as proposed in [32]. The total number of states is then only 64 for the above example. Since the PVA results in linearly increasing complexity with T, which is more important in MIMO systems, it is considered in this work.

In [32] the channel fading is assumed to be quasi-static, where the CIR is assumed to remain constant throughout the transmission of a signal frame, but to vary randomly from frame to frame. Training symbols are used at the beginning of each frame to estimate the CIR using the least-squares (LS) technique. The estimates are then used to equalize the remainder of the frame. Channel tracking is not used in [32] and system performance tends to degrade in a continuously time-varying fading environment, particularly for longer frames.

Here, we extend the PVA-based receiver of [32] to cope with continuous fading by incorporating channel estimation and tracking using the VGRLS/VGLMS channel estimator developed earlier. The resulting receiver is an implementable approximation to MLSE in MIMO channels and is among the first to explicitly incorporate dynamic channel estimation in the context of sequence estimation receivers. For ease of illustration, we will describe the channel estimation in the following sections using the VGRLS estimator, and note that the VGLMS estimator can be used readily in place of the VGRLS estimator.

6.2 SIGNAL MODEL

We employ a MIMO signal model in a spatial multiplexing context similar to the one used in Chapter 5. Therefore we will not repeat the development of the signal

model in this section except to recall the appropriate equations in the subsequent sections when necessary.

6.3 THE INTEGRATED RECEIVER

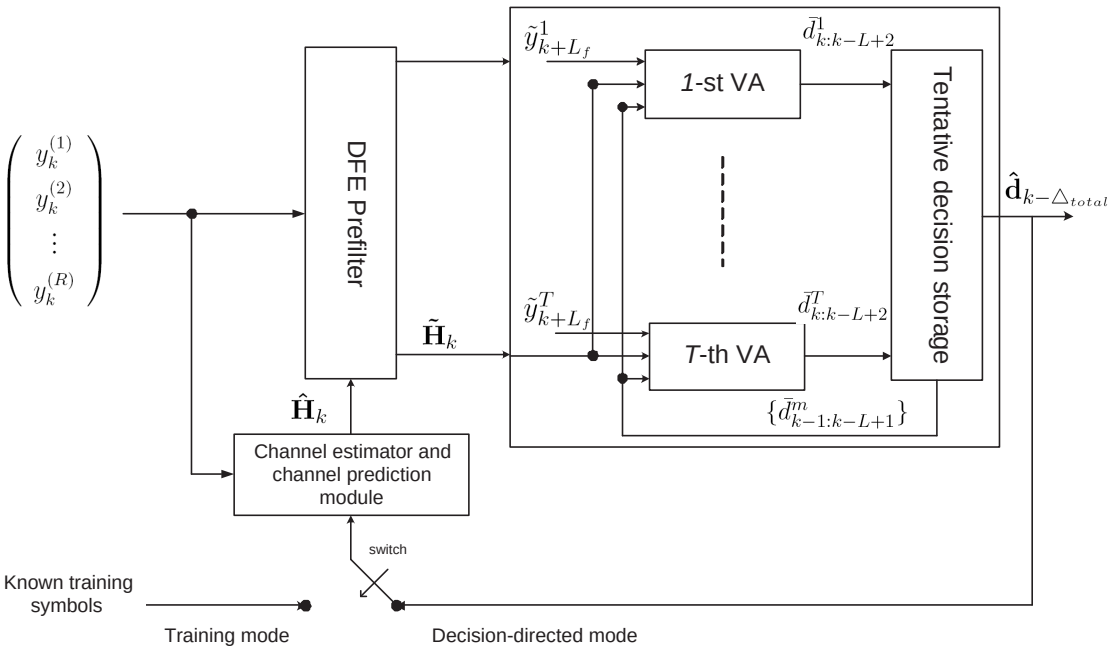

Figure 6.1 The integrated sequence-based receiver using the PVA algorithm with channel estimation and tracking for a continuously frequency selective fading environment.

In this section, we develop the integrated sequence-based receiver using the PVA algorithm with VGRLS channel estimation and tracking. To do this, we replace the non-tracking LS channel estimator of [32] with the VGRLS estimator developed in Chapter 3. The VGRLS estimator tracks the channel variation, so that the channel estimates are time-variant from symbol-to-symbol. There are two aspects that require attention compared to the original structure proposed in [32]: first, the symbol-by-symbol updating of the CIR estimates and the effect on the subsequent PVA operation and second, the effect of the VA and prefilter decision delays on the updating of the CIR estimates. The latter results in the VGRLS estimator producing delayed channel estimates. We consider these in

more detail in the following sections.

6.3.1 Sequence Estimation Based on the PVA

A major component of the PVA is a length L_f prefilter used to provide linear estimates of each of the T transmitted signal streams. Ideally it should be a vector whitened matched filter (WMF) [33]. However, in practice the vector WMF does not always exist [69]. As a result, the feedforward filter of a vector MMSE DFE is used [32]. The benefit of using the DFE prefilter is that it always exists even when the WMF does not. It has been shown to approach WMF performance as the SNR and number of filter taps increases [69]. The prefilter compensates for pre-cursor ISI and decouples the received signal vector into T signal streams. Parallel VAs are then used to process these outputs to obtain equalized data estimates. Tentative decisions are made in each interval and these are exchanged among the parallel VAs. For each transmitted signal stream, feedback terms estimated using the tentative decisions obtained in the previous interval from the other processors are used to cancel 'cross-interferences' effects from the other signal streams.

DFE Prefilter

We describe in this section the derivation of the DFE prefilter. Following [32] we assume that a length L_f fractionally spaced FIR feedforward filter with matrix taps is used as the prefilter, where L_f is the support of the prefilter impulse response in symbols. Each tap is denoted by a T x RN_r matrix $\mathbf{F}_{k,m}$ for $m = 0, 1, \cdots, L_f - 1$. The prefilter matrix-taps can be expressed in a vector as $\mathbf{F}_k = [\mathbf{F}_{k,0}, \cdots, \mathbf{F}_{k,L_f-1}]$. Prefiltering the received signal vector, $\mathbf{y}_k$, of (5.4) gives

$$\begin{aligned}\tilde{\mathbf{y}}_k &= \sum_{m=0}^{L_f-1} \mathbf{F}_{k,m}\mathbf{y}_{k-m} \\ &= \sum_{n=0}^{L_f+L-2} \left(\sum_{m=0}^{L_f-1} \mathbf{F}_{k,m}\mathbf{H}_{k-m,n-m} \right) \mathfrak{d}_{k-n} + \sum_{m=0}^{L_f-1} \mathbf{F}_{k,m}\mathbf{n}_{k-m} \\ &= \sum_{n=0}^{L_f+L-2} \tilde{\mathbf{H}}_{k,n}\mathfrak{d}_{k-n} + \tilde{\mathbf{n}}_k\end{aligned} \tag{6.1}$$

where the prefiltered channel and prefiltered noise are defined as

$$\tilde{\mathbf{H}}_{k,n} = \sum_{m=0}^{L_f-1} \mathbf{F}_{k,m}\mathbf{H}_{k-m,n-m} \tag{6.2}$$

and

$$\tilde{\mathbf{n}}_k = \sum_{m=0}^{L_f-1} \mathbf{F}_{k,m}\mathbf{n}_{k-m} \tag{6.3}$$

respectively.

As shown in Chapter 5, a block of L_f received signal vectors $\mathbf{y}_k$ of (5.4) can be written as

$$\begin{pmatrix} \mathbf{y}_{k+L_f-1} \\ \mathbf{y}_{k+L_f-2} \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{k+L_f-1,0} & \cdots & \mathbf{H}_{k+L_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+L_f-2,0} & \cdots & \mathbf{H}_{k+L_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix} \begin{pmatrix} \mathfrak{d}_{k+L_f-1} \\ \mathfrak{d}_{k+L_f-2} \\ \vdots \\ \mathfrak{d}_{k-L+1} \end{pmatrix} + \begin{pmatrix} \mathbf{n}_{k+L_f-1} \\ \mathbf{n}_{k+L_f-2} \\ \vdots \\ \mathbf{n}_k \end{pmatrix}. \tag{6.4}$$

This may be expressed in the compact form,

$$\mathbf{y}_{k+L_f-1:k} = \mathbf{C}_k \mathfrak{d}_{k+L_f-1:k-L+1} + \mathbf{n}_{k+L_f-1:k} \tag{6.5}$$

where $\mathbf{C}_k$ is the convolution matrix

$$\mathbf{C}_k = \begin{pmatrix} \mathbf{H}_{k+L_f-1,0} & \cdots & \mathbf{H}_{k+L_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+L_f-2,0} & \cdots & \mathbf{H}_{k+L_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix}. \tag{6.6}$$

To facilitate the derivation of the DFE prefilter and the structure of the PVA trellis, we initially assume that the channel responses, i.e. the elements of (6.6), are available. In reality these are obtained through channel estimation which we will deal with later on.

We partition (6.6) into three sections of $T(L_f-1)$, T and $T(L-1)$ columns corresponding to $\mathbf{H}_{k,fut}$, $\mathbf{H}_{k,pres}$ and $\mathbf{H}_{k,past}$. The first $T(L_f-1)$ columns represent the L_fRN_r x $T(L_f-1)$ matrix $\mathbf{H}_{k,fut}$ which represents the filter response from 'future' symbols. The T columns represent the L_fRN_r x T matrix $\mathbf{H}_{k,pres}$ which represent the current transmitted symbols and the last $T(L-1)$ columns represent the L_fRN_r x $T(L-1)$ matrix $\mathbf{H}_{k,past}$ which represent the previously transmitted symbols. Then (6.5) can be written in the form

$$\begin{aligned} \mathbf{y}_{k+L_f-1:k} &= \begin{pmatrix} \mathbf{H}_{k,fut} & \mathbf{H}_{k,pres} & \mathbf{H}_{k,past} \end{pmatrix} \mathfrak{d}_{k+L_f-1:k-L+1} + \mathbf{n}_{k+L_f-1:k} \\ &= \mathbf{H}_{k,fut}\mathfrak{d}_{k+L_f-1:k+1} + \mathbf{H}_{k,pres}\mathfrak{d}_k + \mathbf{H}_{k,past}\mathfrak{d}_{k-1:k-L+1} + \mathbf{n}_{k+L_f-1:k}. \end{aligned} \tag{6.7}$$

In deriving the prefilter, we employ the methodology used in designing a MMSE DFE where the feedforward and feedback filter coefficients are derived and optimized jointly. We assume correct past decisions are available (i.e.,

$\hat{\mathfrak{d}}_{k-1:k-L+1} = \mathfrak{d}_{k-1:k-L+1}$) so that (6.7) can be used to write an FIR vector DFE estimate as

$$\begin{aligned}\hat{\mathfrak{d}}_k &= \mathbf{F}_k \mathbf{y}_{k+L_f-1:k} - \mathbf{F}_k \mathbf{H}_{k,past} \mathfrak{d}_{k-1:k-L+1} \\ &= \mathbf{F}_k \mathbf{H}_{k,fut} \mathfrak{d}_{k+L_f-1:k+1} + \mathbf{F}_k \mathbf{H}_{k,pres} \mathfrak{d}_k + \mathbf{F}_k \mathbf{n}_{k+L_f-1:k} \\ &= [\tilde{\mathbf{H}}_{k,0}, \tilde{\mathbf{H}}_{k,1}, \cdots, \tilde{\mathbf{H}}_{k,L_f-2}] \mathfrak{d}_{k+L_f-1:k+1} + \tilde{\mathbf{H}}_{k,L_f-1} \mathfrak{d}_k + \tilde{\mathbf{n}}_{k+L_f-1:k}.\end{aligned} \tag{6.8}$$

where $\tilde{\mathbf{H}}_k = [\tilde{\mathbf{H}}_{k,0}, \tilde{\mathbf{H}}_{k,1}, \cdots, \tilde{\mathbf{H}}_{k,L_f-1}]$ is the time-variant prefiltered CIR defined by (6.2). The prefiltered noise $\tilde{\mathbf{n}}_{k+L_f-1:k}$ is assumed to be Gaussian and white as in [32]. This is a requirement for the PVA's use of the Viterbi algorithm. We note that in some cases, e.g. in an overloaded system when $T > R$, the noise may not be white. In such a case the colored-noise version of the Viterbi algorithm [121] can be used.

The FIR vector DFE uses the forward filter $\mathbf{F}_k$ that minimizes the mean-square-error (MSE) $E\{||\hat{\mathfrak{d}}_k - \mathfrak{d}_k||^2\}$. It is shown that the MSE is minimized when this filter satisfies [32]

$$([\mathbf{H}_{k,fut}, \mathbf{H}_{k,pres}][\mathbf{H}_{k,fut}, \mathbf{H}_{k,pres}]^H + \sigma_n^2 \mathbf{I}) \mathbf{F}_k^H = \mathbf{H}_{k,pres}. \tag{6.9}$$

Since $([\mathbf{H}_{k,fut}, \mathbf{H}_{k,pres}][\mathbf{H}_{k,fut}, \mathbf{H}_{k,pres}]^H + \sigma_n^2 \mathbf{I})$ is Hermitian and positive definite, (6.9) can be solved efficiently using the Cholesky decomposition.

The advantage of using the DFE prefilter is that the prefiltered channel matrix taps $\tilde{\mathbf{H}}_{k,n}$ for $n = 0, 1, \cdots, L_f - 1$ are T x T matrices independent of the number of receive antennas, R, and oversampling factor, N_r [32]. This means that increasing R or N_r will increase the complexity of solving (6.9) but not the complexity of the Viterbi algorithm. Examination of (6.8) shows that the MSE is minimized when the first $L_f - 1$ prefiltered channel taps $[\tilde{\mathbf{H}}_{k,0}, \tilde{\mathbf{H}}_{k,1}, \cdots, \tilde{\mathbf{H}}_{k,L_f-2}]$ approximate zero-matrices and the tap $\tilde{\mathbf{H}}_{k,L_f-1}$ approximates the identity matrix. With these approximations the τ-th output of the prefilter for $\tau = 1, 2, \cdots, T$ can be written as

$$\begin{aligned}\tilde{y}^{(\tau)}_{k+L_f-1} &= \sum_{m=1}^{T}\sum_{n=0}^{L_f+L-2} \tilde{h}^{(\tau,m)}_{k,n} d^{(m)}_{k+L_f-1-n} + \tilde{n}^{(\tau)}_{k+L_f-1} \\ &\approx \sum_{m=1}^{T}\sum_{n=0}^{L-1} \tilde{h}^{(\tau,m)}_{k,n+L_f-1} d^{(m)}_{k-n} + \tilde{n}^{(\tau)}_{k+L_f-1} \\ &\approx \sum_{n=0}^{L-1} \tilde{h}^{(\tau,\tau)}_{k,n+L_f-1} d^{(\tau)}_{k-n} + \sum_{m\neq\tau}\sum_{n=1}^{L-1} \tilde{h}^{(\tau,m)}_{k,n+L_f-1} d^{(m)}_{k-n} + \tilde{n}^{(\tau)}_{k+L_f-1}.\end{aligned} \tag{6.10}$$

This shows that the τ-th prefilter output value is a function of the most recent L symbols from the τ-th transmitter $d^{(\tau)}_{k:k-L+1}$ and (L - 1) delayed symbols $\{d^{(m)}_{k-1:k-L+1}\}_{m\neq\tau}$ from each of the other transmitters. This is used to generate inputs to the PVA algorithm and the structure of (6.10) is exploited to develop the PVA algorithm.

Trellis Structure

As described in the previous section, the inputs to the PVA algorithm are the prefiltered received signal ($\tilde{\mathbf{y}}_{k+L_f}$) and prefiltered channel responses ($\tilde{\mathbf{H}}_k$). Since there are T outputs from the prefilter, T VAs are employed to process them in parallel [32]. Consider an estimate of the τ-th transmitted symbol sequence $\hat{d}^{(\tau)}_{k:k-L+1}$, for $\tau = 1, 2, \cdots, T$. Divide the estimate into two overlapping sections which define the M^{L-1} "states", $\hat{d}^{(\tau)}_{k-1:k-L+1}$ and $\hat{d}^{(\tau)}_{k:k-L+2}$, where each state corresponds to a possible symbol combination. Let $\mathbf{i}$ indicate a particular previous state $\hat{d}^{(\tau)}_{k-1:k-L+1}$ and let $\mathbf{j}$ indicate a particular current state $\hat{d}^{(\tau)}_{k:k-L+2}$. The branch metric of the VA used for the τ-th trellis at each time k corresponding to the state transition ($\mathbf{i}$, $\mathbf{j}$) is then given by

$$\lambda^{\tau}(\mathbf{i},\mathbf{j},k) = \left| \tilde{y}^{(\tau)}_{k+L_f-1} - \sum_{n=0}^{L-1} \tilde{h}^{(\tau,\tau)}_{k,n+L_f-1} \hat{d}^{(\tau)}_{k-n} - \phi(k,\tau) \right|^2 \tag{6.11}$$

where

$$\phi(k,\tau) = \sum_{m\neq\tau}\sum_{n=1}^{L-1} \tilde{h}_{k,n+L_f-1}^{(\tau,m)} \bar{d}_{k-n}^{(m)} \tag{6.12}$$

is the feedback term estimated using the tentative decisions $\{\bar{d}_{k-n}^{(m)}\}_{m\neq\tau}$ made by the other trellises during the previous symbol time. The summation in (6.11) is an estimate of the τ-th transmitter's contribution to the observed value in $\tilde{y}_{k+L_f-1}^{(\tau)}$.

6.3.2 VGRLS Channel Estimation and Tracking

We integrate the PVA algorithm with both the VGRLS and VGLMS channel estimators. For ease of description, here we use the VGRLS algorithm to describe the channel estimation part of the resulting receiver. Apart from the initialization stage (i.e. the off-line Riccatti computation), we note that the VGLMS estimator can be used readily in place of the VGRLS estimator.

Recall from Chapter 3 that the VGRLS algorithm may be described by the equations

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{d}_k^H(\mathbf{I}_{RN_r} + \mathbf{d}_k\mathbf{P}_{k/k-1}\mathbf{d}_k^H)^{-1} \tag{6.13}$$

$$\mathbf{P}_{k/k} = (\mathbf{I}_{RN_rTLP} - \mathbf{K}_k\mathbf{d}_k)\mathbf{P}_{k/k-1} \tag{6.14}$$

$$\hat{\mathbf{h}}_{k/k} = \hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}). \tag{6.15}$$

$$\hat{\mathbf{h}}_{k+1/k} = \mathbf{U}\hat{\mathbf{h}}_{k/k} \tag{6.16}$$

$$\mathbf{P}_{k+1/k} = \lambda^{-1}\mathbf{U}\mathbf{P}_{k/k}\mathbf{U}^H \tag{6.17}$$

where $\hat{\mathbf{h}}_{k/k-1}$ is the estimate of the channel state vector at time k based on (k-

1) prior received samples, λ is the RLS "forget factor", $\mathbf{K}_k$ is analogous to the Kalman gain vector [49] and $\mathbf{P}_{k/k}$ is the so-called 'intermediate' matrix[1]. More details on the VGRLS estimator are available in Chapter 3.

In decision-directed mode, the VGRLS estimator in the receiver employs the detected symbols[2], $\hat{\mathbf{d}}_k$, from the PVA output in place of $\mathbf{d}_k$, the known training symbol vector. Due to the decision delay, $\triangle$, of the VAs, these are delayed by $\triangle$ symbol periods. Recall from Chapter 5 that as the received signal $\mathbf{y}_k$ passes through the length-L_f DFE prefilter a decision delay of $L_f - 1$ is introduced. Therefore there is a total delay of $\triangle_{total} = L_f + \triangle - 1$ symbols with respect to the input of the receiver as well as the estimator (see Fig. 6.1).

Using these delayed symbols, together with the received vector $\mathbf{y}_{k-\triangle_{total}}$ and P previously estimated channel vectors, the VGRLS estimator produces a delayed channel estimate[3] $\hat{\mathbf{h}}_{k-\triangle_{total}+1}$. However, to calculate the length L_f prefilter taps of the PVA at time k+1, up-to-date estimated channel vectors corresponding to the most recent L_f symbols should be used. The VGRLS estimator provides one of these L_f estimates and the subsequent $L_f - 1$ estimates are still required. A simple method is to assume that the channel remains constant over these L_f symbol periods so that $\hat{\mathbf{h}}_{k-\triangle} = \cdots = \hat{\mathbf{h}}_{k-\triangle_{total}+2} = \hat{\mathbf{h}}_{k-\triangle_{total}+1}$ where $\hat{\mathbf{h}}_{k-\triangle_{total}+1}$ is available from the VGRLS estimator. However, this may apply only to a very slowly fading channel.

An alternative approach is to predict the $L_f - 1$ channel vectors. Here we employ a vector polynomial channel prediction module similar to the one used in Chapter 5 to predict the estimated channel vectors. Since the underlying structure of the VGRLS estimator uses a t-power series expansion [23] to model the channel fading process as an N-th order polynomial series, the polynomial-based state transition matrix $\mathbf{U}$ is available. It is then straight forward to compute the predicted channel estimates, as in Chapter 5, as

[1] $\mathbf{P}_{k/k}$ is the inverse input autocorrelation matrix in a conventional RLS algorithm.

[2] Here the detected symbol vector $\hat{\mathfrak{d}}_k$ is rearranged into $\hat{\mathbf{d}}_k$, a matrix with appropriate dimension as required by the VGRLS estimator. See (3.37) in Chapter 3 for more detail.

[3] For brevity we simplify the notation $\hat{\mathbf{h}}_{k-\triangle_{total}+1/k-\triangle_{total}}$ to $\hat{\mathbf{h}}_{k-\triangle_{total}+1}$.

$$\begin{aligned} \hat{\mathbf{h}}_{k-\triangle_{total}+2} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle_{total}+1} \\ \vdots &= \vdots \\ \hat{\mathbf{h}}_{k-\triangle} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle-1}. \end{aligned} \tag{6.18}$$

The complexity of the VGRLS estimator can be reduced significantly by replacing the online 'Riccatti' computation of (6.14) with an offline pre-computed matrix. This reduces complexity from $\mathcal{O}\big((RN_rTLP)^3\big)$ in the highest order terms to $\mathcal{O}\big((RN_rTLP)^2\big)$ and results in a reduced complexity algorithm known as the VGLMS estimator. We have, therefore, also simulated an integrated receiver employing the VGLMS estimator and PVA algorithm. We note that apart from the offline matrix computation, the VGLMS estimator can be used in place of the VGRLS estimator albeit with some loss in receiver performance.

Upon obtaining new channel estimates at time k, we formulate $\hat{\mathbf{C}}_k$ the estimated convolution matrix of (6.6) and used it to solve for the prefilter coefficients in (6.9). We process a block of L_f received signal vectors to obtain (6.10) and use the channel estimates to obtain the prefiltered channel taps $\tilde{\mathbf{H}}_{k,n}$ for $n = 0, 1, \cdots, L_f - 1$. Using these, we compute the branch metrics for each of the T parallel VAs as in (6.11) and (6.12). After a VA decision delay of $\triangle$, an estimate of the transmitted vector $\hat{\mathbf{d}}_{k-\triangle_{total}}$ is produced[4] as the PVA output.

6.4 RECEIVER OPERATION

The receiver is operated in two modes, a training mode or a decision-directed data transmission mode. Each transmitted frame consists of L_t training symbols followed by L_d data symbols. The receiver operation is described as follows:

6.4.1 Training Mode

Here, only the VGRLS estimator is operated using a training sequence of length L_t, according to the following:

[4]Note that $\triangle_{total} = \triangle + L_f - 1$.

Step 1: Initiate the VGRLS algorithm with an all-zero estimated channel vector, $\hat{\mathbf{h}}_{1/0}$, and 'intermediate' matrix, $\mathbf{P}_{1/0} = \delta^{-1}\mathbf{I}_{RN_rTLP}$, where δ is a small positive real constant and $\mathbf{I}_{RN_rTLP}$ is an identity matrix with dimension, RN_rTLP x RN_rTLP. Using the observation vector $\mathbf{y}_k$, compute the Kalman gain of (6.13), update the 'intermediate' matrix of (6.14), and update the estimated channel vector of (6.15).

Step 2: Compute the one-step predicted channel vector of (6.16) and one-step predicted 'intermediate' matrix of (6.17).

Step 3: With each observation vector, $\mathbf{y}_k$, $k \geq 2$, repeat steps 2 and 3 until the end of the training sequence.

Following training, the receiver switches to decision-directed operation. This procedure is identical to that in Chapter 5.

6.4.2 Decision-directed Mode

In this mode, the DFE prefilter and PVA algorithm are operated in tandem with the VGRLS estimator and the channel prediction module during the L_d data transmission period. In this mode, the feedback terms are calculated using tentative decisions from the VA's. Assuming $\hat{\mathbf{d}}_{k-\triangle_{total}}$ to be the data vector at time k, the receiver operation may be described as:

Step 1: With $\hat{\mathbf{h}}_{k-\triangle_{total}}$ available at time k, operate the VGRLS estimator to produce $\hat{\mathbf{h}}_{k-\triangle_{total}+1}$ at time k+1 using the PVA output vectors $\hat{\mathbf{d}}_{k-\triangle_{total}}, \cdots, \hat{\mathbf{d}}_{k-\triangle_{total}-L+1}$, received vector $\mathbf{y}_{k-\triangle_{total}}$ and P previously estimated channel vectors.

Step 2: Predict the next $L_f - 1$ channel vectors as in (6.18). Recall that each estimated vector $\hat{\mathbf{h}}$ follows the structure of (3.43) with P vector elements, $\hat{\mathfrak{h}}$, where each element or sub-vector has RN_rTL components as shown in (3.7).

Step 3: Following (3.2) and (3.6), the estimated convolution matrix, $\hat{\mathbf{C}}_{k+1}$, at time $k+1$, can be obtained and structured.

Step 4: Calculate the prefilter coefficients, $\mathbf{F}_{k+1}$, of the vector DFE in (6.9) and the prefiltered CIR estimate as $\tilde{\mathbf{H}}_{k+1} = \hat{\mathbf{C}}_{k+1} * \mathbf{F}_{k+1}$ where $*$ is the convolution operator.

Step 5: Prefilter the received signals to obtain the T outputs of (6.10).

Step 6: Calculate the feedback terms in (6.12) using tentative decisions from the previous symbol period.

Step 7: For each of the T parallel VAs calculate the branch metric of (6.11) and advance the algorithms by one time step.

Step 8: Output the data decisions $\hat{\mathbf{d}}_{k-\triangle_{total}+1}$.

Step 9: At each succeeding time instant, repeat steps 1 to 9 to the end of the frame.

6.5 SIMULATION RESULTS AND DISCUSSIONS

We now evaluate the performance of the integrated PVA-based receiver. We assume a similar uncoded, VBLAST-type [16], MIMO system as employed in Chapter 5, where each transmitter uses the same M-ary modulation, pulse-shape, carrier frequency and transmit power. This is considered to be one of the more difficult detection scenarios since only channel differences can be used to separate the spatially-multiplexed co-channel signals [32]. We assume independent, wide-sense-stationary uncorrelated-scattering (WSSUS) subchannels with similar fading conditions on each. The fading processes are assumed to follow Clarke's fading model [43] and are simulated according to [44]. Each subchannel is assumed to have a uniform power delay profile with $L = 3$ rays.

Independent QPSK signal streams are transmitted from each of T antennas. A raised cosine filter with a roll-off of 0.99 is used at the transmitter with its response truncated to $\pm 2T_s$. An ideal low pass filter with sufficient bandwidth to

accommodate the Doppler faded signal is employed at the receiver inputs and an oversampling factor of $N_r = 2$ is used. We mimic GSM specifications where each frame at each antenna consists of $L_t = 26$ training symbols and $L_d = 116$ data symbols, except for the first frame[5] where $L_t = 78$. We also include $L_f + L - 2 = 7$ termination symbols to ensure a known trellis end state. Both the training symbols and data symbols are randomly generated. We use a prefilter with $L_f = 6$ taps, and assume a VA decision delay of $5L = 15$ symbol periods. For each SNR point, the simulation is carried out until 200 errors are accumulated in each subchannel.

The SNR is defined as the received $\frac{E_b}{\sigma_n^2}$ per receive antenna. The average received energy per bit E_b is defined as

$$E_b = \frac{\sigma_d^2 \cdot \sigma_h^2 \cdot T}{log_2 M} \tag{6.19}$$

where $\sigma_d^2 = 1$ is the transmitted symbol power and $\sigma_h^2 = 1$ is the normalized variance of the composite channel responses. For a given SNR, the complex AWGN variance of σ_n^2 can be calculated as

$$\sigma_n^2 = \frac{E_b}{10^{\frac{SNR}{10}}}. \tag{6.20}$$

We compare the performance of the integrated PVA receiver with a VVA [35] receiver also operating with a VGRLS estimator to provide the CIR for MLSE. A VGRLS estimator with a predictor length of $P = 3$ and polynomial order of $N = 2$ is used. Both receivers are operated under the same fading conditions. Using QPSK ($M = 4$) with a channel delay spread, $L = 3$ symbol periods, the number of states required by the PVA and VVA receivers are plotted in Figure 6.2 for comparison. For simplicity we restrict most of the simulations to a (2,2) system. For this scenario, the VVA receiver requires 4096 states while the PVA receiver requires a total of 128 states; a 32 fold reduction in complexity.

Figure 6.3 illustrates the effect of VGRLS channel estimation error upon

[5]Similar to the condition used in section 5.6.1.

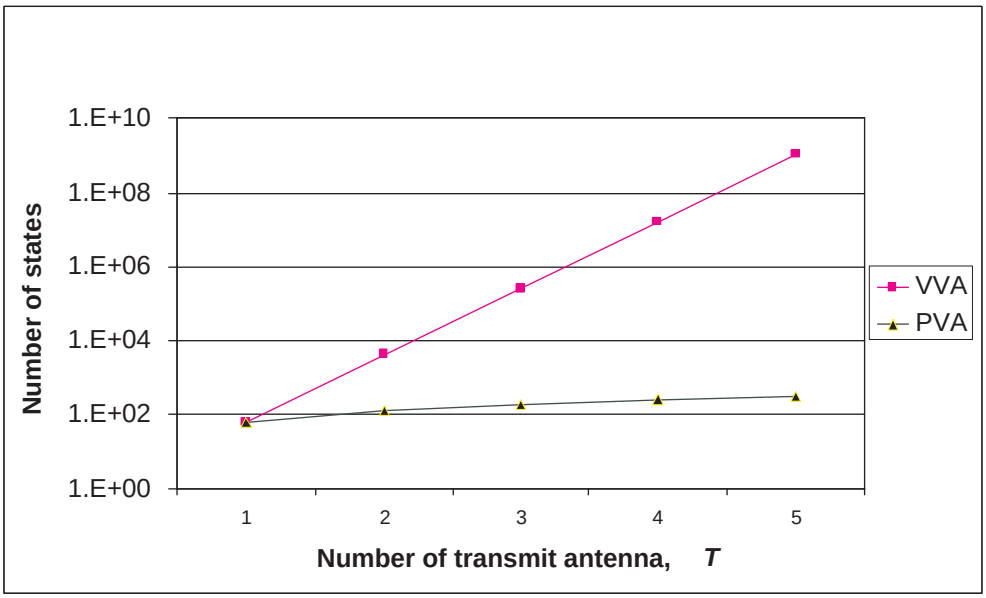

Figure 6.2 A comparison of the number of states required in the trellis search for PVA and VVA based receiver, for a (T,R) MIMO systems for $T = R$, with QPSK modulation $M = 4$ and channel delay spread $L = 3$.

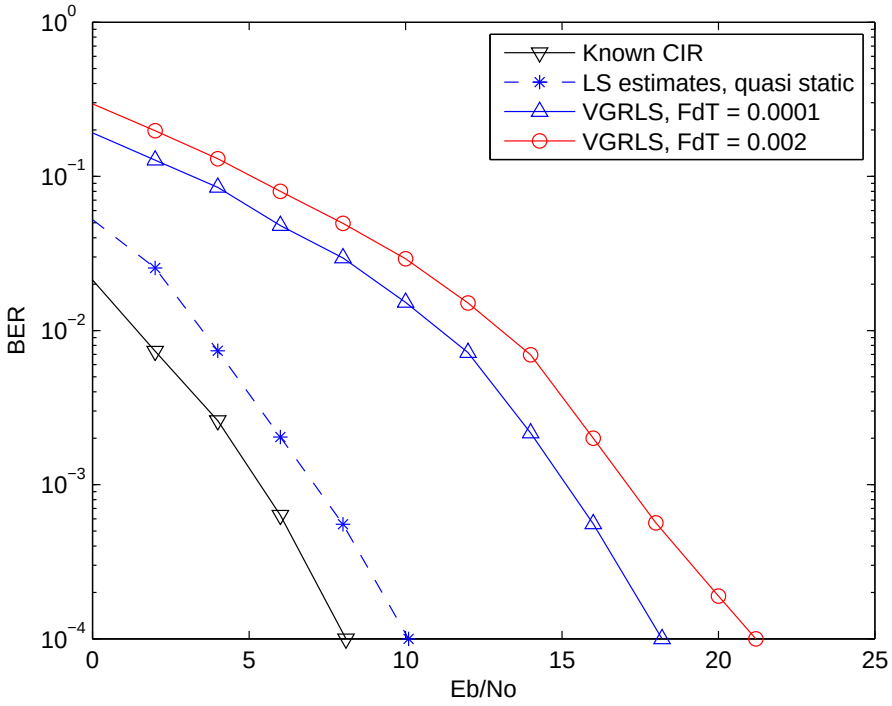

Figure 6.3 BER of the PVA receiver when operating with the VGRLS estimator for a (2,2) MIMO system with a normalized Rayleigh fade rate of 0.0001 and 0.002. The performance of the PVA with known CIR and when using LS channel estimates in a quasi static fading channel is also plotted for references.

the average bit error rate (BER) performance of the integrated PVA receiver in a (2,2) MIMO fading channel with normalized fade rates, $f_D T_s$, of 0.002 and 0.0001, where f_d is the maximum Doppler frequency. The performance of the PVA receiver with known CIR and with LS channel estimation in a quasi static fading channel [32] is also included for reference. For $f_D T_s = 0.002$, we note that the VGRLS channel estimation error degrades the performance by about 12dB at an BER of 10^{-3} compared to the known CIR case. At a BER of 10^{-4}, the performance is slightly worse as the curve is seen to start flooring although very gradually. This degradation is largely due to the dynamics of the fading as it introduces errors in the channel estimates, which in turn affects the accuracy of the DFE prefilter calculation and subsequently the prefiltered received signal and estimated CIR, and lastly the tentative decisions. This is evident in slower fading with $f_D T_s = 0.0001$ where the error floor disappears and the degradation is reduced to only 10dB at both the above BER values. Compared to the quasi static fading case, a continuously time-varying fading channel has a significant effect on the performance of the integrated PVA receiver.

Figure 6.4 shows the BER performance of the integrated PVA receiver and a VVA receiver, both for a (2,2) MIMO system, each operating with a VGRLS estimator in similar Rayleigh fading channels with $f_D T_s$ values of 0.002 and 0.0001 as before. The performance of a (2,2) system using LS channel estimates in a quasi-static fading channel [32] is also simulated for reference. We observe that the difference between the PVA and VVA receivers at $f_D T_s = 0.002$ is about 4dB at an BER of 10^{-4}. This difference is only 2dB [32] at the same BER value in a quasi static fading channel. The additional degradation is due to the time-varying channel estimation error (on prefilter calculation and the subsequent effects) because at $f_D T_s = 0.0001$, the error floor disappears and the difference is reduced to less than 3dB. The results show that the integrated PVA receiver is capable of near MLSE detection in a continuous fading environment, and that it achieves this at significantly lower complexity.

The performance of the PVA receiver for a (4,4) MIMO system, together with

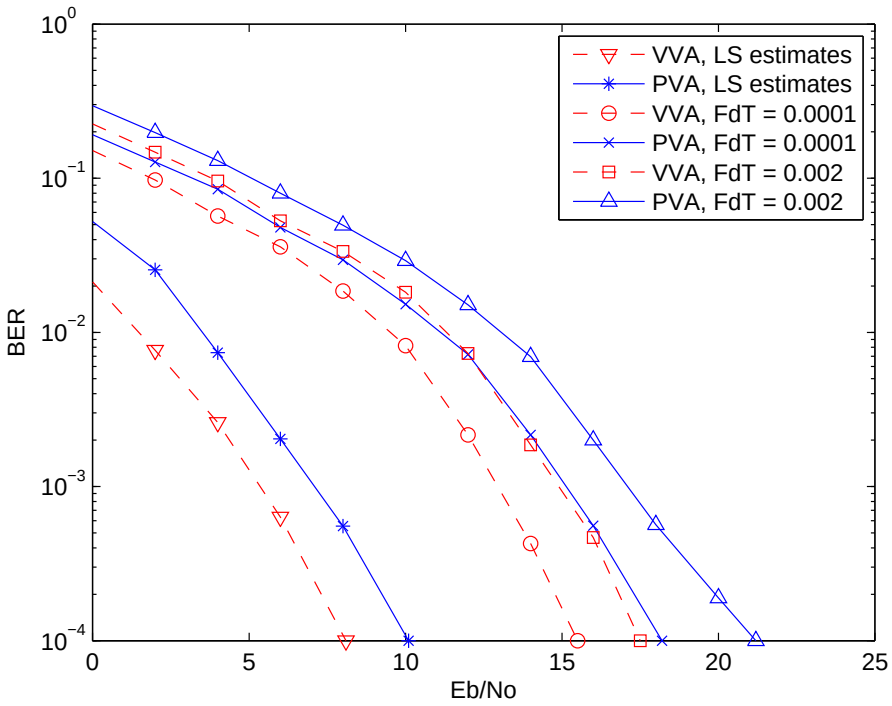

Figure 6.4 BER of the PVA and VVA receivers when operating with the VGRLS estimator for a (2,2) MIMO system with normalized Rayleigh fade rates of 0.0001 and 0.002. The performance of PVA and VVA using LS estimates in a quasi static fading channel is also plotted for comparison.

the performance of (2,2) system, with a normalized fade rates of 0.002 and 0.0001 is shown in Fig. 6.5. In general, the performance for a (4,4) system is about 2.5 to 3dB worse than that of the (2,2) system.

The decision delay $\triangle$ of the VA is typically set at $5L$ symbol periods [36]. For our simulations with $L = 3$, the VA decision delay is therefore 15 symbol periods. We investigate two options to overcome this long latency: one is to use a shorter VA decision delay, which we propose to be $2L = 6$ symbol periods, while the other is to employ the tentative decisions, corresponding to a zero VA decision delay, as reference signals for the VGRLS estimator in each symbol period. Note that in both cases there is still a DFE prefilter decision delay of $L_f - 1 = 5$ symbol periods where channel prediction is required.

The BER performance of the these two options at $f_D T_s = 0.0001$ is shown in Fig. 6.6. We note that performance, when using a VA decision delay of 6 symbols, is slightly better than when using the zero-delay VA tentative decisions,

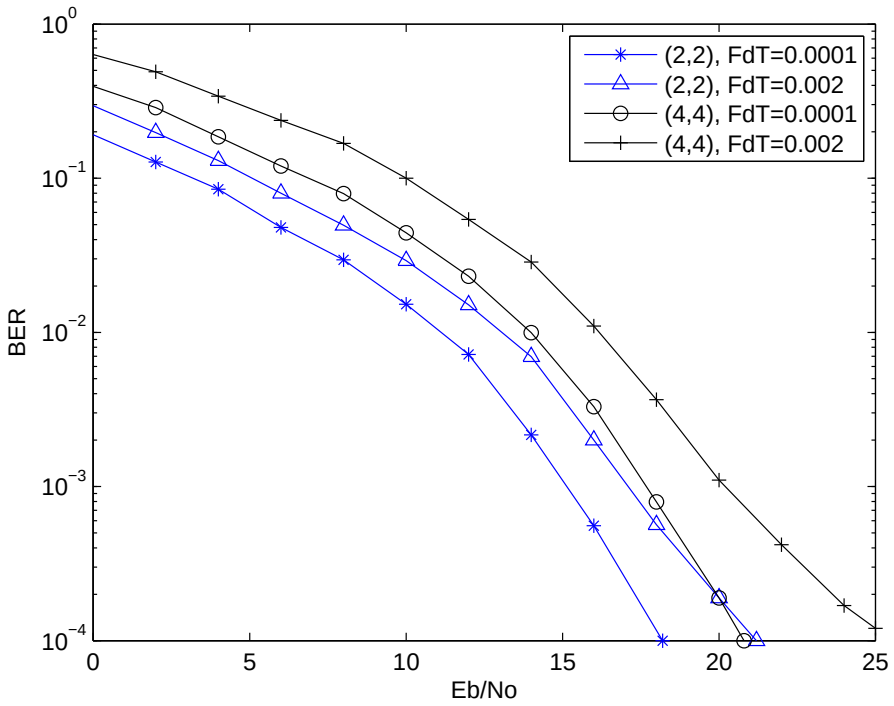

Figure 6.5 Performance comparison of a PVA receiver for a (2,2) and (4,4) MIMO system with a normalized fade rates of 0.002 and 0.0001.

although the difference is only about 0.2 dB. Both performances are, as expected, very slightly worse than when using the original VA delay of 15 symbols. Nevertheless, this very small degradation shows that zero-delay VA tentative decisions can be used and that there is little justification for the use of the longer VA decision delay. Besides this, we have also shown in Fig. 6.6 that the performance degradation without employing the channel prediction module is negligible, at about 0.5dB. This is consistent with the result shown in section 5.6.1 where the performance degradation without using the channel prediction module in a slowly fading channel is less than 1dB.

The complexity of channel estimation and tracking can be further reduced by using the VGLMS estimator at least in slow fading. To see this, we evaluate the performance of an integrated PVA receiver using the VGLMS estimator. The receiver is, as before, operated for a (2,2) Rayleigh fading with normalized fade rates of 0.0001 and 0.002. As can be seen in Fig. 6.7, the VGLMS-based receiver performs 2 - 3 dB worse than the VGRLS-based receiver. This degradation in

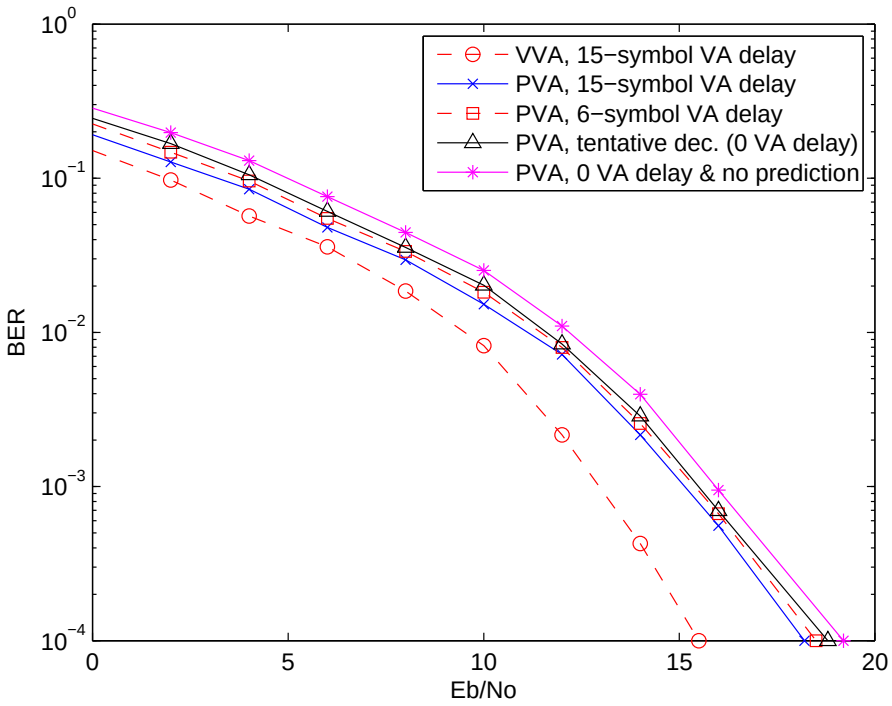

Figure 6.6 BER of the PVA and VVA receivers when operating with the VGRLS estimator for a (2,2) MIMO system with a normalized Rayleigh fade rate of 0.0001. The decision delays of the VAs are as indicated. Using tentative decisions (i.e., zero delay) is shown to have negligible degradation on performance, as well as without using the channel prediction module.

performance is to be expected because of the simplification in the estimation process. We also note that for the VGLMS-based receiver at a fade rate of 0.002, the error rate performance is seen to start gradually flooring more obviously than in the case of the VGRLS-based receiver. However, within the range of SNR and conditions studied, the VGLMS-based receiver offers an attractive trade-off of performance versus complexity.

The PVA algorithm, being a sequence-based detection method using the Viterbi algorithm, is more complicated than the symbol-based detection method such as the vector decision feedback equalizer (DFE). As such the performance of the integrated PVA receiver is expected to be better than that of a DFE-based integrated receiver also employing the VGRLS estimator under the same channel condition. Fig. 6.8 shows a comparison of the average symbol error rate (SER) performance of the integrated PVA receiver with the integrated DFE receiver of Chapter 5 in a Rayleigh fading with a normalized fade rate $f_D T = 0.002$. It

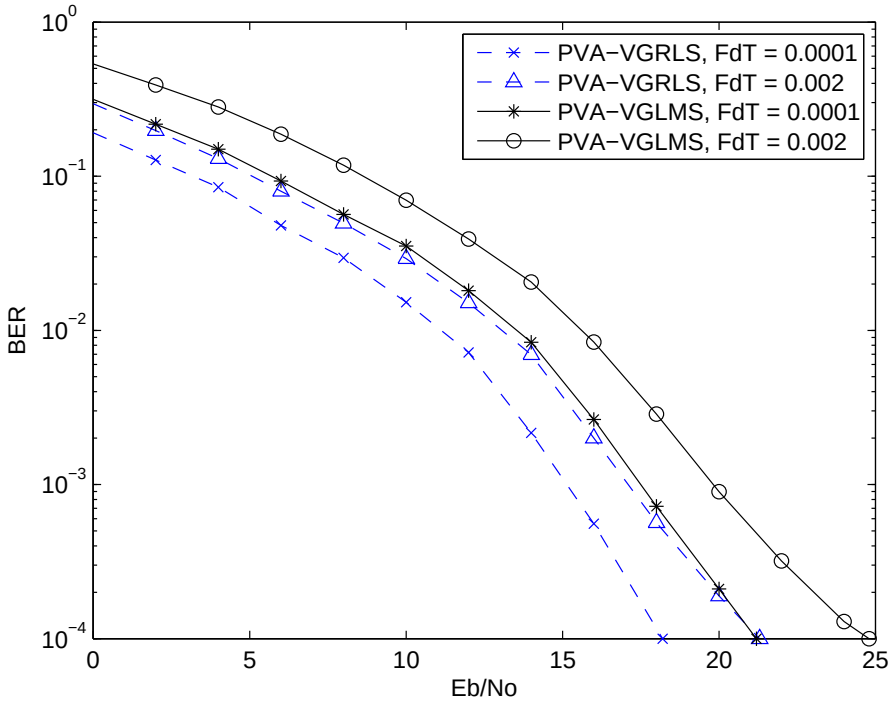

Figure 6.7 BER of a PVA-based receiver using VGRLS and VGLMS estimators for a (2,2) MIMO system with normalized Rayleigh fade rates of 0.0001 and 0.002.

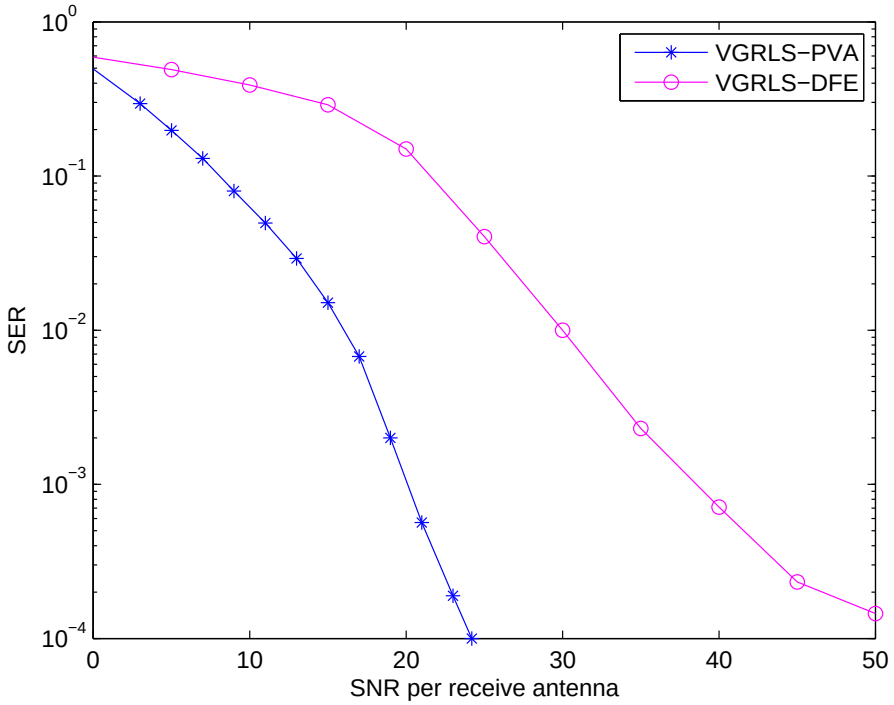

Figure 6.8 Performance comparison of the integrated PVA receiver with a DFE-based receiver of Chapter 5 operating in a MIMO system with normalized Rayleigh fade rate of 0.002.

clearly shows the superiority of the PVA receiver over the DFE receiver where the performance is about 17 dB better at an SER of 10^{-3}. The SER results in that chapter have also shown that the performance of the DFE receiver using the VGRLS estimator is within 1 - 3 dB of that obtained using an optimum KF-based estimator. The performance difference attributed to using VGRLS and KF-based estimators for a PVA receiver is within the same margin.

6.6 SUMMARY

We have developed a reduced complexity sequence-based receiver for a MIMO system that can operate in a continuously time-varying fading environment. The overall receiver, which is an implementable approximation to MLSE in MIMO channels, is implemented by combining the PVA algorithm with the VGRLS/VGLMS channel estimator. It is among the first to explicitly incorporate dynamic channel estimation in the context of sequence estimation receivers. Simulation results show that the integrated PVA-based receiver can offer near MLSE performance when compared to a VVA-based receiver, also using a VGRLS estimator, and at significantly lower complexity in term of the total number of trellis states. Simulations also show that using zero-delay tentative decisions results in negligible performance loss. Furthermore for a very slowly fading channel where the channel is assumed to stay constant, it has been shown that the channel prediction module is not necessary. The complexity in channel estimation can be further reduced by using a VGLMS estimator in place of the VGRLS approach, and the simulation results provided for the range of SNR and conditions studied illustrate the trade-off of performance versus complexity.

Chapter 7

CONCLUSIONS

In this chapter we summarize the contributions of the thesis and suggest some possible future work.

7.1 CONTRIBUTIONS

In the thesis we have presented two receiver structures with integrated channel estimation and tracking for time and frequency selective MIMO fading channels. The receiver structures are based on a symbol-by-symbol equalization technique using a MIMO MMSE DFE, and a sequence-based technique using a PVA which is a suboptimal form of MLSE.

We have shown that when compared to an optimum Kalman estimator, the VGRLS channel estimator developed in Chapter 3 is able to offer comparable performance with a predictor length $P = 3$ and polynomial order, $N = 2$. The baseline complexity of both the estimators is the same. However, the VGRLS estimator has advantages over the Kalman estimator in that it does not require knowledge of the channel statistics to operate, and can readily operate in a Rician fading environment. A Kalman estimator requires the (zero-mean) channel statistics in order to derive the coefficients of the state transition matrix. Therefore in a Rician fading channel, specific re-configuration of the structure of the state transition matrix is necessary.

A simplified version of the channel estimator known as the VGLMS estimator is presented in Chapter 4. By replacing the online recursion of the 'intermediate'

matrix in the VGRLS estimator with an offline pre-computed matrix, complexity is reduced by an order of magnitude. It is shown that when the MIMO dimension in terms of the number of transmit and receive antenna is large, the saving in computational load becomes more significant. The drawback of the reduced complexity is, however, that the performance of the VGLMS estimator is not as good as the VGRLS estimator. However, it can also operate readily in a Rician fading environment. It is also shown to offer sufficiently good performance in slowly Rayleigh fading channels, or Rician fading channel with a strong mean component.

Both estimators are integrated independently with MIMO receivers. In Chapter 5, a MMSE DFE is employed. The channel estimates from the estimator are used to calculate the feedforward and feedback filter tap coefficients. In a decision directed mode, the outputs from the DFE are used by the channel estimator to update the next channel estimates. However, due to the decision delay of the DFE, the outputs are delayed and when this is used by the channel estimator, delayed channel estimates are produced. This poses a problem as the calculation of the DFE filter tap coefficients requires updated channel estimates. To overcome this problem, we propose using a polynomial channel predictor to predict those channel response required by the DFE. However, we have also shown the prediction module may in practice be omitted with negligible performance penalty in slow fading. We have therefore presented a symbol-by-symbol based receiver that can operate in continuously fading environment.

In Chapter 6, a sequence-based receiver capable of operating in a continuously time-varying environment is presented by integrating the PVA algorithm, which is a suboptimal form of MLSE, with the estimators independently. The PVA-receiver is shown to offer near MLSE performance when compared to a VVA-receiver and at significantly lower complexity in terms of the number of trellis states. Due to the decision delay, the output of the VAs are delayed. As such the receiver also encounters a similar problem to the DFE-based receiver in that the channel estimates are not up-to-date. A polynomial channel prediction module

has been proposed to provide these channel estimates. We have also investigated using the zero-delay tentative decisions as feedback to the channel estimator and simulation results indicate negligible loss. We have also shown that the channel prediction module may be omitted in slow fading. The resulting PVA-receiver is among the first to explicitly incorporate dynamic channel estimation in the context of sequence estimation receivers.

7.2 SUGGESTED FUTURE WORK

Our simulation results have shown that the performance of the VGRLS and VGLMS estimators are affected by the value of the parameters used, for example the polynomial order N, and 'forget factor', λ. If the channel conditions, for example the normalized fade rate and SNR, are known, the appropriate values of the parameters may be chosen in advance. However, in reality these may not be readily available. Hence a dynamic selection of the parameters may be needed during a 'start-up' period, perhaps when the estimator is operated in training mode. Our simulation have indicated that for most conditions the VGRLS estimator, for instance, offers good performance with a predictor length $P = 3$, polynomial order $N = 2$ and $\lambda = 0.9$. With dynamic selection, different combination of parameter values may be used and those that offer the best performance selected for use during decision-directed mode. This evaluation may be carried out repeatedly during subsequent channel acquisition periods, if necessary.

In the thesis we have treated the MIMO channel as independent and uncorrelated by assuming sufficiently spaced antennas and a rich scattering environment. However, in reality the MIMO channel may not be spatially uncorrelated when the antennas spacing and angular spread of the incoming rays are small. Further investigation on how this spatial correlation affects the structure, and subsequently the performance of the channel estimator, and that of the integrated receiver, should be done.

Throughout the thesis we have assumed an uncoded MIMO system. In in-

vestigating a way to improve the performance of the symbol-by-symbol based receiver, we have investigated the effect of having different DFE filter lengths. The performance of the integrated receiver may be improved with the use of channel coding and this may be investigated further. Furthermore, we have integrated the receiver in a purely spatial multiplexing context with no successive interference cancellation (SIC). MIMO-DFE based BLAST system with SIC [57],[58] is found to offer improved performance. An adaptation of SIC into our receiver to improve performance may be investigated.

The proposed VGRLS estimator offers an attractive alternative to the optimum Kalman estimator and it may be integrated with other MIMO systems where channel estimation and tracking using the Kalman estimator is used. For example, a Kalman filter is employed for channel tracking in a space-time coded system [122] and instead of Kalman estimator, the VGRLS estimator may be used.

It would be interesting to implement the proposed integrated receiver practically and compare the performance with the simulation results. Of particular interest would be the channel estimator, especially the reduced complexity VGLMS estimator, as it is shown to offer a nice trade-off between performance and complexity in our simulation results.

Appendix A

YULE-WALKER EQUATIONS

If we assume that each of the CIR $h_{k+1,\mu}$ for $\mu = 1, 2, \cdots, RN_rTL$ in (3.7) evolves according to an AR process of order P_a, it can then be represented by the following difference equation

$$h_{k+1,\mu} = \phi_1 h_{k,\mu} + \phi_2 h_{k-1,\mu} + \cdots + \phi_{P_a} h_{k-Pa+1,\mu} + v_{k+1} \tag{A.1}$$

where ϕ_l for $l = 1, 2, \cdots, P_a$ is the AR coefficients and v_k is the zero-mean process noise.

Multiply both side of (A.1) by $h_{k,\mu}$, a lag-1 sample of the CIR,

$$h_{k,\mu} h_{k+1,\mu} = \sum_{p=1}^{P_a} \phi_p h_{k,\mu} h_{k-p+1,\mu} + h_{k,\mu} v_{k+1} \tag{A.2}$$

where k and p are the time and term indices.

Taking expectance, we have

$$E\{h_{k,\mu} h_{k+1,\mu}\} = \sum_{p=1}^{P_a} \phi_p E\{h_{k,\mu} h_{k-p+1,\mu}\} + E\{h_{k,\mu} v_{k+1}\}. \tag{A.3}$$

Note that $\mathrm{E}\{h_{k,\mu} v_{k+1}\} = 0$ as the process noise is assumed to be uncorrelated to the fading process.

Define $\mathrm{E}\{h_{k,\mu} h_{k+q,\mu}\} = r_q$ as the lag-q autocorrelation, (A.3) can be written as

$$r_1 = \sum_{p=1}^{P_a} \phi_p r_{p-1}. \tag{A.4}$$

Following the similar process, next we multiply (A.1) by $h_{k-1,\mu}$, a lag-2 sample of the CIR to obtain

$$r_2 = \sum_{p=1}^{P_a} \phi_p r_{p-2} \tag{A.5}$$

and continue to do so for the next P_a samples. For lag-P_a, we have

$$r_{P_a} = \sum_{p=1}^{P_a} \phi_p r_{p-P_a}. \tag{A.6}$$

Rewriting all these equations together yields

$$\begin{aligned}
r_1 &= \phi_1 r_0 + \phi_2 r_1 + \phi_3 r_2 + \cdots + \phi_{P_a-1} r_{P_a-2} + \phi_{P_a} r_{P_a-1} \\
r_2 &= \phi_1 r_1 + \phi_2 r_0 + \phi_3 r_0 + \cdots + \phi_{P_a-1} r_{P_a-3} + \phi_{P_a} r_{P_a-2} \\
\vdots &= \vdots \\
r_{P_a-1} &= \phi_1 r_{P_a-2} + \phi_2 r_{P_a-3} + \phi_3 r_{P_a-4} + \cdots + \phi_{P_a-1} r_0 + \phi_{P_a} r_1 \\
r_{P_a} &= \phi_1 r_{P_a-1} + \phi_2 r_{P_a-2} + \phi_3 r_{P_a-3} + \cdots + \phi_{P_a-1} r_1 + \phi_{P_a} r_0
\end{aligned}$$

which can also be written as

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{P_a-1} \\ r_{P_a} \end{pmatrix} = \begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_{P_a-2} & r_{P_a-1} \\ r_1 & r_0 & r_1 & \cdots & r_{P_a-3} & r_{P_a-2} \\ & \vdots & & & \vdots & \\ r_{P_a-2} & r_{P_a-3} & r_{P_a-4} & \cdots & r_0 & r_1 \\ r_{P_a-1} & r_{P_a-2} & r_{P_a-3} & \cdots & r_1 & r_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_{P_a-1} \\ \phi_{P_a} \end{pmatrix}.$$

Recalling that $r_0 = 1$, the above equation is also

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{P_a-1} \\ r_{P_a} \end{pmatrix}}_{\mathbf{r}} = \underbrace{\begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{P_a-2} & r_{P_a-1} \\ r_1 & 1 & r_1 & \cdots & r_{P_a-3} & r_{P_a-2} \\ & \vdots & & & \vdots & \\ r_{P_a-2} & r_{P_a-3} & r_{P_a-4} & \cdots & 1 & r_1 \\ r_{P_a-1} & r_{P_a-2} & r_{P_a-3} & \cdots & r_1 & 1 \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_{P_a-1} \\ \phi_{P_a} \end{pmatrix}}_{\Phi}$$

or succinctly

$$\mathbf{R}\Phi = \mathbf{r}. \tag{A.7}$$

Since $\mathbf{R}$ is full rank and symmetric, it is invertible and hence

$$\Phi = \mathbf{R}^{-1}\mathbf{r}. \tag{A.8}$$

The process noise autocovariance can be found by using a lag-0 correlation,

$$r_v = r_0 - \sum_{p=1}^{P_a} \phi_p r_p \tag{A.9}$$

For a given (measured or assumed) autocorrelation process of the fading channel, the AR coefficients and process noise autocovariance can therefore be derived. For a MIMO system, if we assume that the fading condition is similar for each of the subchannels, the above can be easily extend to a vector form as given by (3.36).

REFERENCES

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

[2] S. S. Haykin and M. Moher, *Modern Wireless Communications*. Pearson Prentice Hall, 2005.

[3] W. Webb, *Wireless Communications : The Future*. John Wiley, 2007.

[4] K. Wesolowski, *Mobile Communication Systems*. John Wiley, 2002.

[5] M. Schwartz, *Mobile Wireless Communications*. Cambridge : Cambridge University Press, 2005.

[6] H.-H. Chen and M. Guizani, *Next Generation Wireless Systems and Networks*. John Wiley, 2006.

[7] H.-H. Chen, *Next Generation CDMA Technologies*. John Wiley, 2007.

[8] I. Poole, *Cellular Communications Explained : From Basics to 3G*. Oxford : Newnes, 2006.

[9] S. G. Glisic, *Advanced Wireless Communications : 4G Technologies*. Wiley, 2004.

[10] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to Gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[11] Y. H. Kho, "3G wireless multimedia: The road ahead," 1^{st} *prize, IET Write Around the World competition 2001*, avaliable online at http://www.theiet.org/about/scholarships-awards/ambition/watw-entries.cfm.

[12] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[13] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[14] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[15] G. J. Foschini and M. J. Gans, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs Tech J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[16] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[17] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

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

[20] H. Jafarkhani, *Space-time Coding: Theory and Practice.* Cambridge : Cambridge University Press, 2005.

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

[22] H. Yang, "A road to future broadband wireless access: MIMO-OFDM-based air interface," *IEEE Commun. Mag.*, vol. 43, no. 1, pp. 53–60, Jan. 2005.

[23] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 4, pp. 360–393, Dec. 1963.

[24] W. S. Leon, *The Reception of QPSK Signals Transmitted over Dispersive Fading Channels Using the f-power Series Channel Model.* Masters Thesis, University of Canterbury, 1995.

[25] D. K. Borah and B. D. Hart, "A robust receiver structure for time-varying, frequency-flat, Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 360–364, Mar. 1999.

[26] W. S. Leon, *Equalization and Estimation for Fading Channels.* PhD Thesis, University of Canterbury, 2003.

[27] W. S. Leon and D. P. Taylor, "An adaptive receiver for the time and frequency selective fading channel," *IEEE Trans. Commun.*, vol. 45, pp. 1548–1555, Dec. 1997.

[28] ——, "DPSK receiver with implicit diversity for the linear frequency-selective Rayleigh fading channel," *Int. J. of Wireless Information Networks*, vol. 9, pp. 275–286, Oct. 2002.

[29] ——, "The polynomial-based generalized recursive least squares estimator for Rayleigh fading channels," *Proc. of GLOBECOM*, vol. 5, pp. 2401–2405, 2003.

[30] ——, "The polynomial-based generalized least mean squares estimator for Rician and Rayleigh fading channels," *Proc. of IEEE 60th VTC*, vol. 6, pp. 3861–3865, 2004.

[31] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Proccessing*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.

[32] C. L. Miller, D. P. Taylor, and P. T. Gough, "Estimation of co-channel signals with linear complexity," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1997–2005, Nov. 2001.

[33] G. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Info. Theory*, vol. 18, pp. 363–378, May 1972.

[34] ——, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[35] W. van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. 24, no. 2, pp. 276–283, Feb. 1976.

[36] J. G. Proakis, *Digital Communications*, 4th ed. Boston: McGraw-Hill, 2001.

[37] T. S. Rappaport, *Wireless Communications*. Prentice-Hall, Upper Saddle River, N.J., 1996.

[38] A. Papoulis, *Probability, Random Variable and Stochastic Processes*. McGraw-Hill: New York, third edition, 1991.

[39] D. P. Taylor, G. M. Vitetta, B. D. Hart, and A. Mämmelä, "Wireless channel equalization," *Eur. Trans. Telecom.*, vol. 9, pp. 117 – 143, Mar.-Apr. 1998.

[40] R. S. Kennedy, *Fading Dispersive Communication Channels.* New York: Wiley-Interscience, 1969.

[41] G. Deng, J. Cavers, and P. Ho, "A reduced dimensionality propagation model for frequency selective Rayleigh fading channels," *Proc. IEEE ICC*, pp. 1158–1162, 1995.

[42] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000, 1968.

[43] W. C. Jakes, *Microwave Mobile Communications : Edited by William C. Jakes.* N.Y.: Wiley, 1974.

[44] D. Verdin and T. C. Tozer, "Generating a fading process for the simulation of land-mobile radio communications," *Electronics Letters*, vol. 29, no. 23, pp. 2011–2012, Nov. 1993.

[45] L. M. Correia, *Mobile Broadband Multimedia Networks.* Elsevier, 2006.

[46] S. A. Fechtel and H. Meyr, "An investigation of channel estimation and equalization techniques for moderately rapid fading HF channels," *Proc. of ICC*, vol. 2, pp. 768–772, Jun. 1991.

[47] S. U. H. Qureshi, "Adaptive equalization," *IEEE Proceedings*, vol. 73, pp. 1349–1387, Sept. 1985.

[48] R. W. Lucky, "Automatic equalization for digital communications," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, 1965.

[49] S. S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.

[50] C. A. Belfiore and J. H. Park Jr., "Decision feedback equalization," *IEEE Proceedings*, vol. 67, pp. 1143–1156, Aug. 1979.

[51] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 56–64, Jan, 1971.

[52] J. Salz, "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, pp. 1341–1373, Oct. 1973.

[53] D. Falconer and G. J. Foschini, "Theory of minimum mean-square-error QAM systems employing decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, Dec. 1973.

[54] N. Al-Dhahir and J. M. Cioffi, "MMSE decision feedback equalizers: Finite-length results," *IEEE Trans. Inform. Theory*, vol. 41, pp. 961 – 976, Jul. 1995.

[55] P. A. Voois, I. Lee, and J. M. Cioffi, "The effect of decision delay in finite-length decision feedback equalization," *IEEE Trans. Information Theory*, vol. 42, no. 2, pp. 618–621, Mar. 1996.

[56] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Channel estimation and equalization in fading," *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, pp. 1159 – 1163, Oct. 1999.

[57] A. Lozano and C. Papadias, "Layered space-time receivers for frequency-selective wireless channels," *IEEE Trans. Commun.*, vol. 50, pp. 65–73, Jan. 2002.

[58] X. Zhu and R. D. Murch, "MIMO-DFE based BLAST over frequency selective channels," *IEEE Globecom*, vol. 1, pp. 499–503, Nov. 2001.

[59] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, vol. 43, pp. 421–434, Feb-Mar-Apr 1995.

[60] C. Tidestav, M. Sternad, and A. Ahlén, "Reuse within a cell: Interference rejection or multiuser detection?" *IEEE Trans. Commun.*, vol. 47, pp. 1511–1522, Oct. 1999.

[61] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation for CPM signals transmitted over Rayleigh flat fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, 1990.

[62] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 941–950, Feb./Mar./Apr. 1994.

[63] X. Yu and S. Pasupathy, "Innovation-based MLSE for Rayleigh fading channels," *IEEE Commun. Mag.*, vol. 43, pp. 1534–1544, Feb./Mar./Apr./ 1995.

[64] R. Raheli, A. Polydoros, and C.-K. Tzou, "The principle of per-survivor processing: A general approach to approximate and adaptive MLSE," *Proc. IEEE Globecom*, pp. 1170–1175, 1991.

[65] W. van Etten, "An optimum linear receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. 23, pp. 824–834, Aug. 1975.

[66] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.

[67] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.

[68] J. B. Andersen and S. Mohan, "Sequential coding algorithm: A survey and cost analysis," *IEEE Trans. Commun.*, vol. COM-32, pp. 169–176, Feb. 1984.

[69] C. L. Miller and D. P. Taylor, "Digital prefiltering for multiple-input, multiple-output receivers," *Proc. of ICC*, vol. 6, pp. 1696–1699, Jun. 2001.

[70] G. Paparisto and K. M. Chugg, "PSP array processing for multipath fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 504 – 507, Apr. 1999.

[71] S. N. Diggavi, B. C. Ng, and A. Paulraj, "An interference suppression scheme with joint channel-data estimation," *IEEE J. Select. Areas. Commun.*, vol. 17, pp. 1924 – 1939, Nov. 1999.

[72] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944 – 960, Jul. 1995.

[73] K. Li and K. J. R. Liu, "Adaptive blind source separation and equalization for multiple-input/multiple-output systems," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2864 – 2876, Nov. 1998.

[74] B. Tomiuk and N. Beaulieu, "A new look at maximal ratio combining," *Proc. IEEE Globecom*, vol. 2, pp. 943–948, Dec. 2000.

[75] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," *Proc. IEEE ICC*, vol. 3, pp. 1630–1634, May 1993.

[76] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over rich-scattering wireless channel," *Proc. URSI International Symposium on Signals, Systems, and Electronics*, pp. 295–300, Sept. 1998.

[77] M. Sellathurai and S. S. Haykin, "Turbo-BLAST: performance evaluation in correlated Rayleigh-fading environment," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 340–349, Apr. 2003.

[78] A. Van Zelst, R. Van Nee, and G. A. Awater, "Turbo-BLAST and its performance," *Proc. IEEE VTC*, vol. 2, pp. 1282–1286, 2001.

[79] X. Li, H. Huang, G. J. Foschini, and R. A. Valenzuela, "Effetcs of iterative detection and decoding on the performance of BLAST," *Proc. IEEE Globecom*, vol. 2, pp. 1061–1066, 2000.

[80] W. J. Choi, R. Negi, and J. M. Cioffi, "Combined ML and DFE decoding for the V-BLAST system," *Proc. IEEE ICC*, vol. 3, pp. 1243–1248, 2000.

[81] Y. Li and Z. Q. Luo, "Parallel detection for V-BLAST system," *Proc. IEEE ICC*, vol. 1, pp. 340–344, 2002.

[82] A. Bhargave, R. J. P. de Figueiredo, and T. Eltoft, "A detection algorithm for the V-BLAST system," *Proc. IEEE Globecom*, vol. 1, pp. 494–498, 2001.

[83] B. Hassibi, "An efficient square-root algorithm for BLAST," *Proc. Intl. Conf. on Acoustics, Speech and Signal Processing*, vol. 2, pp. II737–II740, 2000.

[84] K. W. Wong, C. Y. Tsui, and R. S. Cheng, "Low complexity architecture of the V-BLAST system," *Proc. IEEE Wireless Comm. and Networking Conf.*, vol. 1, pp. 310–314, 2000.

[85] X. Li, H. Huang, R. A. Valenzuela, and G. J. Foschini, "Reduced-complexity detection algorithms for systems using multi-element arrays," *Proc. IEEE Globecom*, vol. 2, pp. 1072–1076, 2000.

[86] Z. Chen, J. Yuan, and B. Vucetic, "An improved space-time trellis coded modulation scheme for slow Rayleigh fading channels," *Proc. IEEE ICC*, vol. 4, pp. 1110–1116, Jun. 2001.

[87] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Information Theory*, vol. 45, pp. 1456–1467, Jul. 1999.

[88] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, pp. 1–4, Jan, 2001.

[89] W. Su and X. Xia, "A design of quasi-orthogonal space-time block codes with full diversity," *Proc. IEEE CSSC*, vol. 2, pp. 1112–1116, Nov. 2002.

[90] O. Tirkkonen and A. Hottinen, "Improved MIMO performance with non-orthogonal space time block codes," *Proc. of IEEE Globecom*, vol. 2, pp. 1122–1126, Nov. 2001.

[91] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[92] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1097–1109, Oct. 1986.

[93] N. Bershad and O. Macchi, "Comparison of RLS and LMS algorithms for tracking a chirped signal," *Proc. ICASSP*, vol. 2, pp. 896–899, May 1989.

[94] S. Haykin, A. Sayed, J. Zeidler, P. Yee, and P. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Proc.*, vol. 45, pp. 1118–1128, May 1997.

[95] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, pp. 18–60, Jul. 1994.

[96] D. K. Borah and B. D. Hart, "Frequency-selctive fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, pp. 8962–8973, Jun. 1999.

[97] L. M. Davis, I. B. Collings, and R. J. Evans, "Coupled estimators for equalization of fast-fading mobile channels," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1262–1265, Oct. 1998.

[98] B. D. Hart and D. P. Taylor, "Maximum-likelihood synchronization, equalization, and sequence estimation for unknown time-varying frequency-selective Rician channels," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 211–221, Feb. 1998.

[99] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Proc. of ICASSP*, vol. 2, pp. 1093–1096, May 1996.

[100] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalisation using Kalman estimation," *IEEE Trans. Signal Proccessing*, vol. 50, no. 5, pp. 1065–1076, May 2002.

[101] M. Enescu, M. Sirbu, and V. Koivunen, "Adaptive equalization of time-varying MIMO channels," *Signal Processing*, vol. 85, no. 1, pp. 81–93, Jan. 2005.

[102] A. P. Clark, *Adaptive Detectors for Digital Modems.* London, U.K.: Pentech Press, 1989.

[103] B. Chun, B. Kim, and Y. H. Lee, "Generalization of exponentially weighted RLS algorithm based on a state-space model," *Proc. of ISCAS*, vol. 5, pp. 198–201, Jun. 1998.

[104] P. T. Harju and T. I. Laakso, "Polynomial predictors for complex-valued vector signals," *Electronics Letters*, vol. 31, no. 19, pp. 1650–1652, Sept. 1995.

[105] L. L. Scharf, *Statistical Signal Processing.* Addison-Wesley, 1991.

[106] M. H. Hayes, *Statistical Digital Signal Processing and Modeling.* John Wiley and Son, 1996.

[107] P. Heinonen and Y. Neuvo, "FIR-median hybrid filters with predictive FIR substructures," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 892–899, Jun. 1988.

[108] J. A. Honkanen, T. I. Laakso, S. J. Ovaska, and I. O. Hartimo, "Lowpass IIR predictors for discrete-time signal processing," *Digital Signal Processing*, vol. 5, pp. 133–139, Jul. 1995.

[109] K. Koppinen, J. Yli-Hietanen, and J. Astola, "Optimization of generalized predictors," *IEEE Instrumentation and Measurement Technology Conf.*, vol. 1, pp. 54–59, May 1997.

[110] G. C. Reinsel, *Elements of Multivariate Time Series Analysis.* Springer-Verlag, New York, 1993.

[111] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London*, vol. 226, pp. 267–298, 1927.

[112] G. Walker, "On periodicity in series of related terms," *Proceedings of the Royal Society of London*, vol. 131, pp. 518–532, 1931.

[113] W. S. Leon and D. P. Taylor, "The generalized polynomial predictor based receiver for the nonselective fading channel," *Proc. of GLOBECOM*, vol. 2, pp. 927–931, 2000.

[114] V. Erceg, "Channel models for fixed wireless applications," *IEEE 802.16 Broadband Wireless Access Working Group*, Feb. 2001.

[115] W. H. Tranter, K. S. Shanmugam, T. S. Rappaport, and K. L. Kosbar, *Principles of Communication Systems Simulation with Wireless Applications.* New Jersey: Prentice Hall, 2004.

[116] L. Lindbom, "Simplified Kalman estimation of fading mobile radio channels: High performance at LMS computational load," *Proc. IEEE ICASSP*, vol. 3, pp. 352–355, Apr. 1993.

[117] L. Lindbom, M. Sternad, and A. Ahlén, "Tracking of time-varying mobile radio channels - part i: The Wiener LMS algorithm," *IEEE Trans. Commun.*, vol. 49, pp. 2207–2217, Dec. 2001.

[118] S. Gazor, "Prediction in LMS-type algorithm for smoothly time varying environments," *IEEE Trans. Signal Proc.*, vol. 47, pp. 1735–1739, Jun. 2001.

[119] N. Al-Dhahir, "FIR channel-shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[120] M. N. Patwary, P. B. Rapajic, and J. Choi, "Decision feedback MLSE for spatially multiplexed MIMO frequency selective fading channel," *IEE Proc.-Commun.*, vol. 153, no. 1, pp. 39–48, Feb. 2006.

[121] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, pp. 624–636, May 1974.

[122] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channel," *IEEE Trans. Commun.*, vol. 50, pp. 183–186, Feb. 2002.