On Receiver Design for an Unknown, Rapidly Time-Varying, Rayleigh Fading Channel

Gunther Auer

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
July 2000

Abstract

In this thesis receiver architectures for an unknown, time-varying Rayleigh fading channel are investigated. This includes fast fading scenarios, where the channel impulse response (CIR) can change significantly between two adjacent samples. Channel estimation based on the minimum mean squared error criterion (MMSE) applied to smoothing and linear prediction is considered. One of the key objectives in this thesis is the analysis of error propagation effects due to decision feedback. The studied receiver architectures are divided into two main parts: one-shot receivers which detect the received symbol on a symbol–by–symbol basis, and sequence detectors which jointly estimate and detect the entire received signal sequence.

Considering one-shot receivers, a decision directed receiver is studied using differential modulation (DPSK). The receiver can significantly improve the fast fading performance of conventional DPSK, through linear predictive channel estimation. It is demonstrated through simulations that the performance of the decision directed receiver is better than that of an idealised reference receiver where channel estimation is not corrupted by decision feedback errors (e.g. by means of employing a pilot signal). Furthermore, a receiver employing coherent modulation is considered. The necessary phase reference is provided by time multiplexed pilot symbols. A receiver which exclusively uses these pilot symbols for channel estimation is the pilot aided receiver. The performance for slow fading is excellent, whereas the performance degrades as the Doppler frequency increases. The degradation is proportional to the spacing of the pilots. The performance of both the decision directed and the pilot aided receiver can be significantly improved by employing a second stage channel estimation filter, using a smoothing type estimation filter.

Then, optimum maximum likelihood sequence detection (MLSD) was studied. An important feature of the optimum receiver is that a Wiener filter is the optimum channel estimation filter for detection. Based on a recursive formulation of the optimal receiver the receiver complexity can be drastically reduced by application of the Viterbi algorithm (VA). Performance bounds for the resulting receiver are derived, unfortunately, the obtained bounds are only tight for high SNR. In order to reduce the complexity of MLSD employing the VA further, state reduction techniques are devised. This is firstly, state dependent decision feedback, yielding a receiver based on per-survivor processing (PSP). Secondly, the complexity may be reduced by the list-type Viterbi algorithm, which is a combination of the VA and the M-algorithm (a breadth first implementation for sequence detection). The performance of reduced complexity receivers using a predictive FIR and a first order IIR filter is investigated. Simulation results using time multiplexed pilots as a phase reference, suggest that the performance is dependent on the degree of state reduction applied. It is shown that excessive state reduction can result in stability problems. Surprisingly, slow fading is identified the more critical scenario for the stability problems, whereas with increasing fading rates the receivers are found to be more robust. Close studies indicate that the length and nature of the channel estimation filter is responsible of the observed stability problems. Moreover, an analytical approach is presented, based on a hidden Markov model with two states, i.e. a good state and a burst state. This model, known as the Gilbert–Elliott channel (GEC), was used to analyse error propagation effects and was found to predict the stability problems well. To mitigate the stability hybrid receiver designs were proposed: first, a receiver was developed that consists of a robust reference receiver and the FIR–PSP running in parallel. The stability is maintained with an error propagation detector. Second, a hybrid receiver which switches between the FIR–PSP and a robust receiver, whenever conditions are identified to be unstable. PSP with IIR-type channel estimation offers an appropriate choice for the robust reference receiver, given the filter coefficient is constraint by an upper bound of approximately 0.5.

Finally, the effects of multiple user interference (MAI) were considered, applying a more realistic model of a spread spectrum mobile radio link to the receiver structures mentioned previously. Simulation results for the downlink (the base-station to mobile link) on a frequency selective channel show a drastic degradation in system performance due to the near-far effect, even for modest MAI.

Declaration of Originality

This thesis was composed entirely by myself. The work reported herein was conducted exclusively by myself in the Department of Electronics and Electrical Engineering at the University of Edinburgh. The software written to perform the simulations was written by myself with the following exceptions:

- The routines used to perform an eigenvalue decomposition and to perform the matrix inversion, which were obtained from *Numerical Recipes in C*, reference [69].
- The routines used to generate the impulse response of the time-varying channel described in Chapter 3, which were supplied by D. Laurenson, and are described in [54, 55].
- A program to generate Gold codes, described in [41], to simulate multiple access interference in Chapter 7, was supplied by I. Band.

Gunther Auer. Edinburgh, January 2000.

Acknowledgements

I would like to thank the following people for their invaluable assistance during the work on this thesis:

- Gordon Povey and Dave Laurenson, my supervisors for their continuous support, guidance and useful discussions. Also for reading and checking this thesis.
- The comments of the reviewers: Ian Band, John Thompson, Iain Mann, Andrew and Harald.
- Everyone in the Signal and Systems Group at the University of Edinburgh.

Contents

List of Figures

List of Tables

Abbreviations

ACF	Auto-correlation function.
AR	Auto regressive.
AWGN	Additive white Gaussian noise.
BER	Bit error ratio.
BPSK	Binary phase shift keying.
BSC	Binary symmetric channel.
CCF	Cross correlation function.
CDMA	Code division multiple access.
CIR	Channel impulse response.
DA	Decision aided.
DD	Decision directed.
DPSK	Differential phase shift keying.
DE-PSK	Differential encoded phase shift keying.
DS	Direct sequence.
EPD	Error propagation detector.
FDMA	Frequency division multiple access.
FIR	Finite impulse response filter.
GEC	Gilbert–Elliott channel.
GSM	Global system for mobile.
HMM	Hidden Markov model.
ICE	Iterative channel estimation.
IIR	Infinite impulse response filter.
ISI	Inter symbol interference.
LMS	Least mean squares adaptive algorithm.
LOS	Line of sight.
LVA	List type Viterbi algorithm.
MA	Mooving average.
MAI	Multiple access interference.
MAP	Maximum *a posteriori*.
MAP–SbSD	Maximum *a posteriori* sybol–by–symbol detection.

ML	Maximum likelihood.
MLSD	Maximum likelihood sequence detection.
MMSE	Minimum mean square error.
MRC	Maximal ratio combining.
MSE	Mean square error.
PA	Pilot aided.
PDR	Pilot to data ratio.
pdf	Probability density function.
PN	Pseudo noise.
PSI	Pilot symbol insertion.
PSP	Per-survivor processing.
RF	Radio frequency.
SNR	Signal–to–noise ratio.
SOVA	Soft output Viterbi algorithm.
TDMA	Time division multiple access.
UMTS	Universal mobile telecommunication standard.
UTRA	UMTS terrestrial radio access.
VA	Viterbi algorithm.
WMF	Whitening matched filter.

List of Principle Symbols

$a(k)$	unmodulated information symbol at time kT_s, with discrete values $\{0, \cdots, A_m - 1\}$.
A_m	Size of the symbol alphabet.
$\mathcal{A}_K$	Set of transmitted signal sequence of length K.
c_0	The speed of light.
$c_u(t)$	Spread spectrum code of user u with duration T_c.
$d(k)$	Sampled stream of data symbols $d(t = kT_s)$.
$d^{(\ell)}(k)$	Hypothesis ℓ of transmitted symbol.
$\mathbf{d}$	Data vector.
$\Delta d(k)$	Differentially encoded stream of data symbols.
$\mathbf{D}$	Diagonal data matrix.
$\mathcal{D}$	Symbol alphabet.
$e^{(\ell)}(k)$	Error signal of hypothesis ℓ.
E_s	Energy per transmitted information symbol.
$\mathcal{E}$	Error event.
f_c	RF carrier frequency.
f_s	Sampling frequency.
$g(t)$	Transmition pulse shaping filter.
$h(\tau, t)$	Time variant impulse response of time delay τ and observation time t.
$h_q(k)$	The channel response of the q^{th} tap of the tapped delay line model.
$\mathbf{h}$	Channel vector.
$\mathbf{I}$	Identity matrix.
j	Imaginary unit.
J	Size of a list.
J_0	Bessel function of the first kind.
k	discrete time.
K	Sequence length.
ℓ	number of hypothesis of transmitted data sequence.
L	Total number of data hypothesis.
L_b	length of an error burst.
L_e	length of an error event.

$\mathbf{L}$	Lower triangular matrix.
M	Filter order of the channel estimator.
$n(k)$	k^{th} sample of a complex additive white Gaussian noise (AWGN) process of power N_0.
N_c	Spread spectrum processing gain.
N_0	Complex, single–channel additive white Gaussian noise power.
$\mathcal{N}(\bar{m}, \sigma^2)$	Gaussian pdf with mean $\bar{m}$ and variance σ^2.
P_e	Probability of error.
Q	Number of RAKE filter taps.
R	Pilot symbol insertion rate.
R_b	bit rate in bit/s.
$R(\tau)$	Power profile of the channel of delay τ.
$s(t)$	Transmitted spread spectrum signal.
$S_D(\nu)$	Doppler power spectrum.
$\mathcal{S}$	Set of states at a given time.
t	Time.
T_c	Chip period.
T_s	Symbol period.
$\mathcal{T}_{ij}$	Markov model: transiton from state i to j.
U	Number of active users.
v	Vehicle velocity.
V^2	Mean square error (MSE).
V^2_{min}	Minimum mean square error (MMSE).
W	Signal bandwidth.
$\mathbf{w}$	Channel estimation filter with coefficients $\{w_m\}_{m=1}^{M}$.
$\mathbf{W}$	Filter bank of estimation filters.
$x(t)$	Pre-correlation received signal.
$y(t)$	Post-correlation noise-corrupted underlying baseband signal.
$y_q(k)$	Post-correlation noise-corrupted baseband signal of the q^{th} channel tap and sample k.
$\mathbf{y}_q$	Post-correlation noise-corrupted signal vector of the q^{th} channel tap.
$\mathbf{y}$	Post-correlation noise-corrupted signal vector.
$\alpha_q(k)$	Magnitued of q^{th} channel tap, $\lvert h_q(k)\rvert$.
γ	Signal–to–noise ratio.
$\Delta(i, i')$	Euclidean distance between decoder states i and i'.
$\epsilon(k)$	Estimation error of the k^{th} sample.

ϵ	Vector of estimation errors.
λ_i	The i^{th} eigenvalue.
$\Lambda(\ell, k)$	The decision variable for the k^{th} symbol and hypotheses ℓ.
ν	Doppler frequency.
ν_{max}	Maximum Doppler frequency.
ν'_{max}	Normalised maximum Doppler frequency, $\nu_{max}T_s$.
$\rho_{uv}(p-q)$	Correlation coefficient between code u and v with offset $p-q$.
σ^2	Variance of a random process.
$\mathbf{\Sigma}$	Diagonal matrix of eigenvalues.
τ	Time delay.
$\phi_{xy}(\Delta k)$	Correlation function between the signals $x(k)$ and $y(k-\Delta k)$.
$\boldsymbol{\phi}_{\mathbf{xy}}$	correlation vector between the signal $x(k)$ and the vector $\mathbf{y}$.
$\mathbf{\Phi}_{\mathbf{xy}}$	Covariance matrix of vectors $\mathbf{x}$ and $\mathbf{y}$.
$\chi(k)$	State in finite state machine at time k.

Operators

a^*	Complex conjugate operation.
$\text{sgn}(a)$	Sign of a.
$\widehat{a}$	Estimator of a.
$\bar{a}$	Mean of a.
$\mathbf{A}^T$	Matrix transpose operation.
$\mathbf{A}^H$	Matrix Hermitian transpose, defined as the conjugate transposed operation, $\mathbf{A}^{*T}$.
$\mathbf{A}^{-1}$	Matrix inverse operation.
$\det \mathbf{A}$	Matrix determinant operation.
$\text{Re}\{\cdot\}$	Real part of a complex number.
$\text{Im}\{\cdot\}$	Imaginary part of a complex number.
$\arg(\cdot)$	Argument of a complex number.
$\lvert\cdot\rvert$	Magnitude.
$\lVert\cdot\rVert$	Euclidean norm.
$\lfloor\cdot\rfloor$	truncate to the nearest lower integer.
$\mathbb{R}$	Real space.
$\mathbb{C}$	Complex space.
$\text{E}[\cdot]$	Statistical expectation operator.
$p(\cdot)$	Probability density function (pdf).

$p(\cdot\|\cdot)$	Conditional probability density function.
$P(\cdot)$	Probability.
$\ln(\cdot)$	Natural logarithm.
$\log(\cdot)$	Logarithm of base ten.
$\log_a(\cdot)$	Logarithm of base a.

Chapter 1
Introduction

1.1 Wireless communication systems

In recent years the development of VLSI (very large scale integration) technology made it possible to implement complex signal processing techniques within portable radio devices, which are fast, small and reliable. This has caused an explosion in interest of wireless communication services during the last decade. Wireless communication systems are now at the forefront of current research activities. There is a large number of communication products for a wide variety of applications, from wireless telephones to high speed wireless data networks, some of which are already reality. Current digital wireless systems were designed for low data rate speech services (e.g. 8 kbit/s), while future systems will include higher data rate services [1]. The motivation of wireless is to allow any one to communicate without being limited to a fixed network. On the other hand, transmitting over a radio frequency (RF) channel implies that there is only a finite frequency band available. One of the key elements in designing a wireless communication system is the effective use of the available bandwidth, in order to maximise the capacity of the system. There are three major multiple access schemes which allow the same channel to be shared among several users:

Frequency Division Multiple Access (FDMA) assigns individual RF channels to individual users. Each user is allocated an unique frequency band or channel.

Time Division Multiple Access (TDMA) systems share the same radio spectrum at different time slots, and in each slot only one user is allowed to either transmit or receive. Each user occupies a cyclically repeating time slot.

Code Division Multiple Access (CDMA) systems identify each user with an unique code, which is used to increase the bandwidth of the signal. This so called spreading code is usually chosen orthogonal to all other codewords. All users in a CDMA cell use the same carrier frequency and may transmit simultaneously. For detection of the message signal the receiver must separate the signal of interest.

While FDMA is commonly used for analog transmission, digital data and digital modulation are more popular with TDMA and CDMA.

The need for a more efficient use of RF bandwidth prompted the adaptation of a cellular structure for wireless communication systems. The main advantage of dividing a certain area into cells is that a subset of the allocated frequency range can be re-used in each cell. Thus, the cellular concept offers high capacity in a limited spectrum allocation. It is a system level idea which calls for replacing a single, high power transmitter (large cell) with many lower power transmitters (smaller cells), each providing

coverage to only a small portion of the service area. Cell sizes differ from macro cells with a typical diameter of several miles, through micro cells with a diameter of a hundred metres or so, to pico cells which are the size of a room. To allow a mobile user to move to an adjacent cell without disconnecting, the call needs to be switched to another cell, termed *hand-over.*

In a wireless communication system, it is often desirable to allow the subscriber to send and receive simultaneously information to the base station. This is called *duplexing* and may be done using frequency or time domain techniques. *Frequency division duplex* (FDD) provides two distinct frequency bands for every user. *Time division duplex* (TDD) uses time instead of frequency. If the split between the forward and reverse time slot is small, then the transmission and reception of data appears simultaneous to the user. The *downlink* provides traffic from the base station to the mobile, while the *uplink* provides traffic from the mobile to the base.[1]

The first generation cellular networks employed analogue FDMA. The number of RF channels specified the capacity of the system. The development of the second generation digital networks, such as the European GSM and the American IS-136 standards, permitted large capacity increases. These digital systems employ a combination of TDMA with multiple RF carriers. Later, the American IS-95 standards was proposed, based on CDMA (see e.g. [2] for further discussion on second generation systems). Currently third generation mobile systems, known as UMTS in Europe and IMT-2000 in the ITU [3, 4] are being developed. The UMTS terrestrial radio access (UTRA) includes two operating modes: UTRA–FDD utilising wideband CDMA in FDD mode [5]; and UTRA–TDD employing time division CDMA (TD/CDMA) [6], which is a hybrid of TDMA and CDMA, operating in TDD mode. While second generation systems were designed mainly for voice-band applications, third generation are designed to offer variable data rates ranging from voice-band up to 144 kbit/s for high-mobile users with wide area coverage, and up to 2 Mbit/s for low-mobility users with local coverage.

Future broadband mobile systems under investigation for the fourth generation will offer data rates up to 150 Mbit/s [1]. Alongside the desire for ever increasing data rates goes the requirement for the use of higher frequency bands. While third generation systems operate at frequencies around 2 GHz, systems utilising spectrum up to 60 GHz are being investigated [1]. A significant amount of spectrum is allocated for mobile communication services in these higher frequency bands.

1.2 Objectives of the work

Commensurate with the higher frequencies for future mobile systems are higher fading rates. Fading rates are proportional to both vehicle speed and carrier frequency; therefore, future systems could experience fading rates one or two orders of magnitude higher than those experienced by current wireless systems. Another key parameter is the symbol rate f_s; it classifies whether time variations of the channel are slow or fast. In terms of the symbol duration $T_s = 1/f_s$ the time variations are specified via the normalised

[1] Another notation commonly used in literature is reverse link for the uplink and forward link for the downlink.

Doppler frequency

$$\nu' = \nu T_s \tag{1.1}$$

So, for low transmission rates, there is fundamentally a higher time variability in the channel.

In this thesis, fast fading channels shall be considered, which are defined as channels where the phase of the fading process can vary significantly over a symbol period, i.e. where the symbol duration is a significant fraction of the coherence time of the channel [7]. This work will concentrate on channel estimation techniques in order to determine how the behaviour of communication systems are modified under such conditions.

In order to allow a meaningful comparison between competing techniques, normalised performance measures need to be defined. For example, it would not be very meaningful to compare the symbol error probability as a function of the signal-to-noise ratio (SNR) unless this comparison was made on the basis of a fixed bandwidth, or equivalently a fixed data rate. The definition of *spectral efficiency*

$$S = \frac{R_b}{W} \qquad [\text{bit/s} / \text{Hz}] \tag{1.2}$$

relates the transmission bitrate R_b to the required bandwidth W. On the physical layer, the receiver causes reduction of spectral efficiency due to bandwidth allocation for channel estimation. Furthermore a reduction of power efficiency is observed, in terms of the required increase in SNR, associated with imperfect reconstruction of the channel parameters. Another distinguishing factor in the receiver design is the computational complexity of different algorithms and their robustness in worst case conditions. We shall investigate ways to optimise both the spectral and power efficiency, while also addressing complexity and stability issues.

1.2.1 Separation principle

In the general case the received signal contains two random sequences from the receiver point of view: the useful data and the randomly varying channel parameters. Non-coherent detection techniques do not require information about the channel parameters. Therefore, they have no loss in spectral efficiency. However, the power efficiency may be poor, especially in a fast time-varying environment. On the other hand, coherent techniques, which may have favourable performance, cannot separate the two random elements of the received signal. For phase modulated signals with constant amplitudes, this results in a phase ambiguity at the receiver. This is termed the *phase ambiguity problem* and it necessitates use of a phase reference. Techniques to provide a phase reference are:

Differential encoding A differential encoded signal [7], $d(k)$, gets its phase reference from the last transmitted symbol, such that $d(k) = \Delta d(k) d(k-1)$. The k^{th} information bit is recovered by the demodulator: $\Delta d(k) = d(k) d^*(k-1)$. Since differential encoding is an non-coherent technique, there is no phase ambiguity in the received signal. The number of decision errors, however, is basically doubled. This is easily seen, because of the differential decoding, every error in $d(k-1)$ induces an error in the

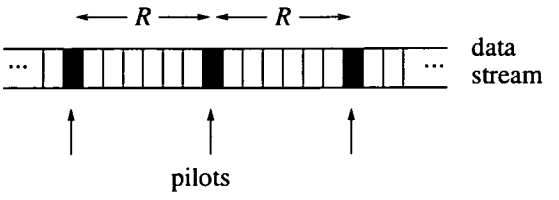


Figure 1.1: *Pilot symbol insertion technique.*

next sample through $d(k)$.

Pilot channel A coherent reference can be made available to the receiver by transmitting a sounding signal or pilot tone (see [8] and references therein). A common technique on the downlink for a mobile multiple access system is to allocate a pilot channel transmitting only known symbols. The pilot signal is shared for all users and is solely used for channel estimation. In a direct sequence (DS) CDMA system this would be a different spreading sequence than for any of the traffic channels. Both the information and pilot signal are transmitted through the same RF channel at the same frequency and they will therefore experience the same fading [9], i.e. the channel impulse response of the pilot and traffic channel will be identical. For example, a pilot channel is employed in the IS-95 standard. Its drawback is that the pilot channel requires about 20% of the total transmit power of the system.

Time multiplexed pilot symbols are another possibility to solve the separation problem. Here pilot symbols are periodically inserted in the data stream, giving the transmitted signal a frame or slot structure. This technique is used for the up- and downlink for UMTS. Another slot structure will be used in this thesis, proposed by Moher [10] and Cavers [11]. The basic idea is to multiplex one pilot symbol into the data stream by the rate $1 : R$, i.e. one pilot symbol is succeeded by $R-1$ data symbols, as illustrated in Figure 1.1. Thus, the loss in spectral efficiency is $1/R$.

1.2.2 Information theoretic system model

Before proceeding with studying receiver designs for the physical layer, it is instructive to look at the information theoretical approach, introduced by Shannon in 1948 [12]. Together with the availability of VLSI technology, the implementation of complex digital algorithms have become feasible. Viterbi [13] summarised the three basic lessons learned from information theory:

1.) Completely separate techniques for digital source compression (source coding) from those for channel transmission (channel coding), even though the first removes redundancy and the second inserts it.

2.) Never discard information prematurely that may be useful in making a decision until after all decisions related to that information have been completed.

3.) The performance of the transmission system is optimised if the signal is corrupted by white Gaussian noise only. In other words, the receiver of the physical layer's optimal strategy is to make the

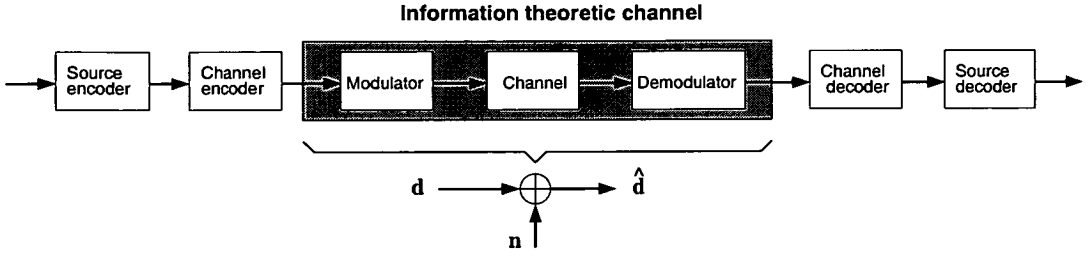


Figure 1.2: *Information theoretic system model.*

channel look like an additive white Gaussian noise (AWGN) channel.

These lessons immediately translate into the basic structure of the information theoretic system model, depicted in Figure 1.2. According to lesson 3, the task of the modulator/demodulator is to produce a sequence of symbols $\mathbf{d}$ that is seen at the outer receiver (the source and channel decoder) as a sequence, $\hat{\mathbf{d}}$, that is corrupted by white Gaussian noise, $\mathbf{n}$. To accomplish this, the inner transmission system (the physical layer) must be adapted to the channel characteristics, if the physical channel itself is not an AWGN channel. One implication is that a interleaver/deinterleaver should be used to provide errors which are uniformly distributed in time, while a fading channel tends to produce error bursts [7]. Another implication is the use of *diversity*, discussed in the next section.

It should be noted at this point, that according to lesson 2, the tasks of the inner receiver and the channel decoder should not be separated, instead they should be performed jointly.

1.2.3 Diversity techniques

Simple transmission of a carrier frequency over a time-variant multipath channel will suffer from fading effects which result in changes in signal power of as much as 20–30 dB [14]. If the channel attenuation is large, so that the channel is in a deep fade, the occurrence of errors can dramatically increase. Thus, transmitting over such a channel with a narrow-band signal alone is very inefficient. The idea of the diversity concept is to supply the receiver with several replicas of the same information signal, transmitted over independently fading channels. By virtue of its averaging effects *diversity* aids in bridging deep fades and hence approaches compliance with the information theoretic AWGN channel. The reason for this is, if the probability of a deep fade for one out of N independent fading component is p, then the probability that all N components are in a deep fade at the same time is p^N. This means a N–fold diversity gain is achieved. In theory, for an infinite number of diversity signals, the performance approaches transmission over an AWGN channel. A fundamental requirement for this, however, is that the CIR is known perfectly to the receiver.

Diversity may be provided in the form of *time*, *frequency*, *polarisation* or *spatial* diversity [2, 7]. These diversity techniques may be arbitrarily combined with each other. Multipath reception over a frequency selective fading channel is one form of frequency diversity, in which information flows from transmitter to receiver via *natural* diversity [15]. One key advantage of direct sequence CDMA systems is that they can exploit the inherent diversity of a frequency selective fading channel through the use of a

RAKE receiver [7].

1.2.4 Outline of the thesis

Work in this thesis is limited to the physical layer, and so source and channel coding will not be considered. It has been stated before that detection and channel decoding should be performed jointly for optimal performance, so references are given for research aiming to link the detection and decoding problem. Techniques to improve channel estimation and data detection of digital signals transmitted through a fast fading channel are investigated. The focus of the thesis is to minimise the system overhead, while optimising the performance; thus optimising both the spectral and power efficiency. In order to achieve that, algorithms for estimation and detection can be systematically derived based on a mathematical model and a chosen performance criterion. Chapter 2 reviews principles of the estimation and detection theory, providing some performance criteria. Estimators chosen by the MMSE criterion are described, resulting in Wiener filters for the linear case. The detection techniques considered are the MAP and ML criterion. The mathematical model of the system investigated is defined in Chapter 3. Fundamentals of direct sequence (DS) CDMA systems are discussed briefly, followed by a more detailed description of the multipath fading channel and its mathematical model used for simulation work in subsequent chapters. In Chapter 4, one-shot receiver structures are investigated. These receivers require signals which comply with the separation principle, to avoid the phase ambiguity problem. This is achieved by either time multiplexing pilot symbols into the data stream (section 4.2) or differential encoding (section 4.3). Chapter 5 surveys a more systematic approach, using joint channel estimation and data detection techniques. The optimum receiver which maximises the probability of detecting the whole sequence is given and a lower performance bound is derived. The optimum receiver is seen to be of prohibitive complexity, which necessitates sub-optimum derivatives. Receiver structures utilising the Viterbi algorithm, based on the principle of per-survivor processing (PSP) are discussed in Chapter 6. Employing PSP implies decision directed channel estimation to some extent, due to state dependent decision feedback. Thus stability issues are of paramount importance and they are thoroughly analysed. A performance bound based on a Markov model is analytically derived, explicitly taking these stability issues into account. Finally, the effects of multiple access interference (MAI) for receiver structures studied previously are investigated. Finally Chapter 7 draws the conclusions of the work.

Chapter 2
Principles of Bayesian Detection and Estimation

In this chapter the principles of detecting data and estimating parameters are addressed. It is a brief overview of techniques which are used in subsequent chapters, which are applied to digital signalling over a time-variant multipath fading channel. For a more thorough study of detection and estimation theory see for example the textbooks [16–20]. This chapter divides into two main parts, after a brief introduction to the Bayesian philosophy, parameter estimation based on the minimum mean squared error (MMSE) criterion will be considered in section 2.1. In particular, its optimum solution for linear problems, the Wiener filter will be discussed in more detail. The second part of this chapter is dedicated to Bayesian detectors in section 2.2. Detection based on the maximum *a posteriori* (MAP) and the maximum likelihood (ML) decision rules, is considered. These decision rules can be used to either maximise the probability of choosing the correct sequence or a particular data symbol.

The problem of estimation and detection is to extract values of parameters based on continuous time waveforms. Due to the use of digital computers to sample and store the continuous time waveform, the equivalent problem of extracting parameter values from a *discrete time* waveform or a *time series* is obtained. Mathematically, the observation is given by the K-point data set $\{y(1), \cdots, y(K)\}$ which depends on an unknown parameter, denoted by θ. This time series is more conveniently expressed in vector notation as $\mathbf{y} = [y(1), \cdots, y(K)]^T$. The objective of parameter estimation is to determine θ based on the data by means of an *estimator*: $\widehat{\theta} = f[\mathbf{y}]$, where $f(\cdot)$ is some function. The unknown parameter may itself depend on time or the data may be dependent on several parameters, cast in the vector $\boldsymbol{\theta}$.

In determining good estimators the first step is to statistically model the data. Because the data are inherently random, they are described by its *probability density function* (pdf). In *Bayesian* estimation/detection both the data $\mathbf{y}$ *and* the parameter $\boldsymbol{\theta}$ are assumed to be random variables. This is opposed to the *classical* estimation problem, where the parameters of interest are assumed to be deterministic but unknown. The Bayesian approach is so named because its implementation is based directly on Bayes' theorem. As such, the data are described by the *joint* pdf

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \tag{2.1}$$

where $p(\boldsymbol{\theta})$ is the *a priori* pdf, summarising the information about $\boldsymbol{\theta}$ before any data are observed, and $p(\mathbf{y}|\boldsymbol{\theta})$ is a conditional pdf providing information about the observation $\mathbf{y}$ conditioned on knowing $\boldsymbol{\theta}$. The pdf $p(\mathbf{y}|\boldsymbol{\theta})$ is also called the *likelihood function*. In order to estimate $\boldsymbol{\theta}$, the *a posteriori* pdf $p(\boldsymbol{\theta}|\mathbf{y})$ is of particular interest, since it provides information about $\boldsymbol{\theta}$ given the observation $\mathbf{y}$. The *a posteriori*

pdf is again linked with the joint pdf $p(\mathbf{y}, \theta)$ by Bayes' rule

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}, \theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\, p(\theta)}{p(\mathbf{y})} \tag{2.2}$$

So the *a posteriori* pdf is the product of the likelihood function $p(\mathbf{y}|\theta)$ and the *a priori* information $p(\theta)$ about the parameter θ, divided by $p(\mathbf{y})$. Note that the denominator is just a normalising factor, independent of θ.

In order to specify an estimator out of a given pdf, criteria need to be defined for which the estimator is aimed to be optimised. This is achieved by means of a cost function $C(\epsilon)$, where $\epsilon = \widehat{\theta} - \theta$ denotes the estimation error between the parameter θ and its estimate $\widehat{\theta}$. An estimator which minimises the *risk* $\mathrm{E}[\, C(\epsilon)\,]$ is then the optimum estimator. For a quadratic cost function $C(\epsilon) = \epsilon^2$, the so called *minimum mean squared error* (MMSE) estimator minimises the risk. Considering the noise filtering problem this yields the *Wiener filter* in the linear case, which will be discussed in more detail in the following section.

Thus far, only the estimation problem has been considered. Estimation theory is concerned with the problem of finding the best value for an unknown parameter from a continuum of possible values. For detection, on the other hand, data from a finite set $\mathcal{D}$ of discrete values is to be extracted. Hence, the detection outcome may be seen as a random variable taken from the constraint set $\mathcal{D}$. For detection, choosing another cost function $C(\epsilon)$, the Bayesian approach leads to the *maximum a posteriori* (MAP) decision rule. That is, it assigns no cost if the received data is within a certain decision boundary and a cost of 1 for all ϵ in excess of this cost function. Decision rules which can be derived from the MAP criterion are addressed in section 2.2.

2.1 MMSE parameter estimation

One desirable property of an estimator is that it is *unbiased*. An estimator is unbiased, if on the average the estimator will yield the true value of the unknown parameter. Mathematically, an estimator is unbiased if

$$\mathrm{E}[\,\widehat{\theta}\,] = \theta \tag{2.3}$$

That an estimator is unbiased does not necessarily mean that it is a good estimator, nor is a biased estimator necessarily a poor one. In searching for optimal estimators some optimality criterion needs to be adopted. A natural one is mean squared error (MSE), defined as

$$C(\widehat{\theta} - \theta) = \mathrm{E}\Big[\, |\widehat{\theta} - \theta|^2 \,\Big] \tag{2.4}$$

Minimising the MSE with respect to $\widehat{\theta}$ yields the MMSE solution. The MMSE criterion will prove extremely useful in the receiver design in the following chapters and is widely used in many areas of digital signal processing. Since for a Bayesian estimator θ is a random variable, the expectation operator is with respect to the *joint* pdf $p(\mathbf{y}, \theta)$. The optimum estimator in the MMSE sense is obtained by

differentiating (2.4) and setting the result equal to zero, that yields [16]

$$\widehat{\theta} \;=\; \int \theta\, p(\theta|\mathbf{y})\, d\theta \;=\; \mathrm{E}[\,\theta|\mathbf{y}\,] \tag{2.5}$$

It is seen that the optimum estimator is the mean of the *a posteriori* pdf $p(\theta|\mathbf{y})$, which refers to θ after the data have been observed. The *a posteriori* pdf is linked with the joint pdf $p(\mathbf{y},\theta)$ by Bayes' rule in (2.2).

2.1.1 Jointly Gaussian distribution

If the pdf $p(\mathbf{y},\theta)$ is a joint Gaussian pdf then a closed form of the estimator and the corresponding MSE is readily available. Let the multivariate Gaussian pdf be defined by

$$p(\mathbf{y},\theta) = \frac{1}{2\pi\,\det^{1/2}(\mathbf{\Phi})} \exp\left(- \begin{bmatrix} \mathbf{y} - \mathrm{E}[\mathbf{y}] \\ \theta - \mathrm{E}[\theta] \end{bmatrix}^T \mathbf{\Phi}^{-1} \begin{bmatrix} \mathbf{y} - \mathrm{E}[\mathbf{y}] \\ \theta - \mathrm{E}[\theta] \end{bmatrix} \right) \tag{2.6}$$

where det denotes the matrix determinant operation. The covariance matrix $\mathbf{\Phi}$ is given by

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{\mathbf{yy}} & \mathbf{\Phi}_{\mathbf{y}\theta} \\ \mathbf{\Phi}_{\theta\mathbf{y}} & \mathbf{\Phi}_{\theta\theta} \end{bmatrix}$$

The properties of the covariance matrices for the special case of a linear model will be discussed later on. In this case the conditional pdf $p(\theta|\mathbf{y})$ is also Gaussian with the conditional mean and corresponding MSE

$$\widehat{\theta} \;=\; \mathrm{E}[\,\theta|\mathbf{y}\,] \;=\; \mathrm{E}[\theta] + \mathbf{\Phi}_{\theta\mathbf{y}}\mathbf{\Phi}_{\mathbf{yy}}^{-1}(\mathbf{y} - \mathrm{E}[\mathbf{y}]) \tag{2.7}$$

$$\mathbf{\Phi}_{\epsilon\epsilon} \;=\; \mathbf{\Phi}_{\theta|\mathbf{y}} \;=\; \mathbf{\Phi}_{\theta\theta} - \mathbf{\Phi}_{\theta\mathbf{y}}\mathbf{\Phi}_{\mathbf{yy}}^{-1}\mathbf{\Phi}_{\mathbf{y}\theta} \tag{2.8}$$

where $\mathrm{E}[\,\theta|\mathbf{y}\,] = \widehat{\theta}$ is the MMSE estimate of θ and $\mathbf{\Phi}_{\theta|\mathbf{y}} = \mathbf{\Phi}_{\epsilon\epsilon}$ is the covariance matrix of the estimation error $\epsilon = \widehat{\theta} - \theta$.

2.1.2 Wiener filters

Under the jointly Gaussian assumptions the optimum estimator is easily found; in the general case of a non-Gaussian joint pdf, however, they are not. If the Gaussian assumption does not hold, the MMSE criterion leads to non-linear estimators which may be very complex. In this case, the MMSE criterion can be retained under the constraint that the estimator is *linear*. Then an explicit estimator can be determined which depends only on the first two moments of the pdf which are given by(2.7) and (2.8), respectively. Although the filter is sub-optimum it is the best *linear* filter in the MSE sense. This class of filters are generally termed *Wiener filters* [21] and they are extensively utilised in signal processing applications.

The vector form of the linear model is defined by

$$\mathbf{y} = \mathbf{D}\mathbf{h} + \mathbf{n} \tag{2.9}$$

where $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{n}$ represent discrete time waveforms of length K, defined by $\mathbf{x} = [x(1), \cdots, x(K)]^T$, with $\mathbf{x}$ being a placeholder for $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{n}$. The problem is to extract $\mathbf{h}$ from $\mathbf{y}$. In digital communication the linear model is extensively used to describe a received base-band signal $\mathbf{y}$. The random vector $\mathbf{h}$ describes the distortions by the channel which is to be estimated. The matrix $\mathbf{D}$ is a known $K \times K$ transformation matrix and $\mathbf{n}$ is a noise vector with zero mean. The quantities $\mathbf{h}$ and $\mathbf{n}$ are assumed to be uncorrelated. The MMSE criterion applied to this problem is to minimise the estimation error $\boldsymbol{\epsilon} = \mathbf{h} - \widehat{\mathbf{h}}$. Then the linear MMSE estimator $\widehat{\mathbf{h}}$ and the corresponding covariance $\boldsymbol{\Phi}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$ are given by (2.7) and (2.8) respectively, where

$$\begin{aligned} \mathrm{E}[\mathbf{y}] &= \mathbf{D}\,\mathrm{E}[\mathbf{h}] \\ \boldsymbol{\Phi}_{\mathbf{yy}} &= \mathbf{D}\,\boldsymbol{\Phi}_{\mathbf{hh}}\,\mathbf{D}^H + \boldsymbol{\Phi}_{\mathbf{nn}} \\ \boldsymbol{\Phi}_{\mathbf{hy}} &= \boldsymbol{\Phi}_{\mathbf{yh}}^H = \boldsymbol{\Phi}_{\mathbf{hh}}\,\mathbf{D}^H \end{aligned} \tag{2.10}$$

The properties of the estimator are summarised by the Gauss-Markov theorem [16], which states:

(i.) Only the knowledge of the mean and covariance of $\mathbf{h}$ and $\mathbf{n}$ are required, the joint pdf $p(\mathbf{y}, \mathbf{h})$ is otherwise arbitrary.

(ii.) The estimator is unbiased, i.e. $\mathrm{E}[\,\widehat{\mathbf{h}}\,] = \mathbf{h}$.

(iii.) The performance of the estimator is measured by the error $\boldsymbol{\epsilon} = \mathbf{h} - \widehat{\mathbf{h}}$ whose covariance $\boldsymbol{\Phi}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = \boldsymbol{\Phi}_{\mathbf{h}|\mathbf{y}}$ is given by (2.8). The diagonal elements of $\boldsymbol{\Phi}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$ yield the minimum MSE of particular samples of $\boldsymbol{\epsilon}$, which is for the k^{th} entry $\{\boldsymbol{\Phi}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}\}_{kk} = \mathrm{E}[\,|\epsilon(k)|^2\,]$. There is no linear estimator which has a lower MSE, regardless of the distribution of the error vector.

(iv.) If the noise vector $\mathbf{n}$ is zero mean, the estimate of $\mathbf{y}$ is given by: $\widehat{\mathbf{y}} = \mathbf{D}\widehat{\mathbf{h}}$.

(v.) An interesting property of the linear MMSE estimator is that the estimation error is orthogonal to the observations, thus the mean of their inner product is zero: $\mathrm{E}[\,\boldsymbol{\epsilon}^H\mathbf{y}\,] = 0$. Since the estimate $\widehat{\mathbf{h}}$ itself is a linear combination of $\mathbf{y}$, it is also orthogonal to $\boldsymbol{\epsilon}$. This is known as *the principle of orthogonality*, it is equivalent to $\widehat{\mathbf{h}}$ and $\boldsymbol{\epsilon}$ being uncorrelated, which is another interpretation of $\mathrm{E}[\,\boldsymbol{\epsilon}^H\widehat{\mathbf{h}}\,] = 0$.

In the following discussion of Wiener filters some further assumptions for $\mathbf{h}$ and $\mathbf{y}$ are made. For the definition of the noise process, we assume a band-limited AWGN (additive white Gaussian noise) channel with bandwidth W. The power of the noise is therefore $\phi_{nn}(0) = \sigma^2 = N_0 W$ in W, where N_0 denotes the spectral power density of the noise process in W/Hz. With $W \triangleq 1$ Hz the variance σ^2 is normalised with respect to the bandwidth, yielding $\sigma^2 = N_0$. Since an AWGN process is uncorrelated, i.e. $\phi_{nn}(i-j) = 0$ for $i \neq j$, the covariance matrix is a diagonal matrix given by $\boldsymbol{\Phi}_{\mathbf{nn}} = N_0\mathbf{I}$.

It is assumed that the observation $\mathbf{y}$ is a wide sense stationary (WSS) stochastic process with zero mean, hence $\mathrm{E}[\mathbf{h}] = \mathrm{E}[\mathbf{y}] = \mathbf{0}$. As such the complex $K \times K$ covariance matrix takes the *Hermitian*

and *Toeplitz* form[1]

$$\mathbf{\Phi_{uv}} \triangleq \mathrm{E}\left[\,\mathbf{u}\mathbf{v}^H\,\right] = \begin{bmatrix} \phi_{uv}(0) & \phi_{uv}(1) & \cdots & \phi_{uv}(K) \\ \phi_{uv}(-1) & \phi_{uv}(0) & \cdots & \phi_{uv}(K-1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{uv}(-K) & \phi_{uv}(1-K) & \cdots & \phi_{uv}(0) \end{bmatrix} \tag{2.11}$$

with the coefficients $\phi_{uv}(\Delta k) = \mathrm{E}[\,u(k-\Delta k)\,v^*(k)\,]$ for the time delay Δk. That is the cross-correlation function (CCF) of two signal sequences $u(k-\Delta k)$ and $v(k)$. If $u(k) = v(k)$, then $\mathbf{\Phi_{uu}}$ denotes the auto-correlation matrix and $\phi_{uu}(\Delta k)$ is the corresponding auto-correlation function (ACF) of a signal sequence with time delay Δk. The variables $\mathbf{u}$ and $\mathbf{v}$ are placeholders for combinations of $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{n}$. For the further discussion in this section, the known transformation matrix $\mathbf{D}$ is assumed to be a $K \times K$ identity matrix, $\mathbf{D} = \mathbf{I}$. Hence the notation for the covariance matrices of (2.10) simplifies to

$$\mathbf{\Phi_{yy}} = \mathbf{\Phi_{hh}} + N_0\mathbf{I} \qquad \text{and} \qquad \mathbf{\Phi_{hy}} = \mathbf{\Phi_{hh}} \tag{2.12}$$

There are two main problems that will be studied: smoothing and prediction.

Smoothing is considered first. Estimation of $h(k)$ involves the whole data set $\{y(1), \cdots, y(K)\}$, consisting of past, present and future samples. Clearly, for smoothing an estimate cannot be obtained until the entire sequence has been received.

For linear MMSE smoothing, (2.7) can directly be applied, yielding the solution

$$\widehat{\mathbf{h}} \;=\; \mathbf{\Phi_{hy}}\mathbf{\Phi_{yy}^{-1}}\,\mathbf{y} \;=\; \mathbf{W}\,\mathbf{y} \tag{2.13}$$

where the $K \times K$ filter bank $\mathbf{W}$ is termed the *Wiener smoothing matrix.* The Wiener-Hopf equation for the set of estimation errors, $\boldsymbol{\epsilon} = \mathbf{h} - \widehat{\mathbf{h}}$, is in the form

$$\begin{aligned} \mathbf{W} \;&=\; \mathbf{\Phi_{hh}}\,\mathbf{\Phi_{yy}^{-1}} \;=\; \left[\mathbf{\Phi_{yy}} - N_0\mathbf{I}\right]\mathbf{\Phi_{yy}^{-1}} \\ &=\; \mathbf{I} - N_0\mathbf{\Phi_{yy}^{-1}} \end{aligned} \tag{2.14}$$

The Wiener filter matrix $\mathbf{W}$ is Hermitian, since it consists of a superposition of Hermitian matrices.

Prediction is the application of a causal filter to the data. For x–step prediction $h(k+x)$ is estimated, based on the data set $\{y(1), \cdots, y(k)\}$. Similar to (2.5), the MMSE estimate of $h(k+x)$ is the conditioned mean of the observed data up to sample k, that is

$$\widehat{h}(k+x) = \mathrm{E}\left[\,h(k+x)\big|y(1), \cdots, y(k)\,\right]. \tag{2.15}$$

[1] A matrix is referred to as Hermitian if the matrix is equal to its complex conjugate transposed, denoted by $\mathbf{\Phi_{uv}} = \mathbf{\Phi_{uv}^H} = \mathbf{\Phi_{vu}}$. A matrix is referred to as Toeplitz if all elements on its main diagonal are equal, and if the elements on any other diagonal parallel to the main diagonal are also equal.

For $x \geq 1$ only past samples are used, while for $x = 0$ past and present samples are used. The case when $x = 0$ is referred to as the *filtering* problem. Clearly, the larger x the larger becomes the MSE of the predictor, so the causality is traded with a poorer estimator, compared to the smoother.

The estimator to predict $\mathbf{h}$ for 1-step linear prediction can be obtained through triangular decomposition [22]. The derivation is based on the factorisation $\mathbf{\Phi_{yy}} = \mathbf{L\Sigma L}^H$ where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{\Sigma}$ is real diagonal matrix. This factorisation does exist since $\mathbf{\Phi_{yy}}$ is a Hermitian symmetric matrix. Then the estimation error becomes $\boldsymbol{\epsilon} = \mathbf{L}^{-1}\mathbf{y}$. The rows of $\mathbf{L}^{-1}$ are the coefficients of a 1-step predictor for orders 0 through $K-1$ and the elements of $\mathbf{\Sigma}$ are the corresponding error covariances, thus $\{\mathbf{\Sigma}\}_{kk} = \sigma_k^2 = \mathrm{E}[\,|\epsilon(k)|^2\,] + N_0$, i.e. the diagonal matrix $\mathbf{\Sigma}$ contains the MSE for 1-step prediction plus the variance of AWGN. Considering the k^{th} element, the prediction error $\epsilon(k)$ may be viewed as an *innovations process*. The innovations approach [23] is an interpretation of 1-step prediction. According to the principle of orthogonality [18] $\epsilon(k)$ is orthogonal to all past observations $\{y(1), \cdots, y(k-1)\}$ and may therefore be regarded as a measure for new information in the random variable $h(k)$ at time k; hence the name innovation process. Thus, the part of $h(k)$ which is new is contained in the prediction error $\epsilon(k)$.[2] The matrix $\mathbf{L}^{-1}$ may be regarded as a transformation matrix which transforms the observation $\mathbf{y}$ into the innovation $\boldsymbol{\epsilon} = \mathbf{L}^{-1}\mathbf{y}$.

The MMSE estimate of $h(k)$ is very closely related to $\epsilon(k)$, such that the k^{th} row of $\mathbf{L}^{-1}$ contains the coefficients of a $(k-1)^{\text{th}}$ order linear predictor for the k^{th} sample: $\{-w_1^{k-1}, \cdots, -w_{k-1}^{k-1}, 1\}$, where the superscript $k-1$ denotes the filter order. Thus, the following 1-step predictor is obtained

$$\widehat{h}(k) \;=\; \sum_{m=1}^{k-1} \left[w_m^{k-1}\right]^* \, y(k-m)$$

where the filter $\mathbf{w}^{(k)} = [w_1^k, \cdots, w_k^k]^T$ represents the 1–step predictor of order k.

2.1.3 Moving average estimation

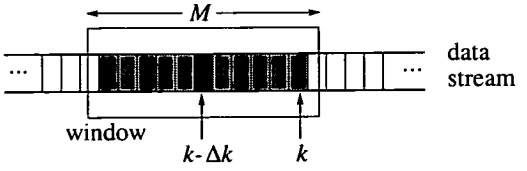
For a long sequence length K neither optimal smoothing nor prediction using all available data is feasible due to the large matrix sizes. Furthermore, optimal estimation implies a time-variant filter response, thus the filter bank $\mathbf{W}$ in (2.14) is not a Toeplitz matrix. A common approach to reduce the filter length, is to assume that $h(k)$ is a *auto-regressive* (AR) process of order M [24]

$$h(k-\Delta k) \;=\; n_p(k) + \sum_{m=1}^{M} w_m^*(\Delta k)\, h(k-m+1) \tag{2.16}$$

where the process noise $n_p(k)$ is complex valued white Gaussian noise. The coefficients $\{w_m(\cdot)\}$ specify the AR process. Now only M snapshots generate $h(k)$, as illustrated in Figure 2.1. If $\Delta k < 0$ the CIR is causal, specifying an AR process for an x–step predictor, with $x = -\Delta k$. Otherwise, for $\Delta k \geq 0$, a non-causal AR process is obtained. Given an AR process the transfer function of the CIR $H(z)$ (the

[2] Because $\epsilon(k)$ is a *whitened* transformation of $y(k)$, the innovations approach is also referred to the *whitening-filter* method. However, this latter term is used only in the context of linear processing of stationary and Gaussian processes; whereas the term innovations approach is applicable to non-linear processing and non-Gaussian processes as well.

(a.) **Smoother**



(b.) **Linear Predictor**

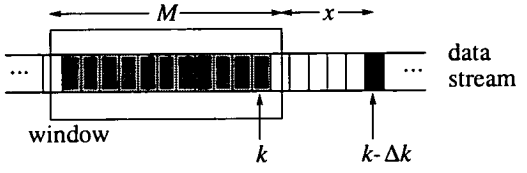


Figure 2.1: *Snapshots for the time frame of an (a.) smoother and (b.) x–step predictor.*

z–transform of $h(k)$) consists of poles only, with M being the number of poles. Given an M^{th} order AR model, the transfer function $H^{-1}(z)$ specifies a M^{th} order moving average (MA) filter [24, 25]. Now only M snapshots are used to estimate $h(k)$, giving the MA filter

$$\widehat{h}(k-\Delta k) \;=\; \sum_{m=1}^{M} w_m^*(\Delta k)\, y(k-m+1) \;=\; \mathbf{w}^H(\Delta k)\, \mathbf{y}(k) \tag{2.17}$$

where the filter $\mathbf{w}(\Delta k) = [w_1(\Delta k), \cdots, w_M(\Delta k)]^T$ corresponds to the AR model in (2.16), being either a smoother ($0 < \Delta k < M$) or a x-step predictor ($\Delta k < 0$), with $x = -\Delta k$, as illustrated in Figure 2.1. The vector $\mathbf{y}(k) \triangleq [y(k), \cdots, y(k-M+1)]^T$ is a sliding window of length $M \ll K$. The incentive in assuming an AR model, is that the optimal filter is an FIR filter with finite length M [24]. If $h(k)$ is not an AR process, the MA filter in (2.17) is an approximation of the optimal filters of length K. The accuracy of the approximation is dependent on the choice of M.

In practice, the the filter $\mathbf{w}(\Delta k)$ is usually unknown and needs to be estimated. In order to determine the tap weights for an arbitrarily distributed CIR, consider the cost function

$$C\big[\mathbf{w}(\Delta k)\big] = \big|h(k) - \mathbf{w}^H(\Delta k)\, \mathbf{y}(k)\big|^2$$

Minimising the risk $\mathrm{E}\big[\, C\big[\mathbf{w}(\Delta k)\big]\,\big]$ is the MMSE criterion for the M^{th} order MA filter, $\mathbf{w}(\Delta k)$. One possibility to minimise the risk is to apply the principle of orthogonality [18], which states that $\mathrm{E}[\, \epsilon^H \mathbf{y}\,] = 0$. The solution is the Wiener-Hopf equation, which comprises both the smoothing and prediction case, having the form

$$\mathbf{w}(\Delta k) = \mathbf{\Phi}_{\mathbf{yy}}^{-1}\, \boldsymbol{\phi}_{h\mathbf{y}}(\Delta k) \tag{2.18}$$

where $\boldsymbol{\phi}_{y\mathbf{y}}(\Delta k) = \mathrm{E}[\, h^*(k-\Delta k)\, \mathbf{y}(k)\,]$ denotes the cross-correlation vector of the time delay Δk. Unless otherwise stated, all vector and matrix quantities will be of dimension M and $M \times M$, respectively.

The MMSE of the estimator $\mathbf{w}(\Delta k)$ becomes

$$\begin{aligned} V_{min}^2(\epsilon, \Delta k) &\triangleq \mathrm{E}\big[\,|\epsilon(k)|^2\,\big] = \mathrm{E}\big[\,|h(k) - \mathbf{w}^H(\Delta k)\,\mathbf{y}(k+\Delta k)|^2\,\big] \\ &= 1 - \boldsymbol{\phi}_{h\mathbf{y}}^H(\Delta k)\,\boldsymbol{\Phi}_{\mathbf{yy}}^{-1}\,\boldsymbol{\phi}_{h\mathbf{y}}(\Delta k) \end{aligned} \tag{2.19}$$

2.1.4 Method of least squares

So far, it has been assumed that the second order statistics in terms of the auto and cross-correlation matrices are perfectly known. In a practical situation this may not be the case, so these statistics need to be estimated. The method of least squares solves the linear filtering problem without invoking assumptions on the statistics of the inputs applied to the filter. Instead there is a given set of say N complex valued measurements $\{y(1), \cdots, y(N)\}$ made at times $\{t_1, \cdots, t_N\}$ and the requirement is to fit these points in some optimum fashion. The criterion minimises the least squares between the desired response and the observed data set. Its solution is very similar to the Wiener-Hopf equation in (2.18), in the sense that the exact covariance matrices $\boldsymbol{\Phi}_{\mathbf{uv}}$ in (2.11) are replaced by estimates $\hat{\boldsymbol{\Phi}}_{\mathbf{uv}}$.[3] Specifically, the coefficient of m^{th} row and n^{th} column is given by

$$\begin{aligned} \mathrm{E}[\,u(k)\,v^*(k-\Delta k)\,] &\approx \big\{\hat{\boldsymbol{\Phi}}_{\mathbf{uv}}\big\}_{mn} = \hat{\phi}_{uv}(\underbrace{m-n}_{\Delta k}) \\ &\approx \frac{1}{N-\Delta k}\sum_{k=\Delta k}^{N} u(k)\,v^*(k-\Delta k) \end{aligned}$$

So far, it has been assumed that the desired response $h(k)$ is available to generate the estimation error ϵ. In many applications however, the receiver does not have access to the desired signal $h(k)$, only to the noise corrupted received signal $y(k) = h(k) + n(k)$. Hence, the auto and cross-correlation matrices $\hat{\boldsymbol{\Phi}}_{\mathbf{hh}}$ and $\hat{\boldsymbol{\Phi}}_{\mathbf{hy}}$ are not available, whereas the receiver may generate estimates of $\hat{\boldsymbol{\Phi}}_{\mathbf{yy}}$ and $\hat{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k)$. The auto-correlation vector $\hat{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k) = \big[\hat{\phi}_{yy}(\Delta k), \cdots, \hat{\phi}_{yy}(\Delta k - M + 1)\big]^T$ denotes the Δk^{th} row of $\hat{\boldsymbol{\Phi}}_{\mathbf{yy}}$. With (2.12), the ACF of $y(k)$ can be expressed as $\phi_{yy}(\Delta k) = \phi_{hy}(\Delta k) + N_0\,\delta_{\Delta k}$, where $\delta_{\Delta k}$ is the Kronecker δ–function. Thus, for $\Delta k < 0$ (the linear predictor case) we have $\hat{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k) \approx \hat{\boldsymbol{\phi}}_{h\mathbf{y}}(\Delta k)$. This is the case for a linear predictor. In this instance, the estimator may be generated on the noisy signal snapshots $y(k)$. Thus, in the decision for the estimation error, $h(k)$ is replaced with $y(k) = h(k) + n(k)$, yielding $\tilde{\epsilon}(k) = \hat{h}(k) - y(k)$. As $n(k)$ is zero mean, the estimate is still unbiased, while its variance will increase due to additional noise. A smoother is on the other hand not implementable when the receiver has to estimate the channel statistics. It can however be modified, in the way that *present* samples (with time delay Δk) are omitted in (2.18) for calculating $\mathbf{w}(\Delta k)$. This can be achieved by cancelling the Δk^{th} row and column of $\hat{\boldsymbol{\Phi}}_{\mathbf{yy}}$ and the Δk^{th} element of $\hat{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k)$ and replace them by zeros, yielding $\bar{\boldsymbol{\Phi}}_{\mathbf{yy}}(\Delta k)$ and $\bar{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k)$, respectively. The smoothing filter is then determined by

$$\begin{aligned} \bar{\mathbf{w}}(\Delta k) &= \big[\bar{w}_1(\Delta k), \cdots, \bar{w}_{\Delta k-1}(\Delta k),\, 0,\, \bar{w}_{\Delta k+1}(\Delta k), \cdots, \bar{w}_M(\Delta k)\big]^T \\ &= \bar{\boldsymbol{\Phi}}_{\mathbf{yy}}^{-1}(\Delta k)\,\bar{\boldsymbol{\phi}}_{\mathbf{yy}}(\Delta k) \end{aligned} \tag{2.20}$$

[3] The placeholders $\mathbf{u}$ and $\mathbf{v}$ take on the values $\mathbf{y}$ or $\mathbf{h}$ and combinations of the two.

Obviously, for prediction type filters with $\Delta k < 0$, the covariances $\bar{\Phi}_{\mathbf{yy}}(\Delta k)$ and $\bar{\phi}_{yy}(\Delta k)$ are identical to $\hat{\Phi}_{\mathbf{yy}}$ and $\hat{\phi}_{yy}(\Delta k)$ respectively.

The filter $\bar{\mathbf{w}}(\Delta k)$ can be generated adaptively, for instance by means of the adaptive least mean squares algorithm (LMS) or the recursive least squares (RLS) [18]. Generally, the RLS algorithm offers superior performance at the expense of significantly increased complexity compared to the LMS algorithm. With these adaptive techniques the matrix inversion required to compute $\bar{\mathbf{w}}(\Delta k)$ can be avoided, furthermore adaptive filters can cope with a non-stationary channel. Adaptive algorithms are distinguished between their performance, in terms of convergence time and the MSE, and their computational complexity. A comparison of the conventional least squares approach with the LMS and RLS algorithms, for fast fading channel estimation was studied in [26]. In general, the RLS algorithm offers better performance, while the LMS algorithm is easier to implement.

2.1.5 IIR-type filtering

Another possibility is to predict $\widehat{h}(k+1)$ dependent on the observations $\mathbf{y}(k)$ *recursively*, i.e. using information about the previous estimates. This type of filter is a generalised form of the Wiener filter, termed *Kalman filter* [27]. It is an application of the innovations approach to non-stationary second order processes. Its significance is the ability to accommodate vector signals and noises which additionally may be non-stationary. The theory of Kalman filters is studied in the textbooks e.g. [16, 18, 20]. If the Gaussian assumption holds, Kalman filters are optimum in the MMSE sense. Otherwise they are the optimum linear estimator, given the linear model in (2.9).

In order to keep the complexity to a minimum only the 1st order IIR filter will be considered. Given a WSS channel, the the 1st order IIR filter is specified by a real valued, scalar parameter α, independent of time. The recursive channel estimator can be expressed as[4]

$$\widehat{h}(k+1) = (1-\alpha)\, y(k) + \alpha\, \widehat{h}(k)\,; \qquad 0 \le \alpha < 1\,. \tag{2.21}$$

The filter has the form of a low-pass filter, thus it reduces the effects of noise at the expense of some imposed pass-band distortions. Thus, for rapid fluctuations of $h(k)$, its estimate $\widehat{h}(k+1)$ experiences a phase lag, which degrades the filter performance. This implies that the actual channel dynamics, in particular a strictly band-limited fading process $\{h(k)\}$, can only be represented with sufficient accuracy by a high-order Gauss-Markov process model, which can be achieved by the use of Kalman filters.

In [28] the gain α is approximated, by replacing the MMSE with a simpler criterion. The MSE is separated into two components: a noise error, σ_n^2, which accounts for the gradient noise inherent to the stochastic-gradient based LMS algorithm, and the lag error, σ_φ^2, which accounts for channel dynamics:

$$V^2 \;=\; \mathrm{E}\Big[\,\big|h(k)-\widehat{h}(k)\big|^2\,\Big] \;=\; \sigma_\varphi^2 + \sigma_n^2$$

[4]The filter is commonly described by the well known least mean square (LMS) adaptive algorithm [18]. In some publications this filter is also referred to as α–tracker.

Generally, σ_φ^2 is a function of the channel dynamics, dependent on the parameter ν'_{max}, which is the maximum normalised Doppler frequency,[5] whereas the gradient noise σ_n^2 is a dependent on N_0.

The gain factor α is to be chosen to optimise the filter design, dependent on ν'_{max} and N_0. Generally speaking, a large α reduces the impact of AWGN on the MSE, but increases the *lag error* induced by the phase lag of the filter, and vice versa. The MMSE V_{min}^2 is approximated if $\sigma_n^2 = \sigma_\varphi^2$. In terms of the fading rate, ν'_{max}, and the noise power N_0, the following approximation is obtained [28]

$$\alpha_{opt} \approx 1 - 3.6\sqrt[3]{\frac{\nu'^2_{max}}{N_0}} \; ; \quad \frac{\nu'^2_{max}}{N_0} \leq 0.01 \tag{2.22}$$

This approximation is based on the assumption that the lag error, σ_φ^2 is an uncorrelated noise process. Since σ_φ^2 is coloured, this approximation is only valid if thermal noise dominates the effects of the channel estimation error. The white noise approximation breaks down for $\nu'^2_{max}/N_0 \leq 0.01$ [28], which effectively means modest fading rates and dominant noise power. For larger fading rates, α_{opt} in (2.22) becomes negative. In this case, setting $\alpha_{opt} = 0$ is the best possible solution. Then $\widehat{h}_q(k+1) = y(k)$, which essentially means the received signal is not filtered at all. Instead only the previous sample is used, which is similar to conventional differential detection. Despite these restrictions, using this approximation for (2.22), α_{opt} is much easier to derive analytically, compared to the derivation using the true MMSE. Furthermore, the system performance is not critically dependent on α, thus the approximation in (2.22) will be used for simulation work in subsequent chapters, bearing in mind that this may be not the optimum choice for α. Finally it should be noted, that α_{opt} can also be calculated adaptively [29, 30].

2.2 Bayesian detectors

In this section the detection problem is considered, specifically a sequence of K data symbols cast in the vector $\mathbf{d} = [d(1), \cdots, d(K)]^T$ transmitted over a radio channel, resulting in the received signal sequence $\mathbf{y} = [y(1), \cdots, y(K)]^T$. The k^{th} information symbol is taken from a finite discrete set $\mathcal{D}$, the symbol alphabet. Let the cardinality of the alphabet be A_m, then the receiver has to pick one out of A_m possibilities or hypothesis, according to some optimality criterion or decision rule. The objective is to minimise the probability of error according to some hypothesis criterion:

- This could either be the probability of a symbol or sequence error.
- The *hypothesis criteria* may be the maximum *a posteriori* (MAP) or the maximum likelihood (ML) decision rule.
- The received signal sequence may either be deterministic or of random nature, the former resulting in a one-shot receiver structure. For the latter case the whole sequence needs to be taken into account for optimum detection.

[5] The channel model is characterised in more detail in section 3.2.

2.2.1 Detection of correlated signals

The detection of random signals is considered first. In Bayesian detection the data $\mathbf{d}$ is regarded as random variable, described by the joint pdf $p(\mathbf{y}, \mathbf{d}) = p(\mathbf{y}|\mathbf{d})\, p(\mathbf{d})$, according to (2.1). Applying the MAP criterion to $p(\mathbf{y}, \mathbf{d})$ yields the MAP sequence detector (MAP–SD). The MAP criterion maximises the probability that the sequence $\widehat{\mathbf{d}}$ was transmitted, conditioned on the observation $\mathbf{y}$. A sequence hypothesis, denoted by $\mathbf{d}^{(\ell)}$, is the MAP estimate of the actual transmitted sequence $\mathbf{d}$, if its *a posteriori* probability is maximum

MAP–SD:

$$\widehat{\mathbf{d}} = \arg\max_{\ell \in \mathcal{A}_K} p(\mathbf{d}^{(\ell)}|\mathbf{y}) = \arg\max_{\ell \in \mathcal{A}_K} \frac{p(\mathbf{y}|\mathbf{d}^{(\ell)})\, p(\mathbf{d}^{(\ell)})}{p(\mathbf{y})} \tag{2.23}$$

where $\mathcal{A}_K$ is the total number of possible symbol sequences, which grows exponentially with the sequence length K. Note that the pdf of the observation $p(\mathbf{y})$ is independent of the transmitted sequence and can therefore be neglegted to find the maximum of (2.23). For uncoded signals with a A_m-ary symbol alphabet $\mathcal{A}_K$ equals A_m^K. The second equality follows from Bayes' rule in (2.2), where terms independent of $\mathbf{d}^{(\ell)}$ have been neglected.

On the other hand, maximising the likelihood function $p(\mathbf{y}|\mathbf{d}^{(\ell)})$ with respect to hypothesis ℓ yields the maximum likelihood sequence detector (MLSD)

MLSD:

$$\widehat{\mathbf{d}} = \arg\max_{\ell \in \mathcal{A}_K} p(\mathbf{y}|\mathbf{d}^{(\ell)}) \tag{2.24}$$

The resulting receiver was proposed by the pioneering work of Kailath [31–34] in the 1960s. Later recursive solutions have been proposed based on the innovations approach [35–37].

It is seen that the MAP and ML criteria are closely related. The difference between the two is that the MAP criterion incorporates *a priori* information $p(\mathbf{d}^{(\ell)})$ about the transmitted sequence. If all sequences are equally likely, as it is most often for uncoded systems, the *a priori* probability is simply $p(\mathbf{d}^{(\ell)}) = 1/\mathcal{A}_K$, independent of ℓ. In this case the MAP and ML criteria are equivalent.

Next symbol-by-symbol detection is considered. The objective is to calculate soft *a posteriori* probabilities for the symbol hypothesis ℓ at time k. The MAP estimate is

MAP symbol-by-symbol detection:

$$p(\widehat{d}(k)|\mathbf{y}) = \max_{\ell \in \mathcal{D}} p(d^{(\ell)}(k)|\mathbf{y}) \tag{2.25}$$

A solution was given by Bahl *et al.* [38]. The resulting receiver structure requires a much higher computational complexity than a corresponding sequence detector. When used for detection of uncoded signals or decoding a single stage code, the performance improvement of MAP symbol-by-symbol detection (MAP–SbSD) compared to MLSD is insignificant, and certainly insufficient to justify the increase in complexity. With the development of *iterative* decoding techniques, MAP–SbSD has become more popular since it offers the possibility of *soft* symbol estimates, in terms of the *a posteriori* probability.

ML symbol-by-symbol detection (ML–SbSD) was described in [39]. It neglects *a priori* information of the symbol hypothesis $d^{(\ell)}(k)$. Therefore it can not be employed for iterative decoding, while its

complexity approaches MAP detection. Hence, MLSD and MAP–SbSD are most often used in practice.[6] Still, the computational complexity of both MLSD and MAP–SbSD are often prohibitive, so sub-optimum derivatives which trade complexity with performance are frequently used. The detection of signal sequences with an unknown CIR is discussed further in Chapter 5.

Log-domain: If the likelihood function $p(\mathbf{y}|\mathbf{d}^{(\ell)})$ is Gaussian distributed, most of the decision problems are solved in the log-domain. That is taking the logarithm and inverting the sign of the likelihood function or *a posteriori* probability, dependent whether ML or MAP detection is used. The decision variable in the log-domain is given by

$$\widehat{\Lambda} = \min_{\ell} p(\mathbf{y}|\mathbf{d}^{(\ell)}) = \min_{\ell} \{ -\ln p(\mathbf{y}|\mathbf{d}^{(\ell)}) \}$$

which is now to be minimised.

2.2.2 One-shot detection

Suppose the received signal is deterministic (and the data sequence is uncorrelated), e.g. transmission over an AWGN channel [7], or a channel with prior knowledge of the CIR. Then the decision on the k^{th} sample can be made on a symbol-by-symbol basis, based on information about sample k only. Thus the name one-shot detector.

Consider the received signal $y(k) = d(k)\,h(k) + n(k)$, where $h(k)$ is an arbitrary but known impulse response and $n(k)$ is AWGN. In this case the likelihood function $p(\mathbf{y}|\mathbf{h}, \mathbf{d})$ conditioned on the transmitted signal $\mathbf{y}$ and the impulse response $\mathbf{h}$, can be used for detection. In the Gaussian case the pdf of the likelihood function is $\mathcal{N}(\mathbf{hd}, \sigma^2\mathbf{I})$. Now the detection of $d(k)$ is statistically independent from the rest of the sequence, giving the pdf

$$p\big(y(k)\big|h(k), d^{(\ell)}(k)\big) = \frac{1}{2\pi\sigma} \exp\left[\frac{-\big|y(k) - d^{(\ell)}(k)\,h(k)\big|^2}{\sigma^2}\right]$$

where $\sigma^2 = N_0$ denotes the variance of $n(k)$. After taking the logarithm, inverting the sign and neglecting terms which are independent of $d(k)$, the decision variable $\Lambda(\ell, k) = \big|y(k) - d^{(\ell)}(k)\,h(k)\big|^2$ is obtained. The decision upon symbol k is made by minimising $\Lambda(\ell, k)$ over the symbol alphabet $\mathcal{D}$. For constant amplitude or phase modulation, the Euclidean distance can be replaced by an inner product to yield the decision

$$\widehat{d}(k) = \min_{\ell \in \mathcal{D}} \Lambda(\ell, k) \propto \max_{\ell \in \mathcal{D}} \mathrm{Re}\big[y^*(k)\,d^{(\ell)}(k)\,h(k)\big]$$

In many practical systems prior knowledge about $h(k)$ is not available. Then a reliable estimate $\widehat{h}(k)$ may be used instead, usually aided by some side information. These sub-optimum receivers are addressed in

[6] Since in most cases MAP and ML sequence detection are equivalent, the term MLSD will be used for sequence detection in the following.

Chapter 4.

Chapter 3
System & Channel Model

This chapter discusses the equivalent base-band model, which provides the framework for subsequent chapters. This chapter is divided into three main parts: first, mobile communication based on spread spectrum techniques will be considered concisely in section 3.1; second, the channel model used to simulate a time-variant multipath channel is addressed in section 3.2; finally, the receiver front-end is specified in section 3.3.

3.1 System model

A rather simple system model is adopted. The discussion is limited to the inner receiver, so source and channel coding are not considered. Digital signalling of linear phase modulated sequences with symbol rate $f_s = 1/T_s$ is considered, which is the reciprocal symbol duration T_s. Specifically, M-ary phase shift keying (MPSK) has been adopted. The transmitted signal of the k^{th} signalling interval can be expressed in the form

$$d(k) = \exp\left(j2\pi\, a(k)/A_m\right) \quad \in \mathbb{C} \tag{3.1}$$

where $a(k) \in \{0, 1, \cdots, A_m - 1\}$ is the k^{th} information symbol containing $\log_2 A_m$ bits.

3.1.1 CDMA — spread spectrum techniques

This work considers direct sequence (DS) spread spectrum employed in a CDMA cellular radio system. The work presented in this thesis focuses on the base station to mobile link, termed the *downlink* or *forward link*. On the downlink all users transmit at the same time, giving a synchronous system. Furthermore, all signals are transmitted through the same channel. A simplified block diagram of the downlink of a DS–CDMA system is depicted in Figure 3.1. Spread spectrum signals are distinguished by the characteristic that their chirp rate per symbol $1/T_c$ is much larger than the information rate $1/T_s$ in symbols/s. In terms of the chip and symbol duration T_c and T_s respectively, the processing gain $N_c = T_s/T_c$ for a spread spectrum signal is much greater than unity. A spreading sequence converts, $d_u(k)$, the k^{th} information symbol for user $u \in \{1, \cdots, U\}$, to a wide-band noise-like signal $\{s_u(t) = d_u(k)\, c_u(t - kT_s)\}$ before transmission. The spreading sequence $c_u(t) = \sum_n c_u[n]\, g(t - nT_c)$ identifies user u, where $T_c = 1/W$ is the duration of one chip and $g(t)$ is the pulse shaping filter. The code for user u is specified by a code word of length N_c, having the coefficients $\{c_u[1], \cdots, c_u[N_c]\}$. As seen in Figure 3.1 the

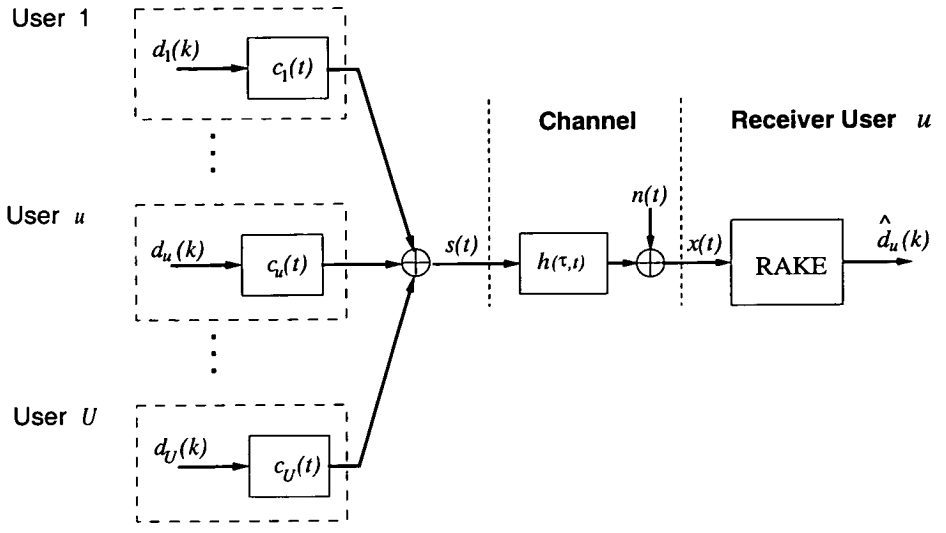


Figure 3.1: *Block diagram of the system model.*

signal of altogether U users are then transmitted simultaneously giving the base-band signal

$$s(t) = \sum_{u=1}^{U} \sum_{n=1}^{N_c} d_u(k)\, c_u[n]\, g(t - nT_c - kT_s) \tag{3.2}$$

where N_c denotes the length of the code. After transmission, the receiver can recover the desired signal by applying the same code to the received signal of the equivalent low-pass channel $x(t)$. This yields the received post-correlation signal, which is subsequently further processed to obtain the data estimate $\widehat{d}_u(k)$.

Spreading codes: There are a number of families of codes which can be used to identify each transmission. The properties of the spreading codes are discussed fully in [40,41]. Although, there exist codes which are orthogonal, there are none which retain this property in presence of a multipath channel. The time dispersive nature of the channel destroys the orthogonality of the codes. Thus the unsynchronised properties of the codes are just as important as their auto and cross-correlation properties when there is no time delay. We distinguish between two types of spreading codes:

Pseudo-random sequences such as m–sequences [42], generated by linear feedback shift registers. Unfortunately, there are only relatively few m–sequences for a certain length which means that they are of only limited use for DS–CDMA systems. To gain increased capacity (at the expense of altering the correlation properties slightly), a pair of m-sequences may be used to construct a set of Gold sequences [43], which have the property that the crosscorrelation is always 1 when the phase offset is zero. Nonzero phase offsets produce a correlation value from one of three possible values. The choice of preferred pairs of m-sequences is described in [41]. Since their synchronised characteristics are good, while their unsynchronised characteristics are not excessive, Gold sequences offer a reasonable choice of spreading sequences for DS-CDMA systems. Since pseudo-random sequences aim to approximate the properties of white Gaussian noise, they are also referred to as

pseudo noise (PN) sequences.

Orthogonal sequences such as Walsh codes [44]. They offer maximum capacity if synchronised, since they are mutually orthogonal. However, the auto- or cross-correlation of Walsh sequences can take on very high amplitudes when correlating them with time delayed versions of the same or different codewords. Thus, Walsh codes are not suitable for transmission through a multipath fading channel, since orthogonality of the codes cannot be maintained.

The non-orthogonality of the spreading codes ultimately limits the channel capacity. The characteristics of spreading sequences can be improved by using a hybrid of PN and orthogonal sequences. For instance, Walsh codes may be scrambled by very long PN or random codes, which improves their cross-correlation properties. For instance, for UMTS the spreading codes on the downlink consist of an orthogonal code scrambled with a very long Gold code of length $2^{18} - 1$ [5]. The orthogonal codes have a variable spreading factor between 4 up to 256, to provide services with different data rates.

3.2 Channel model

This section will seek to define the characteristics of a typical mobile radio channel observed in urban areas. Specifically, environments containing a number of obstacles, such as buildings, walls and traffic, will be considered. For this case modelling the radio channel by a single *line of sight* (LOS) is inadequate, since there are usually many buildings between the transmitter and receiver. The radiated electro-magnetic (EM) waves are reflected randomly by a large number of obstacles until they finally reach the receiver. Many of these local propagation paths contribute to the radiation measured at the receiver's antenna. Moreover, if a EM wave is reflected on a rough surface, the reflected EM wave will be a superposition of many rays. The reflection and defraction on rough surfaces is termed *scattering*. Of all mechanisms by which the signal may propagate through an environment this occurrence is the most difficult to predict analytically [45]. The propagation channel is illustrated in Figure 3.2.

In the following the properties of a multipath fading channel will be discussed. The channel simulator will be described which is used for simulation work in subsequent chapters.

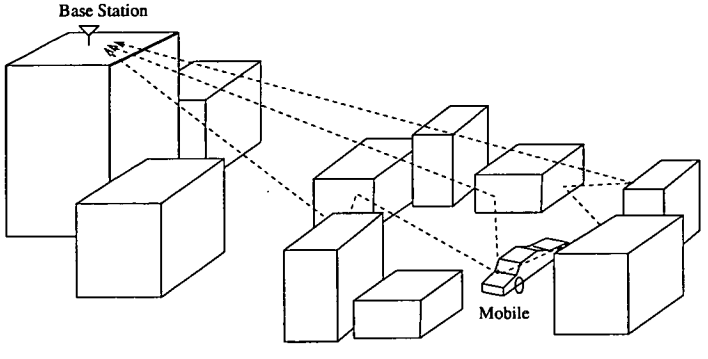


Figure 3.2: *The urban radio channel.*

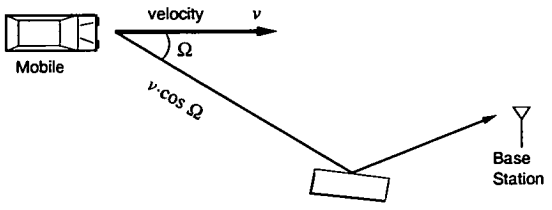


Figure 3.3: *The Doppler effect.*

3.2.1 The Doppler effect

Suppose the mobile is in motion with respect to the base station, the nature of the propagation channel varies with time due to the Doppler effect. Suppose a sine wave with frequency f_c is transmitted from the mobile. The mobile is travelling with velocity v, with angle Ω between the velocity vector and the vector to the first reflector, as illustrated in Figure 3.3. Then the receiver will observe a change in frequency ν called the *Doppler frequency*, given by

$$\nu = \frac{f_c}{c_0} v \cos\Omega \tag{3.3}$$

where c_0 denotes the speed of light. Denoting the maximum Doppler frequency by $\nu_{max} = f_c\, v/c_0$, changes of ν are in the range $[-\nu_{max}, \nu_{max}]$, dependent on Ω.

3.2.2 Describing a multipath fading channel

When transmitting a signal $s(t)$ over a multipath channel, several time delayed copies of the signal will be observed, due to a large number of propagation paths. Hence, one characteristic of the multipath medium is the spread in time. A second characteristic is due to the time variations of the channel caused by the Doppler effect. The propagation paths contributing to the received signal, appear random to the user of the channel. Therefore, it is reasonable to characterise the time-variant multipath channel statistically. The received signal is assumed to be *wide-sense stationary* (WSS). It is assumed that different propagation paths are mutually uncorrelated, which appears reasonable, since these paths are due to reflections located at different places. This effectively means that propagation paths with different time delays $\tau_n(t)$ are uncorrelated. This is called the WSS uncorrelated scattering (WSS–US) assumption [46].

A large number of propagation paths are assumed to contribute to the received signal: associated with each path is a propagation delay $\tau_n(t)$ and an attenuation factor α_n. The overall received base-band signal is the summation over all propagation paths, given by

$$x(t) = \sum_n \alpha_n\, e^{-j2\pi f_c \tau_n(t)}\, s(t - \tau_n(t)) + n(t) \tag{3.4}$$

where $n(t)$ is a complex valued, additive white Gaussian noise (AWGN) of power $\sigma^2 = N_0$. The amplitudes α_n may be modified to include a log-normal term to simulate shadowing effects, due to the movement of the mobile of longer distances [47,48]. This long-term fading characteristic is due to propagation paths which might be suddenly blocked by an obstacle, while others may appear. However,

for time intervals of interest, i.e. only a few data samples T_s, the impact of shadowing is negligible and will not be considered.

According to (3.3), the phase changes of path n are in the form

$$\varphi_n(t) \triangleq 2\pi f_c \tau_n(t) = 2\pi \nu_n t + \phi_n \tag{3.5}$$

where ν_n is the Doppler frequency of path n. The random phases ϕ_n are described by uniformly distributed random variables defined over $[0, 2\pi]$. According to (3.3), the Doppler effect causes a spectral broadening to the received signal within the range $[-\nu_{max}, \nu_{max}]$.

The phase of the n^{th} propagation path $\varphi_n(t)$ will change by 360^o whenever t changes by $1/\nu_n$, which may be only a fraction of a 1 ms for a high carrier frequency f_c and mobile velocity v. For instance, for a carrier frequency of $f_c = 2\,\text{GHz}$ and a mobile velocity $v = 120\,\text{km/h}$ the maximum Doppler frequency becomes $\nu_{max} = 222\,\text{Hz}$, according to (3.3). Thus, the time duration of a phase change of 360^o is at least $t = 4.5\,\text{ms}$.

These phase changes will result in signal fading. That is, the randomly time-variant phases $\varphi_n(t)$ associated with the term $\alpha_n e^{-j\varphi_n(t)}$ may result in adding up destructively. Then, the resultant received signal $x(t)$ is practically zero. At other times, the terms $\alpha_n e^{-j\varphi_n(t)}$ add constructively, so that the received signal amplitude is large.

The time variant channel impulse response (CIR) is obtained from (3.4) as follows

$$h(\tau, t) = \sum_n \alpha_n e^{-j\varphi_n(t)}\, \delta(\tau - \tau_n(t)) \tag{3.6}$$

The Fourier transform of $h(\tau, t)$ can be taken with respect to the time delay τ and the observation time t, yielding the transfer functions $H_f(f, t)$ and $H_D(\tau, \nu)$, respectively. This dependence of $h(\tau, t)$ on two variables implies that the frequency components of the transmitted signal in general suffer from two independent types of fading: first different frequency components will generally not observe the same fading, termed frequency-selective fading. Second the fading is time variant with respect to the observation time t. In the Doppler frequency domain the time variant channel is dependent on the Doppler frequency ν. The time variations caused by the Doppler effect, are characterised by a spectral broadening, termed the *Doppler spread*. Since the Doppler spread is within the range $[-\nu_{max}, \nu_{max}]$, it is a strictly band limited process. The difference in observation time Δt where $H_f(f, t)$, or correspondingly $h(\tau, t)$, does not change significantly is denoted as the *coherence time* Δt_c, given by the reciprocal of the Doppler spread. Clearly, a slowly fading channel has a large coherence time.

The range of frequency translations where the autocorrelation of $H_f(f, t)$ is high in the f variable is defined as the coherence bandwidth Δf_c. Thus, two sinusoids with frequency separation smaller than Δf_c will be affected by the channel in the same way. So, if signal bandwidth W of the transmitted signal is much smaller than the coherence bandwidth of the channel Δf_c the channel is said to be *frequency non-selective* or *flat fading* [7]. This implies that the time-variant transfer function $H_f(f, t)$ is a complex valued constant in the frequency variable, i.e. $H_f(f, t) = H_f(0, t)$. On the other hand, if the

bandwidth W is larger than Δf_c the channel is *frequency selective*. An analogous characterisation of the channel is in the time domain. The *multipath delay spread* τ_d defines the range of time delays τ were the auto-correlation function of the channel impulse response is essentially non-zero, and it is given by the reciprocal of the coherence bandwidth Δf_c.

Mathematically, a fading multipath channel can be described by its auto-correlation function (ACF). Assuming a WSS–US channel the ACF is in the general form [46]

$$\mathrm{E}[\, h(\tau_1, t)\, h^*(\tau_2, t+\Delta t)\,] \triangleq \phi_{hh}(\tau_1, \tau_2; \Delta t) = \phi_{hh}(\tau_1, \Delta t)\, \delta(\tau_1 - \tau_2) \tag{3.7}$$

Note that the ACF can be written in a product form:

$$\phi_{hh}(\tau_1, \Delta t) = \phi_{hh}(\Delta t)\, R(\tau_1) \tag{3.8}$$

where $\phi_{hh}(\Delta t)$ is mathematical equivalent to $\phi_{hh}(0, \Delta t)$. Thus the ACF consists of a term describing the time variations of the channel, due to the Doppler spread, $\phi_{hh}(\Delta t)$; and the *power profile* of the channel $R(\tau_1)$.

3.2.3 Tapped delay line model of the channel

Using PN sequences to modulate the information bearing signal means that the assumption of a narrow-band signal no longer holds. In this case, the bandwidth of the transmitted signals $s(t)$ is significantly larger than the coherence bandwidth of the channel

$$W \gg \frac{1}{\tau_d} \approx \Delta f_c$$

As the bandwidth of the signal $s(t)$ increases, the behaviour of the frequency components tend to become uncorrelated, because the electrical length of the propagation paths is different. This results in a dispersion of time delays of the propagation paths contributing to the received signal, which is larger than the delay spread τ_d of the channel. In other words, if two sinusoids, separated by a finite frequency range propagate in a medium, they will not be distorted in the same way.

However, this does not affect two frequency components with only a small difference in frequency [15]. With the assumption that the duration of one chip-length $T_c = 1/W$ is much smaller than the delay spread τ_d, $s(t)$ can be approximated as a constant within T_c, such that

$$s(t - \tau_n(t)) \approx s(t - q/W)\,; \qquad \textit{for} \quad \frac{q}{W} \leq \tau_n(t) < \frac{q+1}{W}\,.$$

Thus, the set of all τ_n's in (3.4) can be partitioned into Q disjoint sets $\mathcal{Q}_q$, $q = \{1, \ldots, Q\}$. Since the total multipath spread is τ_d, for all practical purposes the number of sets can be truncated at $Q = [\tau_d W] + 1$

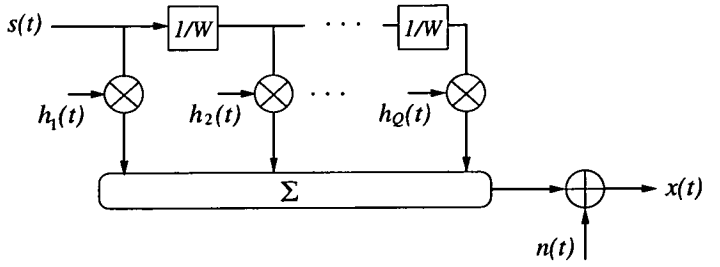


Figure 3.4: *Tapped delay line model of a frequency selective fading channel.*

taps. The standard frequency-selective multipath description of the fading channel obtained is

$$\begin{aligned} x(t) &= \sum_{q=1}^{Q} s(t-\tau_q) \sum_{n=1}^{N_q} \alpha_n \, e^{-j(2\pi\nu_n t+\phi_n)} + n(t) \; ; \qquad \tau_q = \frac{q-1}{W} \\ &= \sum_{q=1}^{Q} s(t-\tau_q) \; h_q(t) + n(t) \end{aligned} \tag{3.9}$$

where $h_q(t)$ is defined as the CIR of subset $\mathcal{Q}_q$. A frequency selective channel can be modelled by Q independent and time delayed flat fading channels. Thus, the channel may be modelled with a finite impulse response (FIR) filter having Q fingers or *taps*. This is called the tapped delay line (TDL) model of the channel and it is illustrated in Figure 3.4. In terms of the $h_q(t)$, the overall CIR of the channel becomes

$$h(\tau,t) = \sum_{q=1}^{Q} h_q(t) \; \delta(\tau-\tau_q)$$

It is assumed that the CIR consists of a large number of propagation paths, so that the central limit theorem can be applied. That is, the number of paths approach infinity $N_q \rightarrow \infty$, yielding a complex valued random variable $h_q(t) = \alpha_q(t) \, e^{j\psi(t)}$.

Where a line-of-sight (LOS) path exists between the transmitter and receiver, the CIR $h_q(t)$ follows a *Rician distribution* [14]. Mathematically, the contribution of the LOS path corresponds to the mean of the CIR, $\mathrm{E}[\, h_q(t) \,]$. On the other hand, if there is no LOS path the CIR is zero mean and $h_q(t)$ is *Rayleigh distributed*, which is the worst case situation. Moreover, since $\mathrm{E}[\, h_q(t) \,] = 0$, Rayleigh fading is easier to describe mathematically. Thus, only Rayleigh fading will be considered in this thesis. The Rayleigh distribution is the most common statistical characterisation of the fast-fading envelope of the CIR, $\alpha_q(t) = |h_q(t)|$, and is given by

$$p(\alpha_q(t)) = \frac{1}{2\bar{\alpha}_q^2} e^{-\alpha_q^2(t)/2\bar{\alpha}_q^2} \tag{3.10}$$

where $\bar{\alpha}_q^2 = \mathrm{E}[\, \alpha_q^2(t) \,]$. The phases $\psi_q(t) = \arg[h_q(t)]$ are uniformly distributed over $[0, 2\pi]$. Assuming independent Rayleigh fading of the diversity branches, the composite path weight $\alpha(t) = \sum_q \alpha_q(t)$ follows a χ^2–distribution with $2Q$ orders of freedom [7]. Another fading distribution, termed Nakagami–

m distribution [49, 50] fit some experimental results more closely.

3.2.4 Power profile of the channel

In general, as the time delay increases the measured mean power of the received signal will decline. The mean power of a propagation paths measured at the receiver, arriving with relative time delay τ with respect to the first path, is given by the power profile of the channel. The power profile determines the attenuation of tap q, which is $R(\tau_q)$ according to its definition in (3.8), for a time delay of $\tau_q = (q-1)/W$. Physically speaking, the power profile represents the mean attenuation factor of a group of propagation paths impinging within the time range $[q/W, (q+1)/W]$. That in turn allows the determination of the tap weights, i.e. the mean loss of power a received signal suffers at the q^{th} tap. It also allows the determination of the number of significant taps Q, which is the time range over which $R(\tau_q)$ is effectively non-zero. Note that $h_q(t)\, h_q^*(t) = \alpha_q^2(t)$, thus the power profile simplifies to $R(\tau_q) = \bar{\alpha}_q^2$.

The channel model proposed for the UMTS terrestrial radio access (UTRA) [51] describes a number of different scenarios. For the vehicular model, these are "Vehicular A" and "Vehicular B". For Vehicular A the delay spread is $\tau_d \approx 2.7\,\mu$s, the number of significant taps are $Q_0 = 8$. However, the number of channel taps can be lowered to 4–6 without significantly affecting the performance [51]. For Vehicular B the delay spread is significantly larger, $\tau_d \approx 20\,\mu$s, whereas the number of significant taps are $Q_0 = 6$, with the number of significant taps being much smaller than $Q = \tau_d W$.

3.2.5 Fading distributions

In this section modelling a fading channel will be considered, in order to describe the fading effects mathematically, in terms of the observation time t, or equivalently, the Doppler frequency ν. The Fourier transform of $\phi_{hh}(\tau_q, \Delta t)$ yields the *Doppler power spectrum* of the channel $S_D(\tau_q, \nu)$, which may be expressed in the product form $S_D(\tau_q, \nu) = S_{D,q}(\nu)\, R(\tau_q)$. The Doppler power spectrum of the q^{th} tap $S_{D,q}(\nu)$ represents the Fourier transform of the ACF $\phi_{hh}(\Delta t)$ in (3.8).

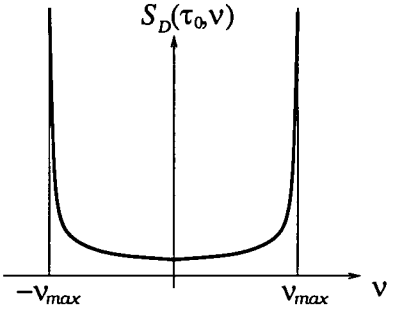


Figure 3.5: *Power spectrum of the classical Doppler model.*

Classical Doppler model: In the UTRA channel model [51] the model due to Clarke [52] has been chosen for all channel taps, termed the *classical Doppler model*. For a carrier frequency of $f_c = 2\,$GHz and a mobile speed of $v = 120\,$km/h, resulting in a maximum Doppler frequency of $\nu_{max} = 222\,$Hz.

Other studies of the profile, such as the COST–207 report [53], show two independent modes of scattering: near-in and far-out scattering. Near-in scattering refers to propagation paths due to buildings in the near vicinity of the mobile. The model assumes a carrier frequency of $f_c = 900$ MHz. For this frequency, near-in scattering is adopted for time delays $\tau < 500$ ns, or correspondingly the first two taps at the given frequency.

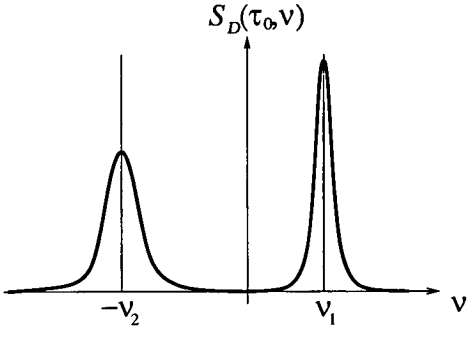
Mathematically, the power spectrum has the following form [47,52]

$$S_D(\tau_q, \nu) = \frac{\bar{\alpha}_q^2}{1 - \left(\frac{\nu}{\nu_{max}}\right)^2}; \qquad |\nu| \leq \nu_{max} \tag{3.11}$$

The Doppler power spectrum of the classical Doppler model is depicted in Figure 3.5 This equation gives rise to singularities at $\nu = \pm\nu_{max}$; however, the general form of the spectrum fits closely to observed power spectra. The Fourier transform yields the auto-correlation function of the impulse response at delay τ_q

$$\phi_{hh}(\tau_q, \Delta t) = \bar{\alpha}_q^2 \, J_0(2\pi \, \nu_{max} \Delta t) \tag{3.12}$$

This model is by far the most commonly used and will be employed to model the statistics of the CIR in the remainder of this thesis.



	GAUSS1	GAUSS2
$\bar{\nu}_1$	$-0.8\nu_{max}$	$0.7\nu_{max}$
σ_1	$0.05\nu_{max}$	$0.1\nu_{max}$
$\bar{\nu}_2$	$0.4\nu_{max}$	$-0.4\nu_{max}$
σ_2	$0.1\nu_{max}$	$0.15\nu_{max}$
A_2 [dB]	-10 dB	-15 dB

Figure 3.6: *Doppler power spectrum of the Gaussian model.*

Gaussian fading model: A model for the far-out scattering is described in the COST-207 model [53]. The far-out scattering model is based on RF measurements on physical mobile channels. Measurements suggest that as time delay τ increases a transition in the Doppler profile occurs. That is, most of the propagation paths are caused by isolated reflections such as large buildings or hills. In the COST-207 model peaks in $S_D(\tau_q, \nu)$ are approximated as a Gaussian distribution $\mathcal{N}(\bar{\nu}, \sigma^2) = \exp[-(\nu - \bar{\nu})^2/2\sigma^2]$, where $\bar{\nu}$ denotes the mean Doppler frequency and σ^2 is the standard deviation, which specifies the sharpness of the peak. Fourier transform with respect to ν yields the ACF

$$\phi(\Delta t) = \exp\left[j\bar{\nu}\,\Delta t - \sigma^2\,\Delta t/2\right] \tag{3.13}$$

The COST-207 models describe two types of Gaussian model, each of which is the summation of two Gaussian profiles $\mathcal{N}(\bar{\nu}, \sigma)$, dependent on the time delay τ_q (see Figure 3.6), defined by

$$S_D(\tau_q, \nu) = \bar{\alpha}_q^2 \left[\mathcal{N}(\bar{\nu}_1, \sigma_1) + A_2\, \mathcal{N}(\bar{\nu}_2, \sigma_2)\right] \tag{3.14}$$

The type 1 model, termed GAUSS1, is deemed appropriate for time delays of $\tau_q = [500\,\text{ns}, 2\,\mu\text{s}]$, while the type 2 model (GAUSS2) is used for excess time delays greater than $2\,\mu$s. The parameters for the models GAUSS1 and GAUSS2 are given in the table of the right hand side of Figure 3.6.

Note, this type of filtering is much more difficult to simulate, as its Doppler power spectrum is not symmetrical about the y–axis, giving rise to a complex ACF in (3.13).

Rectangular spectrum: The simplest approximation of the Doppler power spectra is to assume a uniform distribution within the interval $\nu = [-\nu_{max}, \nu_{max}]$, defined by

$$S_D(\tau_q, \nu) = \begin{cases} \frac{\bar{\alpha}_q^2}{2\nu_{max}}\,; & |\nu| \leq \nu_{max} \\ 0\,; & \text{elsewhere} \end{cases} \tag{3.15}$$

The ACF becomes $\phi_{hh}(\tau_q, \Delta t) = \bar{\alpha}_q^2 \sin(2\pi\,\nu_{max}\Delta t)/(2\pi\,\nu_{max}\Delta t)$. The rectangular Doppler power spectrum is commonly used to model the indoor multipath fading channel.

In general however, the distribution of the Doppler frequency is of less importance to assess the system performance. What matters is the maximum Doppler frequency ν_{max}; it specifies the variations in time of the received signal and hence indicates how fast the fading is.

3.2.6 Simulating the channel response

For computer simulations carried out in this thesis, the fading distributions discussed previously were approximated by filtering a noise source according to [54, 55]. Let $G_q(p)$ denote the transfer function of $h_q(t)$ (the Laplace transform with respect to the observation time t), then the Doppler power spectrum can be expressed as [25]

$$S_D(\tau_q, \nu) = G_q(p)\, G_q^*(-p) \quad \Big|_{p=j\,2\pi\nu}$$

Following [54,55], the transform function is chosen such that $S_D(\tau_q, \nu)$ is closely approximated, using a 4^{th} order IIR filter. As the fading distributions are a strictly band limited processes, $G_q(p)$ is clearly a low-pass filter. In particular, the filter employed to adopt the classical Doppler model is constructed as a cascade of two second order Butterworth low-pass filters. One is a standard filter, while the other is a modified version that rings at the cut-off frequency, to accommodate the singularities at $\pm\nu_{max}$.

Modelling discrete-time waveforms In order to implement digital filters for the use in computer simulations, the continuous-time waveform $h_q(t)$ is modelled as a discrete-time waveform $h_q(k) = h_q(t = kT_s)$. The term $1/T_s$ denotes the sampling frequency.

Since $G_q(p)$ is the transfer function in the Laplace domain, it needs to be adopted for the implementation as a digital filter. The *bilinear z–transform* [25] was used to transform $G_q(p)$ into the z–domain, yielding $\tilde{G}_q(z)$. The filter coefficients $\{g_{n,i}\}$ and $\{g_{d,i}\}$, representing the nominator and denominator of $\tilde{G}_q(z)$, respectively, yield the 4^{th} order IIR filter which generates the CIR

$$h_q(k) = \sum_{i=1}^{4} g_{d,i}\, h_q(k-i) + \sum_{i=0}^{4} g_{n,i}\, n_p(k-i)$$

where $n_p(k)$ is a complex valued white Gaussian noise process. The above equation implies that the CIR is modelled by a 4^{th} order auto-regressive moving average process (ARMA).

It is useful to define the normalised Doppler frequency in terms of the sampling frequency $1/T_s$ as $\nu' = \nu T_s$. This notation will be used throughout this thesis. For discrete-time waveforms the ACF of the channel from (3.8) can be rewritten as

$$\mathrm{E}\big[\, h_q(k)\, h_q^*(k+\Delta k)\,\big] \;\triangleq\; \phi_{hh,q}(\Delta k) \;=\; \bar{\alpha}_q^2\, \phi_{hh}(\Delta k) \tag{3.16}$$

where $\Delta k = \Delta t/T_s$. For instance, assuming a classical Doppler model from (3.12), the ACF becomes $\bar{\alpha}_q^2\, J_0(2\pi\nu'_{max}\,\Delta k)$. This notation for the ACF will be extensively used in this thesis to generate the covariance matrix $\mathbf{\Phi_{hh}}$ defined in (2.11). For a flat fading channel $\bar{\alpha}_1^2$ can be normalised to one, hence $\phi_{hh}(\Delta k)$ suffices to describe the channel statistics. For the frequency selective case the tap weight of the qth tap relative to the first one was set to $\bar{\alpha}_q^2 = \bar{\alpha}_1^2\, 2^{1-q}$, i.e. the average power of this tap is 3 dB less than tap $q-1$. To allow fair comparison between channels with different numbers of diversity taps, the mean $\bar{\alpha}_q^2$ is chosen such that the the sum of all tap weights added together always equals the signal energy E_s of duration T_s:

$$E_s \;\triangleq\; \sum_{q=1}^{Q} \bar{\alpha}_q^2 \;=\; \bar{\alpha}_1^2 \sum_{q=1}^{Q} 2^{1-q} \tag{3.17}$$

3.3 RAKE Receiver front-end

Multipath reception over a frequency selective fading channel is one form of diversity reception, in which information flows from transmitter to receiver via *natural* diversity. Thus instead of regarding the multipath receiver as a nuisance disturbance whose effects are to be suppressed, it should be regarded as an opportunity to improve system performance. A RAKE receiver is an application of a diversity receiver used in spread spectrum systems. RAKE reception has received much attention since its introduction in the late 1950s by Price and Green [56]. Theoretical and practical aspects have been extensively studied since, the results are summarised in e.g. [7, 15]. A RAKE filter can be split into two main parts: first a tapped delay line (TDL) to despread the received signal; and second a weighted combination of the diversity branches. If the number of multipath component is higher than the receiver can process, the TDL approach becomes too complex. This problem may be handled by some sort of selection diversity, where RAKE fingers are placed on a subset of decision delays. In order to do that, the delay of the main

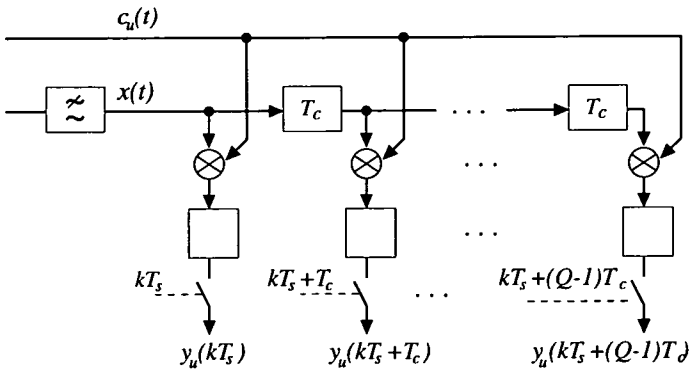


Figure 3.7: *Block diagram of the receiver front-end.*

multipath components needs to be estimated.

In this section the receiver front-end including the despreading unit using a TDL is described. The RAKE combining of the diversity taps will be discussed in subsequent chapters.

After low-pass filtering the received signal $x(t)$ at the input of the despreader consists of several time delays copies of the transmitted signal $s(t)$ which fade independently, due to the frequency selective channel. Taking the spread transmitted signal $s(t) = \sum_u d_u(k)\, c_u(t-kT_s)$ and put it into (3.9), which represents the tapped delay line model of the channel, yields the pre-correlation received signal

$$x(t) = \sum_{p=1}^{Q} \sum_{v=1}^{U} h_p(t)\, d_v(k)\, c_v(t-kT_s-\tau_p) + n(t)\ ; \qquad \tau_p = (p-1)T_c \tag{3.18}$$

The signal of the desired user u can be recovered by applying the spreading sequence $c_u(t)$ to time delayed copies of the received signal. This may be achieved by feeding $x(t)$ through a TDL [56] shown in Figure 3.7. The signal after despreading and sampling is for RAKE finger $q = \{1, \cdots, Q\}$

$$y_{qu}(k) \triangleq y_u(kT_s-\tau_q) \quad = \quad \int_{T_s(k-1)}^{kT_s} x(t)\, c_u(t-kT_s-\tau_q)\, dt\ ; \qquad \tau_q = (q-1)T_c$$

The code sequences are constructed such that they are cyclo-stationary, that is $c_u(t \pm N_c T_c) = c_u(t)$, with $N_c T_c = T_s$ being the symbol duration. For the evaluation of the integral it is assumed that $h_p(t)$ from (3.18) is approximately constant for one symbol interval T_s. This idealised assumption allows all subsequent processing to be carried out on the symbol level. Although, for very fast fading this assumption does not hold any longer, thus $y_{qu}(k)$ must be processed at chip level. This however, increases the computational complexity by the processing gain N_c. Therefore, chip level processing will not be considered further on. Furthermore, inter-symbol interferences (ISI) of the data bearing symbol are neglected, which is a reasonable assumption for modest delay spreads and large processing gains. Now the

post correlation received signal can be expressed as

$$y_{qu}(k) = \sum_{p=1}^{Q}\sum_{v=1}^{U} d_v(k)\, h_p(k) \underbrace{\int_0^{T_s} c_v(t-\tau_p)\, c_u(t-\tau_q)\, dt}_{\rho_{uv}(p-q)} + n'(k) \tag{3.19}$$

where $n'(k) = \int n(t)\, c_u(t-kT_s-\tau_q)\, dt$ is a white Gaussian noise process with $\mathcal{N}[0, N_0\rho_{uu}(0)]$. The term $\rho_{uv}(p-q)$ represents the correlation coefficient between codes u and v with the relative delay $(\tau_p-\tau_q)$. Recall the definition of the spreading sequence $c_u(t) = \sum_n c_u[n]\, g(t-nT_c)$ from section 3.1.1. Given the pulse shaping filter $g(t)$ satisfies the Nyquist criterion, the code correlation can be expressed as

$$\rho_{uv}(p-q) = \sum_{n=1}^{N_c} c_v[n-p]\, c_u[n-q] \tag{3.20}$$

Note that cases where $n+p-q \leq 0$ and $n+p-q > N_c$ result in ISI. This means there are Q chips out of N_c chips per symbol involved. MAI, i.e. the interference from other users, ultimately limits the system performance, which results in an error floor. Therefore the choice of code families with low correlation coefficients $\rho_{uv}(\cdot)$ becomes crucial. The effects of MAI will be further discussed in section 6.6.

3.3.1 Single user case

Consider the case when there is no MAI present. This may be achieved by a genuine random code which has the properties of white Gaussian noise, that is

$$\rho_{uv}(i-q) = \begin{cases} 1\,, & u = v \quad \text{and} \quad i = q\,; \\ 0\,, & \text{elsewhere.} \end{cases}$$

Unfortunately, there exists no such code. Thus, this idealised assumption serves as a lower bound for systems with non-zero correlation coefficients $\rho_{uv}(\cdot)$. If all interference due to other users is cancelled out the received signal simplifies to

$$y_q(k) = d(k)\, h_q(k) + n(k) \tag{3.21}$$

where the subscript u to identify a certain user has been dropped. The single user case, which is a general order Q diversity system will be assumed in the Chapters 4–6.

For realisable codes the effects of MAI can be taken into account by assuming that the interference is a white noise process in (3.21). Then the variance of the noise term $n(k)$ is adapted to include thermal noise and the interference caused by other users. This simplified assumption is known as the *Gaussian assumption.*

To allow fair comparison between receivers with different numbers of diversity taps, the tap weights $\bar{\alpha}_q^2 = \mathrm{E}\left[\, |h_q(k)|^2 \,\right]$ are chosen such that the average signal-to-noise ratio (SNR) always is $\bar{\gamma} = E_s/N_0$.

This is achieved by defining

$$\bar{\gamma} \triangleq \frac{E_s}{N_0} = \frac{1}{N_0} \sum_{q=1}^{Q} \bar{\alpha}_q^2 \tag{3.22}$$

where the symbol energy E_s was defined in (3.17), and the noise power $\sigma^2 = N_0$ was defined in section 2.1.2.

Chapter 4
One Shot Receivers

In this chapter sub-optimum but realisable receiver structures are discussed. Here detecting the received signal is carried out independently from estimating the CIR. The correlation of adjacent samples is exploited for estimation. These estimates are used to decorrelate the received signal, such that detection can be performed on a symbol-by-symbol basis. The receiver follows the separation principle for detection and estimation discussed in section 2.2.2.

4.1 The RAKE receiver

In this section receiver structures are discussed for diversity reception of Q independent fading taps of the received signal. The single user case is considered, which according to section 3.3.1 yields the received signal of the q^{th} diversity tap $y_q(k) = d(k)\,h_q(k) + n(k)$ with $q \in \{1, \cdots, Q\}$. Assuming that all Q taps fade independently, the estimate of the k^{th} information bit $d(k)$ is given by

$$\widehat{d}(k) \;=\; \max_{\ell \in \mathcal{D}} \Lambda(\ell, k) \;=\; \max_{\ell \in \mathcal{D}} \left\{ \sum_{q=1}^{Q} \mathrm{Re}\big[y_q^*(k)\; d^{(\ell)}(k)\, \widehat{h}_q(k) \big] \right\} \tag{4.1}$$

where $\Lambda(\ell, k)$ denotes the decision variable and $\mathcal{D}$ represents the symbol alphabet. For binary modulation, the decision of (4.1) can be achieved simply by extracting the sign, that is

$$\widehat{d}(k) \;=\; \mathrm{sgn} \left\{ \sum_{q=1}^{Q} \mathrm{Re}\big[y_q(k)\, \widehat{h}_q^*(k) \big] \right\} = \{-1, 1\}$$

where sgn denotes the sign operator. For the case where the receiver has perfect measurement of the CIR, (4.1) corresponds to maximal ratio combining (MRC) [57]. The considered receiver structures are deduced from the conventional RAKE receiver [7], where $h_q(k)$ is replaced by its estimate $\widehat{h}_q(k)$, termed generalised MRC [58]. A block diagram of the RAKE receiver for binary signalling is depicted in Figure 4.1.

Due to imperfect channel estimation, the SNR of the q^{th} tap, $\gamma_q(k)$, decreases in terms of the prediction error $\epsilon_q(k) = \widehat{h}_q(k) - h_q(k)$. The average SNR of the signal of the q^{th} diversity tap, after multiplication with $\widehat{h}_q^*(k)$, but before diversity combining, has the mean

$$\bar{\gamma}_q \;=\; \frac{\mathrm{Re}\big\{ \mathrm{E}\big[\, h_q(k)\, \widehat{h}_q^*(k) \,\big] \big\}}{N_0 + \mathrm{E}\big[\, |\epsilon_q(k)|^2 \,\big]} \tag{4.2}$$

Optimising the estimate $\widehat{h}_q(k)$ in the MMSE sense calls for minimising the estimation error $\mathrm{E}\big[\, |\epsilon_q(k)|^2 \,\big] = V_{min}^2$. The tap weights $\bar{\alpha}_q^2 = \mathrm{E}\big[\, |h_q(k)|^2 \,\big]$, which accounts for the tap weights of the tapped delay line

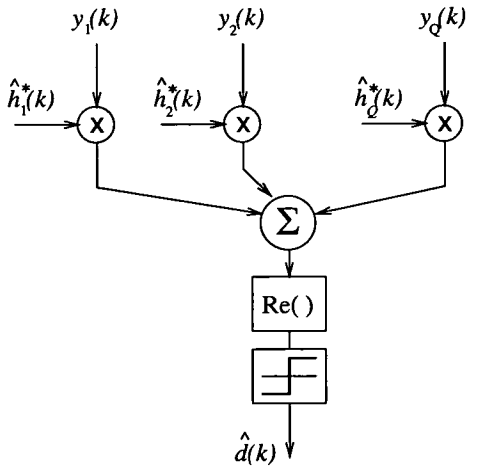


Figure 4.1: *Block diagram of a RAKE receiver with Q diversity taps for binary signalling.*

model of the channel, described in section 3.2.4, is chosen such that the average input SNR always is $\bar{\gamma} = E_s/N_0$, according to (3.22). After diversity combining the average SNR at the receiver output is in the form

$$\bar{\gamma}_{out} = \sum_{q=1}^{Q} \bar{\gamma}_q \tag{4.3}$$

Given the receiver has perfect knowledge of the CIR, i.e. $\widehat{h}_q(k) = h_q(k)$ and $\epsilon_q(k) = 0$, the output equals the input SNR, i.e. $\bar{\gamma}_{out} = E_s/N_0$. The SNR per tap from (4.2) accordingly simplifies to $\bar{\gamma}_q = \bar{\alpha}_q^2/N_0$. Otherwise, if only imperfect estimates of the CIR are available, the average output SNR $\bar{\gamma}_{out}$ is degraded by the MSE $\mathrm{E}\left[\,|\epsilon_q(k)|^2\,\right]$ of the estimation error.

With the output SNR, $\bar{\gamma}_{out}$ defined in (4.3), it is possible to evaluate the probability that $\widehat{d}(k)$ is detected as an error, following e.g. [7, chp. 14]. The probability of a decision error $P_e(\gamma_{out})$, conditioned on γ_{out}, is given by the probability of a received signal error transmitted over an AWGN channel, with the SNR γ_{out}. The SNR γ_{out} is a sum of Q squared complex valued Gaussian distributed random variables, $\gamma_q(k)$, described by a non-central χ^2–distribution with $2Q$ degrees of freedom [7], $p_{\chi^2}(\gamma_{out})$. To obtain the average probability of error, the conditional error probability, $P_e(\gamma_{out})$, is averaged with the pdf of the multipath fading channel, $p_{\chi^2}(\gamma_{out})$, over all $\gamma_{out} \geq 0$, given by

$$P_e = \int_0^\infty p_{\chi^2}(\gamma_{out})\, P_e(\gamma_{out})\, d\gamma_{out} \tag{4.4}$$

Considering BPSK, the error probability conditioned on γ_{out} is[1] $P_e(\gamma_{out}) = \frac{1}{2}\mathrm{erfc}(\sqrt{\gamma_{out}})$. After

[1]The complementary error function is defined in [7]: $\mathrm{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-u^2)\, du$.

solving the integral in (4.4), the average probability of a bit error for BPSK is obtained

$$P_e = \frac{1}{2}\sum_{q=1}^{Q}\Pi_q\left\{1-\sqrt{\frac{\bar{\gamma}_q}{1+\bar{\gamma}_q}}\right\}, \qquad \Pi_q = \prod_{\substack{i=1\\ i\neq q}}^{Q}\frac{\bar{\gamma}_q}{\bar{\gamma}_q-\bar{\gamma}_i}; \qquad \text{for BPSK} \tag{4.5}$$

The case where the CIR is known *a priori* serves as a lower bound which will be used for comparison purposes throughout this thesis. The effects of imperfect channel estimation can be taken into account by inserting $\bar{\gamma}_q$ from (4.2), into the above equation [59]. Alternatively, the bit error probability for binary modulation can be determined by using the characteristic function of (4.4). A general error formula for the pairwise error probability of a random signal disturbed by Gaussian noise was given by Barrett [60]. Barrett's formula will be applied to determine the error probability for BPSK in section 4.2.1.

Two cases are considered for DPSK: the DPSK demodulation can be performed before or after the diversity combining in the RAKE receiver. The solution of the former case is described in e.g. [7, chp. 14], which will be coined *conventional DPSK* in the following. Since demodulation is done before combining, the receiver is non-coherent. Hence, no channel estimation is required. The channel estimate is simply given by the received signal of the previous sample $\widehat{h}_q(k) = y_q(k-1)$. Thus the correlation $\mathrm{E}\left[\, h_q(k)\,\widehat{h}_q^*(k)\,\right]$ equals the correlation of the CIR $\mathrm{E}\left[\, h_q(k)\,h_q^*(k-1)\,\right] = \phi_{hh,q}(1)$ with the MSE $\mathrm{E}\left[\,|\epsilon_q(k)|^2\,\right] = 1-\left|\phi_{hh,q}(1)\right|$. Consequently, the mean SNR of tap q from (4.2) becomes

$$\bar{\gamma}_q = \frac{\mathrm{Re}\{\phi_{hh,q}(1)\}}{N_0+1-\left|\phi_{hh,q}(1)\right|}$$

With the conditional error probability for DPSK, $P_e(\gamma_{out}) = e^{\gamma_{out}}/2$, the average probability of a bit error, after averaging with $p_{\chi^2}(\gamma_{out})$ according to (4.4), becomes [7]

$$P_e = 2^{1-2Q}\sum_{m=0}^{Q-1}\sum_{n=0}^{Q-1-m}\binom{2Q-1}{n}\sum_{q=1}^{Q}\frac{\Pi_q}{\bar{\gamma}_q}\left(\frac{\bar{\gamma}_q}{1+\bar{\gamma}_q}\right)^{m+1}, \qquad \text{conventional DPSK} \tag{4.6}$$

where Π_q was defined in (4.5).

In the latter case the differential decoding is performed *after* diversity combining. In this case detection is performed in analogy to coherent PSK and the receiver output $\widehat{d}(k)$ of (4.1) is then differentially decoded. This approach will be termed *differentially encoded PSK (DEPSK)* in the following. The diversity taps are combined coherently, thus a reliable estimate of the CIR is essential. The benefit of this approach is that, given a sufficient channel estimate, the receiver performance can be improved. The performance improvements of such receiver structures will be assessed in section 4.3. An approximation of the error probability for that case was derived in [61].

Numerical results for BPSK and DEPSK with perfectly known CIR (labels "ideal BPSK" and "ideal DEPSK" respectively), and for a conventional DPSK receiver (label "DPSK"), are shown in Figure 4.2. The statistics of the channel are specified by the classical Doppler power spectra, having the ACF $\phi_{hh,q}(1) = J_0(2\pi\nu'_{max})$, due to Clarke [52] (see also section 3.2.6). Figure 4.2 shows the conventional DPSK receiver of (4.6) for some fading rates ν'_{max}. It is seen that the penalty of differential encoding of

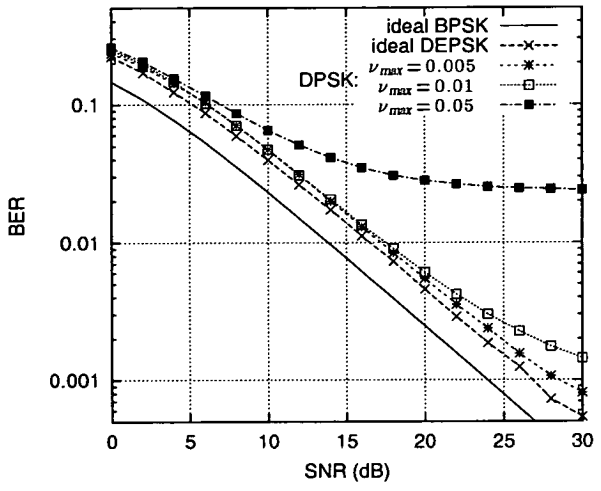


Figure 4.2: *BER vs SNR for some receiver types with different fading rates, ν'_{max}; $Q=1$.*

the data bits is roughly 2.5 dB. For slow fading, there is little difference between the coherent combining DEPSK receiver with perfect knowledge of the CIR and the conventional DPSK receiver. For fast fading however, the conventional DPSK receiver experiences an error floor, which increases with growing ν'_{max}.

It is the purpose of this chapter to lower the error floor of the conventional DPSK receiver with receiver structures which do not assume *a priori* knowledge of the channel response. The CIR can be estimated in various ways, dependent on the system model. If a pilot is provided, the detection and estimation task can be completely separated. This leads to pilot aided channel estimation discussed in the following section. For differential encoding of the transmitted signal, on the other hand, the CIR can be estimated in a decision directed manner, discussed in section 4.3.

4.2 Pilot aided channel estimation

In this section techniques where channel estimation is performed by using pilots are considered, i.e. signals without data modulation. The CIR estimate, $\widehat{h}_q(k)$, is not dependent on the decisions of past or future samples. The pilot symbols may be provided either in form of a parallel pilot channel or pilot symbols are multiplexed in the data stream, as described in section 1.2.1. If a pilot channel is employed, channel estimation is straightforward; the received pilot signal is identical to $y'_q(k)$, given the pilot and the traffic signal are transmitted through the same physical channel and therefore undergo the same fading distortion. Performance comparisons between time-multiplexed pilot channel and parallel pilot channel were carried out for DS-CDMA systems in [62]. No considerable differences were found between the two techniques. However, one feature of the 3rd generation WCDMA that is distinct from the IS-95 CDMA standard is that instead of a separate pilot channel several dB stronger than the user data channels, pilot symbols are embedded in the data stream, in both the up- and downlink.

This section concentrates on *pilot symbol-assisted channel estimation* schemes where known symbols are periodically time multiplexed in the data stream for carrier recovery. So, one in R symbols is known

to the receiver followed by $R-1$ data symbols, corresponding to the ratio $1:R$. Proposed pilot aided receivers were based on a low order Gaussian interpolation [10, 63, 64] and Wiener filtering [11]. A pilot aided receiver applicable to 3rd generation mobile standards such as UMTS/IMT-2000 was investigated in [65]. There a short sequence of pilot symbols is embedded in the data stream, instead of a single pilot. The channel estimation technique involves using pilot symbols of several slots and weighing them dependent on the distance from the time slot to be detected.

From the sampling theory point of view, the sampling rate of the received pilots is $f_P = 1/(RT_s)$. This is R times less the sampling rate of the received signal, $f_s = 1/T_s$. Provided that the CIR is a band-limited process, the sampling theorem [66] states that the CIR $h_q(k)$ can be recovered perfectly from noise free samples $h_q(\kappa R)$, given that the samples are taken at least the Nyquist rate $R\nu_{max}T_s < 0.5$. Then the fading characteristics are determined by the interpolation filter

$$h_q(k) = \sum_{\kappa=-\infty}^{\infty} h_q(\kappa R)\ \mathrm{sinc}(\pi[k/R-\kappa]) \tag{4.7}$$

where $\mathrm{sinc}(x) \triangleq \sin(x)/x$. If $\{h_q(\kappa R)\}$ is replaced by the noisy pilots $\{y_{Pq}(\kappa)\}$ some degree of oversampling must be allowed to average over the noise. Introducing an oversampling factor of $\beta \geq 1$, R can be determined by [66]

$$R \leq \frac{1}{2\beta\,\nu'_{max}} \tag{4.8}$$

where $\nu'_{max} = \nu_{max}T_s$ is the maximum normalised Doppler frequency. An oversampling factor of $\beta = 1$ means that $h_q(\kappa R)$ is sampled at Nyquist rate. Hence, there is no redundancy but the noise bandwidth is large. For instance for a maximum normalised Doppler frequency of $\nu'_{max} = 0.005$ the maximum spacing of two adjacent pilots is $R = 100$ symbol intervals. Practically however, R should be much smaller than that, to allow some degree of oversampling in order to average over the additive Gaussian noise. The optimisation of the pilot symbol symbol spacing R is analysed in [67]. As a rule of thumb $\beta = 10$ offers a good compromise between redundancy and noise reduction. For the above example an oversampling of $\beta = 10$ leads to a multiplexing rate of $R = 10$. In [64] the CIR estimate $\widehat{h}_q(k)$ is obtained by applying an interpolation filter (4.8) to the pilots $\{y_{Pq}(\kappa)\}$. This is clearly sub-optimum for filtering of noisy signals, furthermore the filter is non causal, thus inducing a long decision delay.

4.2.1 Pilot symbol based Wiener filtering

A more sophisticated approach to estimate $\widehat{h}_q(k)$ is to use a Wiener filter, which estimates $h_q(k)$ based on the pilot symbols only. Let the received signal without data modulation be defined by $y'_q(k) = h_q(k) + n(k)$. To describe pilot symbol-assisted channel estimation it is useful to define a subset of the received signal sequence containing only the pilots, $\{y_{Pq}(\kappa)\} = \{y'_q(\kappa R)\}$, sampled at a R times lower rate $\kappa = \lfloor k/R \rfloor$. After the reception of a pilot symbol, a whole block of $R-1$ data samples, $\{y_q([\kappa-\Delta\kappa]R+r)\}$ with $r = \{1, \cdots, R-1\}$, can be processed. One pilot followed by a block of $R-1$ data samples is called a *frame*. The time delay $\Delta\kappa$ specifies which frame is to be processed. Thus, the

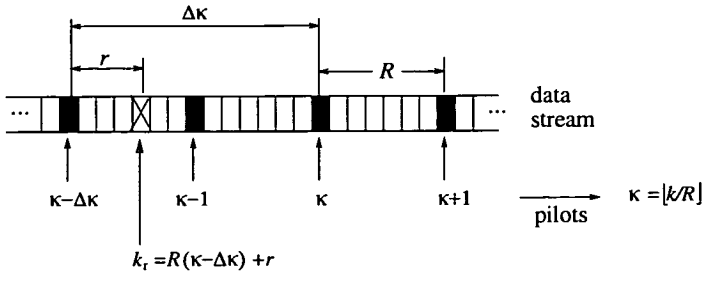


Figure 4.3: *Pilot symbol insertion (PSI) technique.*

desired bit which is to be detected, is accordingly $k_r = (\kappa - \Delta\kappa)r$. Figure 4.3 illustrates how the pilots are placed in the data stream.

Let the vector containing the recent M received pilots be denoted by $\mathbf{y}_{Pq}(\kappa) = [y'_{Pq}(\kappa - M + 1), \cdots, y'_{Pq}(\kappa)]^T$, of dimension $\mathbb{C}^M$. With this definition, the covariance matrix of the pilots $\mathbf{\Phi}_P = \mathrm{E}[\,\mathbf{y}_{Pq}(\kappa)\,\mathbf{y}^H_{Pq}(\kappa)\,]$ can be specified. The entry of the m^{th} row and n^{th} column of the covariance matrix is given by

$$\{\mathbf{\Phi}_P\}_{m,n} = \mathrm{E}[\,y_{Pq}(\kappa - n)\,y^*_{Pq}(\kappa - m)\,] = \mathrm{E}[\,y'_q(k - Rn)\,y'^*_q(k - Rm)\,]\;; \qquad k = \kappa R \tag{4.9}$$

Furthermore, define the cross correlation vector between the desired sample $y'_q(k_r)$ and the pilots $\mathbf{y}_{Pq}(\kappa)$. The m^{th} entry of the cross correlation vector $\phi^{(\Delta\kappa)}_{Pr} = \mathrm{E}[\,y'^*_q(k_r)\,\mathbf{y}_{Pq}(\kappa)\,]$ can be expressed as

$$\{\phi^{(\Delta\kappa)}_{Pr}\}_m = \mathrm{E}[\,y'^*_q(k_r)\,y_{Pq}(\kappa - m)\,] = \mathrm{E}[\,y'^*_q(k_r)\,y'_q(k - Rm)\,]\;; \qquad k_r = (\kappa - \Delta\kappa)R + r \tag{4.10}$$

The quantities (4.9) and (4.10) are necessary to evaluate the channel estimator.

A positive $\Delta\kappa$ imposes a time delay of $\Delta\kappa R$ symbols at the receiver output. Then the estimation filter is a smoothing type filter. On the other hand, setting $\Delta\kappa = 0$ specifies a linear prediction receiver without an induced time delay due to channel estimation, at the expense of a somewhat poorer estimate of the CIR. With these definitions the CIR at time k_r can be estimated as follows

$$\widehat{h}_q(k_r) = \sum_{m=0}^{M-1} w^{*(\Delta\kappa)}_{mr}\, y_{Pq}(\kappa - m) = \mathbf{w}^{H(\Delta\kappa)}_r\, \mathbf{y}_{Pq}(\kappa)\;; \qquad \begin{aligned} \kappa &\triangleq \lfloor k/R \rfloor \\ k_r &\triangleq (\kappa - \Delta\kappa)R + r \\ r &= \{1, \cdots, R-1\} \end{aligned} \tag{4.11}$$

where $w^{(\Delta\kappa)}_{mr}$ is a coefficient of the matrix $\mathbf{W}^{(\Delta\kappa)} = [\mathbf{w}^{(\Delta\kappa)}_1, \cdots, \mathbf{w}^{(\Delta\kappa)}_{R-1}]$ of dimension $\mathbb{C}^{M \times R-1}$. The filter bank $\mathbf{W}^{(\Delta\kappa)}$ is the Wiener filter matrix of the data block to be processed. The column vectors $\mathbf{w}^{(\Delta\kappa)}_r = [w^{(\Delta\kappa)}_{1r}, \cdots, w^{(\Delta\kappa)}_{Mr}]^T$ are the estimators for the r^{th} symbol in the block, corresponding to the r^{th} column of $\mathbf{W}^{(\Delta\kappa)}$. In analogy to 2.19, the MSE of the estimation errors is minimised, $\mathrm{E}[\,|\epsilon_q(k_r)|^2\,]$, as a function of the weight vector $\mathbf{w}^{(\Delta\kappa)}_r$, with

$$\epsilon_q(k_r) \triangleq y'_q(k_r) - \widehat{h}_q(k_r) = y'_q(k_r) - \mathbf{w}^{H(\Delta\kappa)}_r \mathbf{y}_{Pq}(\kappa)$$

According to (2.18) the solution in the MMSE sense is, $\mathbf{w}_r^{(\Delta\kappa)} = \mathbf{\Phi}_P^{-1} \cdot \boldsymbol{\phi}_{Pr}^{(\Delta\kappa)}$, being the well known Wiener-Hopf equation. The auto- and cross-correlation quantities are defined by (4.9) and (4.10). It is desirable to find an expression for detection of an entire frame. The received signal of one frame is then processed by the filter bank $\mathbf{W}^{(\Delta\kappa)}$. Now the estimation of a whole data block with delay $\Delta\kappa$ is given by

$$\left[\widehat{h}_q\big((\kappa-\Delta\kappa)R\big), \cdots, \widehat{h}_q\big((\kappa-\Delta\kappa+1)R\big)\right]^T = \mathbf{W}_r^{H(\Delta\kappa)}\, \mathbf{y}_{Pq}(\kappa) \tag{4.12}$$

The MMSE of the channel estimator, in analogy to (2.19) is

$$\begin{aligned} V_{min,q}^2(r, \Delta\kappa) &= \mathrm{E}\big[\,|\epsilon_q(k_r)|^2\,\big] = \mathrm{E}\Big[\,|y'_q(k_r) - \mathbf{w}_r^{H(\Delta\kappa)}\, \mathbf{y}_{Pq}(\kappa)|^2\,\Big] \\ &= 1 - \boldsymbol{\phi}_{Pr}^{H(\Delta\kappa)}\, \mathbf{\Phi}_P^{-1}\, \boldsymbol{\phi}_{Pr}^{(\Delta\kappa)} \end{aligned}$$

The smallest MSE is obtained by estimating the channel response with a delay of $\Delta\kappa = M/2$ pilots. The MSE monotonically increases towards estimating the channel response when $\Delta\kappa$ tends to zero.

Performance analysis In this section the evaluation of the error probability for pilot aided channel estimation is addressed. Its application to pilot aided channel estimation for a similar receiver was carried out by Kaasila and Mämmelä [68]. The derivation of the error probability due to Barrett [60], was originally derived for maximum likelihood sequence detection (MLSD) for binary modulated signals. However, Barrett's formula can be applied to a wide range of different receiver structures. For a more thorough discussion of Barrett's formula, see the performance analysis of MLSD in section 5.1.5.

Note, the probability of a bit error is dependent on its position relative to a pilot, since the MSE of the channel estimate, $V^2(r)$, is a function of r. For binary modulation we have $d(k) = \{-1, 1\}$. The decision variable from (4.1) can be rewritten as $\Lambda(k, r) = \sum_q \mathrm{Re}\{y_q(k)\, \widehat{h}_q^*(k)\}$, and choosing $\widehat{d}(k) = 1$ if $\Lambda(k, r)$ is positive and $\widehat{d}(k) = -1$ otherwise. Assuming that $d(k) = 1$ was transmitted, a negative $\Lambda(k, r)$ equals a decision error. Thus, the probability of error is the probability that $\Lambda(k, r) < 0$, i.e. $P_e(r) = P\big(\Lambda(k, r) < 0\big)$. In order to determine the error probability, the decision variable is cast into a quadratic form

$$\Lambda(k, r) = \sum_{q=1}^{Q} \mathrm{Re}\{y_q(k)\, \widehat{h}_q^*(k)\} = \mathbf{u}^H(k, r)\, \mathbf{Q}(r)\, \mathbf{u}(k, r) \tag{4.13}$$

The column vector $\mathbf{u}(k, r) = [\mathbf{u}_1^T(k, r), \cdots, \mathbf{u}_Q^T(k, r)]^T$ of dimension $\mathbb{C}^{Q(M+1)}$, contains the bit to be detected and the pilots used for channel estimation for all Q diversity taps. Its q^{th} entry $\mathbf{u}_q(k, r)$ of dimension $\mathbb{C}^{M+1}$ is defined by $\mathbf{u}_q(k, r) = [y_q(k - M_f R), \mathbf{y}_{Pq}(\kappa)]^T$. The filter matrix $\mathbf{Q}(r)$ as well as $\mathbf{\Phi}_u(r) = \mathrm{E}\big[\,\mathbf{u}(k, r)\, \mathbf{u}^H(k, r)\,\big]$, which is the covariance matrix of $\mathbf{u}(k, r)$, are derived in Appendix A.3.1.

Applying Barrett's formula determines the probability of a bit error as a function of r, yielding [60]

$$P_e(r) = \sum_{\substack{n=1 \\ \lambda_n(r)<0}}^{N} \prod_{\substack{\nu=1 \\ \nu \neq n}}^{N} \frac{1}{1-\lambda_\nu(r)/\lambda_n(r)} \tag{4.14}$$

where the set $\{\lambda_n(r),\ n = 1, \cdots, N\}$ are the eigenvalues of the matrix $\Phi_u(r)\mathbf{Q}(r)$. The average probability of error can be obtained by averaging $P_e(r)$ over r, that is

$$P_e = \frac{1}{R-1} \sum_{r=1}^{R-1} \sum_{\substack{n=1 \\ \lambda_n(r)<0}}^{N} \prod_{\substack{\nu=1 \\ \nu \neq n}}^{N} \frac{1}{1-\lambda_\nu(r)/\lambda_n(r)} \tag{4.15}$$

Note that (4.15) is not valid if two or more eigenvalues are equal. For instance, this case occurs if the average powers of the diversity taps are equal and they have the same statistics, and therefore filter coefficients. It is assumed that the sequences are equally likely, so the error probability is the same for all hypotheses. The dimension of the matrices is $N \times N$, with $N = Q(M+1)$. It should also be noted, that $\Phi_y(0)\mathbf{Q}$ is in general *not* a Hermitian matrix. The eigenvalue decomposition of a non-Hermitian matrix is a non trivial task. A solution to this problem is described in Appendix A.2. The eigenvalue decomposition was evaluated with a C function given in [69].

Numerical Results Some numerical results for the pilot aided receiver are presented in this section. The error probabilities are computed using (4.14) and (4.15) and are compared with Monte Carlo simulations to prove their validity. Unless otherwise stated the results presented in this section, are based on the specification in Table 4.1 operating in a complex baseband urban channel according to section 3.2.6. The performance of the system was evaluated for one and two diversity taps. The statistics of the qth diversity tap are specified by the classical Doppler power spectra from (3.12), due to Clarke [52], described by the ACF $\phi_{hh,q}(\Delta k) = J_0(2\pi\nu'_{max}\Delta k)$. Generally, curves labelled "ideal" show the results when the CIR is known *a priori* to the receiver, given by (4.5).

Filter order	M	4
PSI rate	R	10
Number of diversity taps	Q	2
Mean SNR	$\bar{\gamma}$	10 dB
normalised Doppler	ν'_{max}	0.005
Modulation		BPSK

Table 4.1: *System & simulation parameters for the pilot aided RAKE receiver.*

In Figure 4.4 the average bit error rate (BER) is presented as a function of the average signal-to-noise ratio (SNR) in a system with $Q = 1$ and $Q = 2$ diversity taps. The results which are obtained by applying (4.15), are compared with Monte Carlo simulations. It is seen that the performances improves significantly when the delay $\Delta\kappa$ is increased, since additional future samples are taken into account, which improve the channel estimate. Diversity is seen to improve the performance significantly. The difference between the lower bound where the CIR is known to the receiver and the PA–RAKE is seen

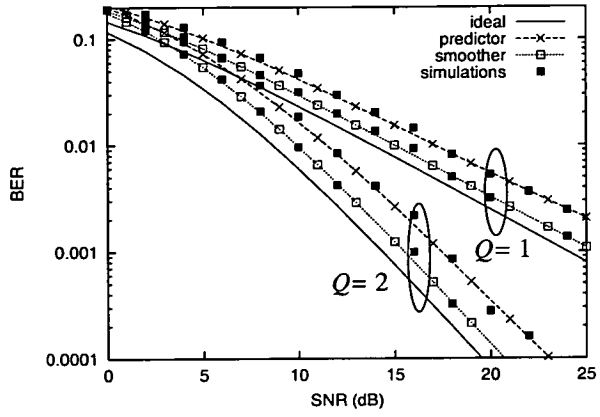


Figure 4.4: *BER vs SNR for different numbers of diversity taps Q of a smoother ($\Delta\kappa = M/2$) and linear predictor ($\Delta\kappa = 0$); $\nu'_{max} = 0.005$.*

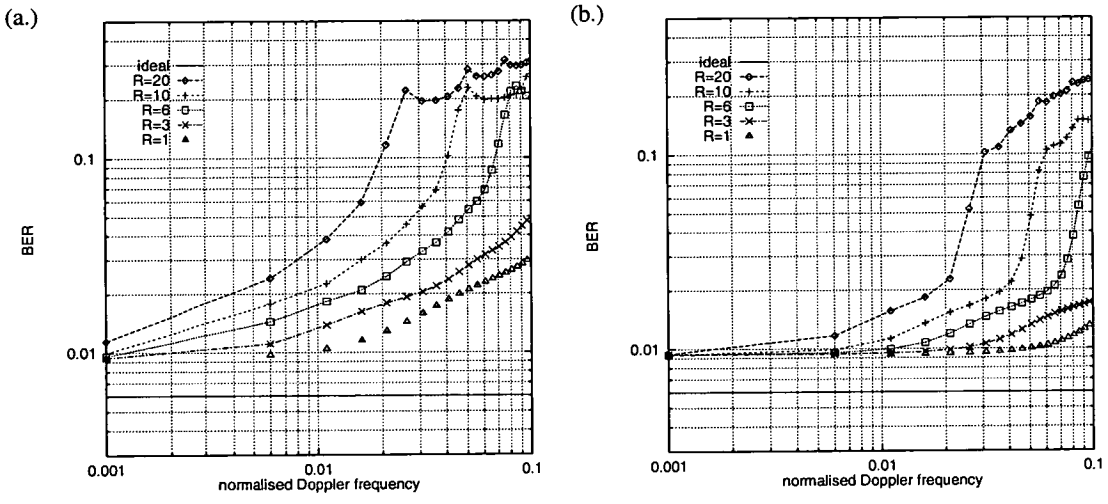


Figure 4.5: *BER vs normalised Doppler frequency, ν'_{max}, for pilot multiplexing rates, R; (a.) $\Delta\kappa = 0$ (linear predictor) and (b.) $\Delta\kappa = M/2$ (smoother).*

to be constant relative to Q, if the SNR is reasonably high ($\bar{\gamma} \geq 5$ dB), being approximately 3 dB for $\Delta\kappa = 0$ and 1 dB for $\Delta\kappa = M/2$. For low SNR values ($\bar{\gamma} < 5$ dB), however, the difference between the PA–RAKE and the lower bound increases as diversity is introduced to the system, comparing the graphs for $Q = 1$ and $Q = 2$. This is because, increasing the number of diversity taps Q, effectively means a decrease of the SNR per tap, $\bar{\gamma}_q$ of (4.2), due to a constraint input SNR $\bar{\gamma} = E_s/N_0$. This results in an increased MSE $\mathrm{E}\left[|\epsilon_q(k)|^2\right]$, since channel estimation is performed for every tap separately, which will in turn reduce $\bar{\gamma}_{out}$ according to (4.3). This affects the system performance of the PA–RAKE for low SNR, $\bar{\gamma} < 5$ dB.

The bit error probability degrades, however, when the maximum Doppler frequency, ν_{max}, is increased, as shown in Figure 4.5 for different numbers of multiplexing rates R. For low Doppler frequencies ($\nu'_{max} < 0.01$ for $R = 10$), the dependence on R is more pronounced for the linear predictor in Figure 4.5.a, with respect to the smoother in Figure 4.5.b. So, less system overhead, i.e. a lower multiplexing rate R, may be achieved at the expense of a larger decision delay $\Delta\kappa R$ by using a smoother

instead of a predictor. For large Doppler ν'_{max}, however, the PA–RAKE breaks down. The reason is, the PA–RAKE performs channel estimation with a R times lower sampling rate, i.e. if the data rate is $1/T_s$ the sampling rate for the pilots becomes $1/(RT_s)$. Moreover, if the Doppler frequency exceeds $\nu'_{max} \geq 1/(2R)$, aliasing effects occur. It is seen that by doubling R, the maximum Doppler frequency, ν'_{max}, to achieve the same error probability, is approximately halved. On the other hand, for low Doppler ($\nu_{max} \leq .001$) there is little difference between the smoother and linear predictor and the performance is also hardly dependent on R. The curve labelled $R = 1$ shows the performance of a reference system which uses every sample for channel estimation, i.e. each sample is a pilot (this may be achieved by a pilot channel, or to neglect decision feedback effects), serving as a lower bound for the system performance.

4.2.2 Iterative channel estimation

Iterative channel estimation (ICE) is a method that uses both pilot and data symbols to estimate the CIR. The word iterative indicates that the received signal is processed in two (or more) stages. In the 1st stage an initial channel estimate is made using only the pilot symbols, and then channel estimates are refined in one or more iterations by using both pilot and data symbols. Iterative channel estimation for a frame structure used for UMTS was developed by Schidl *et al* [70], and it was shown that the quality of the channel estimates can be significantly improved with ICE compared to non-interative channel estimation. Due to the sampling rate of the pilots $f_P = 1/(RT_s)$, being a factor R under-sampled with respect to the data rate, channel estimation may be improved by incorporating data symbols in the estimation process. In order to do that, the modulation of the data symbols is to be removed by means of decision feedback, thus introducing error propagation.

With a 2–stage receiver, data decisions concerning the future symbols of the 2nd stage are provided by the output of the 1st stage. In the 1st stage the pilot aided receiver discussed in the previous section 4.2 is employed, to make tentative decisions on the symbols. For the 2nd stage a smoothing type FIR filter is employed, using $M/2$ tentative decisions $\{\widehat{d}^{(1)}(k)\}$ of the 1st and the final decisions $\{\widehat{d}^{(2)}(k)\}$ of the 2nd stage, respectively, given by

$$\begin{aligned} \widehat{h}_q^{(2)}(k-M/2) &= \sum_{\substack{m=1 \\ m \neq M/2}}^{M} w_m^{(2)*}\, \widehat{d}^{(i)}(k-m+1)\, y_q(k-m+1)\;; \quad i = \begin{cases} 1, & m < M/2 \\ 2, & m > M/2 \end{cases} \\ &= \mathbf{w}^{(2)T} \tilde{\mathbf{y}}_q(k) \end{aligned} \tag{4.16}$$

where the superscript $i = \{1, 2\}$ denotes the decision output of the 1st and 2nd stage, respectively. The filter $\mathbf{w}^{(2)} = [w_1^{(2)}, \ldots, w_{M/2-1}^{(2)}, 0, w_{M/2+1}^{(2)}, \ldots, w_{2,M}]^T$ denotes the vector containing the filter weights of the 2nd stage estimation filter of dimension $\mathbb{C}^{M+1}$. The 2nd stage smoother can be calculated using the Wiener-Hopf equation in (2.20). According to the discussion in section 2.1.4 the filter in (4.16) omits the coefficient $w_{M/2}^{(2)}$. This enables the receiver to generate the filter $\mathbf{w}^{(2)}$ based on noisy snapshots. A block diagram of the filter is shown in Figure 4.6. Finally, the decision variable of the 2nd stage is

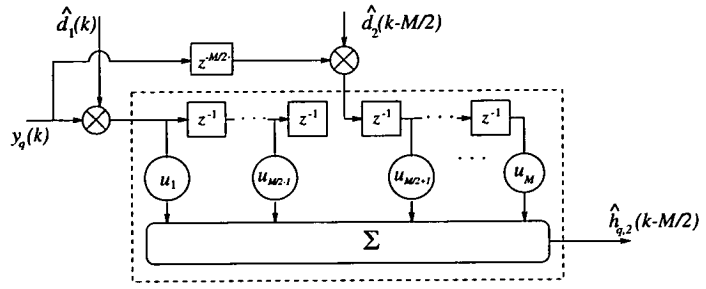


Figure 4.6: *The 2nd stage channel estimation filter.*

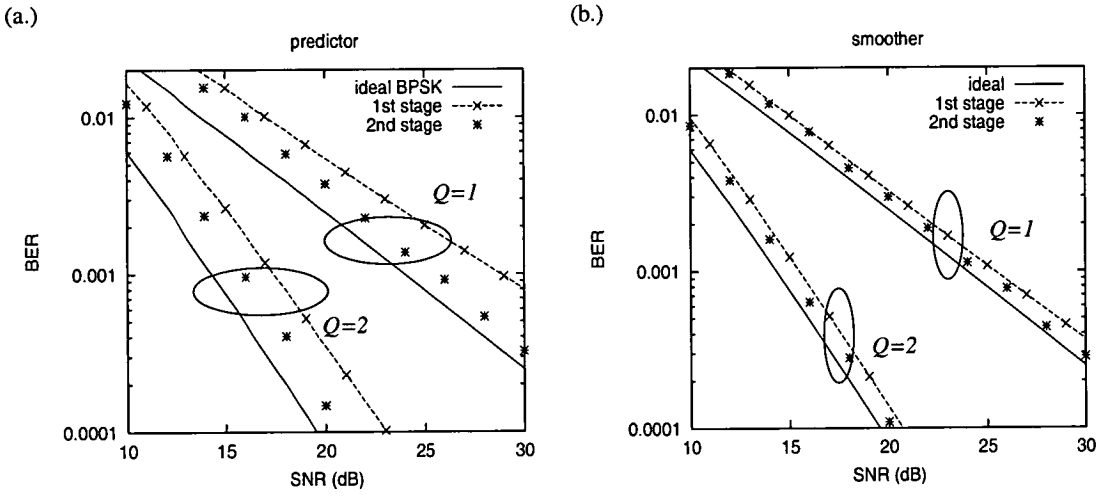


Figure 4.7: *BER vs SNR of an PA-RAKE with ICE for different numbers of diversity taps Q. (a.) $\Delta\kappa = 0$ (linear predictor) and (b.) $\Delta\kappa = M/2$ (smoother); $M = 4$, $M_2 = 8$.*

$$\widehat{\Lambda}^{(2)}(k) = \sum_{q=1}^{Q} \mathrm{Re}\{y_q^*(k)\,\widehat{h}_q^{(2)}(k)\} \tag{4.17}$$

and is then passed to the decision circuit to obtain $\widehat{d}^{(2)}(k)$.

Simulation Results For simulation work the same system model as in the previous section was used, defined in Table 4.1 and section 3.2.6. The benefit of ICE is assessed with computer simulations and compared with results from the previous section. For the 1st stage a PA-RAKE receiver with $M = 4$ coefficients was employed. The estimation filter was either (a.) a predictor with $\Delta\kappa = 0$, or (b.) a smoother with $\Delta\kappa = 2$. For the post-processing in the 2nd stage a smoothing type filter with $M_2 = 8$ coefficients for (a.) and (b.) was used. Simulation results of the system performance against the SNR are shown in Figure 4.7. Iterative channel estimation particularly improves the BER at high SNR. This is an expected result since the decisions of the 1st stage are more reliable and decision errors have less impact on the processing in the subsequent stages. The improvement of the receiver labelled (a.) is particularly impressive, bearing in mind that the linear predictor in (a.) of the 1st stage has no decision delay, compared to the smoother in (b.) which induces $R\Delta\kappa = 20$ symbols decision delay. However, in order to improve the performance the BER of the 1st stage needs to be below approximately $P_{e,1st} \leq 10^{-2}$.

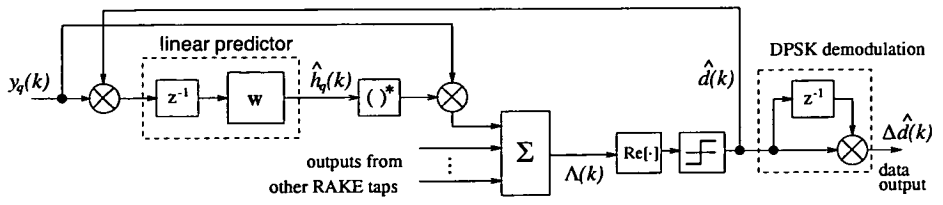


Figure 4.8: *Block diagram of the q^{th} tap of a decision directed (DD) diversity receiver.*

Consequently, with ICE the performance cannot be improved at low SNR or fast fading.

It was pointed out in [70] that the extra complexity to implement ICE is very modest compared to other DSP requirements in a mobile handset. Hence, ICE is easy to accommodate with today's DSP technology and is very attractive for applications in future mobile systems.

4.3 Decision directed channel estimation

If the transmitted signal is differentially encoded, a decision directed type of receiver may be used. Clearly, differential encoding of the data bits does not essentially require any knowledge of the CIR. However, in a fast fading environment, a conventional differential receiver suffers from an irreducible bit error rate (IBER), due to the induced phase lag of two adjacent samples (see Figure 4.2). The performance of conventional DPSK receiver in a fast fading environment may be significantly improved by means of DEPSK, if an accurate estimate of the CIR is available. For DEPSK detection is performed in analogy to coherent PSK and the receiver output $\widehat{d}(k)$ of (4.1) is then differentially decoded. Employing a decision directed receiver, the data modulation of the received signal is removed by using *decision feedback*. The motivation behind this is, that for sufficiently high SNR, virtually all decisions are correct in the feedback loop. Using decision feedback the pre-multiplied received signal is $y'_q(k) = \widehat{d}(k)\, y_q(k)$, where $\widehat{d}(k)$ is the output from the decision circuit after the diversity combining. The pre-multiplied received signal, $y'_q(k)$, is then used for channel estimation.

4.3.1 Linear predictive channel estimation

For channel estimation, a 1–step linear predictor from (2.17) is employed, having the form

$$\widehat{h}_q(k) \;=\; \sum_{m=1}^{M} w_m^*\, y'_q(k-m) = \mathbf{w}^H \mathbf{y}'_q(k-1) \tag{4.18}$$

where the filter $\mathbf{w} = [w_1, \cdots, w_M]^T$ is determined by the Wiener-Hopf equation in (2.18). The necessary phase reference is given by differentially encoding of the transmitted signal. Thus, the information bit is extracted from $\widehat{d}(k)$ in (4.1) by $\Delta\widehat{d}(k) = \widehat{d}(k)\;\widehat{d}^*(k-1)$. The obtained receiver is termed decision directed RAKE receiver (DD–RAKE). A block diagram of the DD–RAKE is shown in Figure 4.8. Another possibility of performing an one-step prediction is to use an infinite impulse response (IIR) filter (see section 2.1.5). The 1^{st} order IIR filter applied to decision directed diversity reception has been

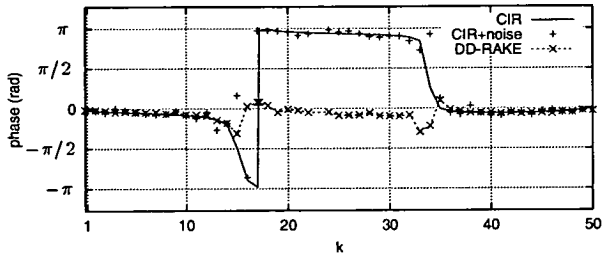


Figure 4.9: *Phase of the channel estimate* $\arg\left[\widehat{h}(k)\right]$ *vs time* k*, compared to the CIR with and without AWGN.*

studied for example in [61]. Unfortunately, the IBER of a conventional differential receiver cannot be lowered with that technique [61].

In the following, the work of Laurenson and Povey [55,71] is continued to study the effects of error propagation, due to decision feedback. In order to assess the decision feedback effects, the DD–RAKE is compared with a receiver where all the decisions are correct in the feedback loop, i.e. $y'_q(k) = d(k)\, y_q(k)$. The hypothetical decision $\widehat{d}(k)$ has been replaced with the true transmitted symbol, $d(k)$. This is termed the decision aided RAKE receiver (DA–RAKE). The assumption $\widehat{d}(k) = d(k)$ is often made for analysis purposes, as decision feedback effects are difficult to analyse. In a practical situation this may be achieved by employing a pilot channel.[2]

The phase slip effect Since channel prediction is performed on a decision directed basis, error propagation effects occur, which may lead to a so-called *phase slip*. Figure 4.9 illustrates a phase slip of the estimated CIR, $\arg\left[\widehat{h}(k)\right]$, in the time domain, compared to the true CIR with and without AWGN. It is seen that the receiver can be locked in a false state, where the channel estimator phase is flipped (shifted 180^o relative to the CIR phase) i.e. $\arg\left[\widehat{h}(k)\right] = \arg\left[h(k)\right] + \varphi_\epsilon \pm \pi$, where φ_ϵ denotes the prediction error phase. This can be observed in the graph between samples $k \approx [17, 33]$. In this interval the sign of the detected bit is the conjugate of the actual transmitted bit, that is $\widehat{d}(k+n) = -d(k+n)$. The receiver is entering the false state after an error burst, i.e. a series of decision errors, which may occur during a deep fade. Such an error burst causes the decision directed predictor in (4.18) to lose track of the CIR phase. After the occurrence of a phase slip the predictor remains locked until the following phase slip, as shown in the graph. While this effect causes severe degradation for a coherent receiver, no such implications are observed for differential encoding, since only the difference of the phases of two consecutive bits are considered and not its absolute values.

Simulation Results The theoretical analysis of a decision directed receiver is somewhat difficult as the number of decision errors and their distribution is of crucial importance for the system performance. Thus, the error probability is assessed by means of Monte Carlo simulations. As a lower bound a DD–RAKE receiver with *a priori* knowledge of the CIR was employed (for a closed form expression see

[2]Note, although such a receiver does not need to employ differental modulation as it is assumed here. This has been done solely for comparison purposes with the DD–RAKE.

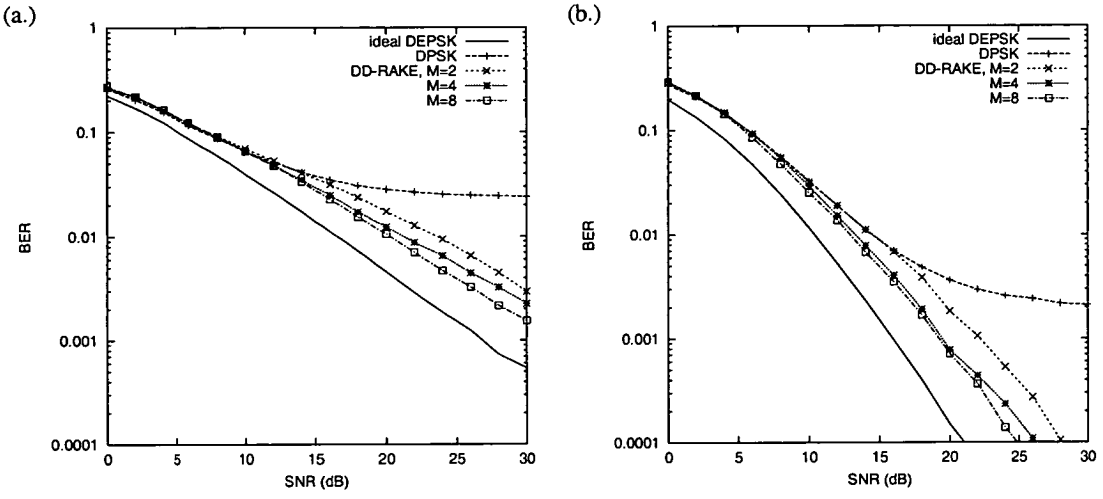


Figure 4.10: *BER vs SNR for different receiver realisations; (a.) $Q = 1$, and (b.) $Q = 2$.*

e.g. [61]). Simulation work is based on the specification in Table 4.2 for a complex baseband urban channel, defined in section 3.2.6. The performance of the system was evaluated for one and two diversity taps. The statistics of the qth diversity tap are specified by the classical Doppler power spectra from (3.12), due to Clarke [52], described by the ACF $\phi_{hh,q}(\Delta k) = J_0(2\pi\nu'_{max}\Delta k)$. The normalised maximum Doppler spread was chosen to be $\nu'_{max} = 0.05$. Generally, curves labelled "ideal DEPSK" show the results when the CIR is known *a priori* to the receiver, which serves as a lower bound. An upper bound is given by a conventional DPSK receiver in (4.6) (label "DPSK").

Number of diversity taps	Q	{1,2}
Mean SNR	$\bar{\gamma}$	15 dB
Doppler frequency	ν'_{max}	0.05
Modulation		DEPSK

Table 4.2: *System & simulation parameters for the decision directed (DD) receiver.*

The performance drawn against the SNR, of a receiver using a linear prediction filter, for $Q = 1, 2$ diversity taps, is considered in the Figures 4.10 and 4.11. It can be observed from Figure 4.10 for both parts (a.) and (b.), that the DD–RAKE significantly lowers the error floor of the conventional DPSK receiver. If the SNR exceeds $\bar{\gamma} \approx 10$ dB the $M = 4$ predictor performs considerably better than the one with $M = 2$ coefficients. Only for high SNR ($\bar{\gamma} > 20$ dB) can an improvement be achieved for $M > 4$. For low SNR the DD–RAKE cannot improve the performance of the conventional DPSK receiver, until the latter runs into its irreducible error floor.

By comparing Figure 4.10 part (a.) for a flat fading channel and (b.) for channel with double diversity, it is seen that the BER performance of the system can be improved by employing diversity techniques. For low SNR values ($\bar{\gamma} < 10$ dB), however, the difference between the DD–RAKE and the receiver with *a priori* knowledge of the CIR increases as diversity is introduced to the system, in analogy to the PA–RAKE plot in Figure 4.4. This is because, increasing the number of diversity taps Q, effectively means a decrease of the SNR per tap $\bar{\gamma}_q$ of (4.2), due to a constraint input SNR $\bar{\gamma} = E_s/N_0$. This results in an increased MSE $\mathrm{E}\left[\,|\epsilon_q(k)|^2\,\right]$, since channel estimation is performed for every tap separately, which will

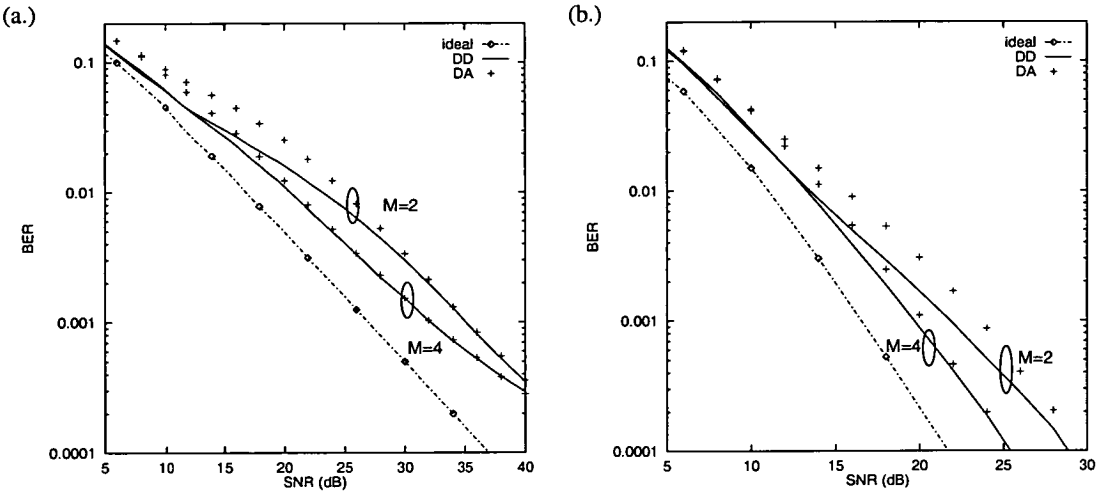


Figure 4.11: *BER vs SNR performance of a decision directed DD and DA–RAKE receiver; (a.) $Q = 1$, and (b.) $Q = 2$.*

in turn reduce $\bar{\gamma}_{out}$. Moreover, a decreasing $\bar{\gamma}_q$ results in more and longer error bursts, leading to more severe error propagation. Thus, the benefit of diversity is partly cancelled out by the enlarged estimation error. For larger SNR values, on the other hand, error propagation is not a major source of decision errors. The MSE of the channel estimate $\widehat{h}_q(k)$ is dominated by the induced phase lag of $\widehat{h}_q(k)$ relative to the CIR $h_q(k)$. Thus for high SNR, the difference between the DD–RAKE and the receiver with *a priori* knowledge of the CIR, does not significantly increase as diversity is introduced to the system.

The receiver performance of a DD–RAKE receiver with filter order $M = 2$ and $M = 4$ is shown in Figure 4.11, for $Q = 1$ and $Q = 2$ taps. It is seen that the DD–RAKE performance with the same number of coefficients is never worse than the corresponding DA–RAKE. Furthermore, the DD–RAKE performance is significantly better for low SNR values ($\bar{\gamma} \leq 20$ dB for $M = 4$). In particular, the fewer coefficients M the predictor has, the more significant is the difference between the DD and DA–RAKE. This effect was investigated in [72, 73] and will be analysed in the following. Note that for differential modulation, an error in $\widehat{d}(k)$ is likely to cause two consecutive errors in $\Delta\widehat{d}(k)$ and $\Delta\widehat{d}(k+1)$, since $\Delta\widehat{d}(k) = \widehat{d}(k)\,\widehat{d}(k-1)$. However, for the DD–RAKE, the subsequent error caused by $\widehat{d}(k+1)$ may be cancelled out by a phase slip. Recall that for binary modulation a phase slip translates into an alternation of the sign, such that $\widehat{d}(k+n) = -d(k+n)$ for $n > 0$. So, consider the detection of $\Delta\widehat{d}(k+1) = \widehat{d}(k+1)\,\widehat{d}(k)$: here the *additional* error induced by differential modulation due to $\widehat{d}(k)$, may be cancelled out by a phase slip in $\widehat{d}(k+1)$; since 2 consecutive errors result in a correct detection of $\Delta\widehat{d}(k+1)$. In other words, a single error may be sufficient to cause a phase slip, particularly for a short predictor. Thus, the phase slip effect is more pronounced for $M = 2$ compared to $M = 4$. For higher SNRs, a phase slip becomes less likely due to the decreased occurrence of error bursts. Hence the DD and DA–RAKE curves merge.

Figure 4.11.b shows the BER against the SNR of a DD and DA–RAKE receiver with $Q = 2$ taps. Similar results are observed as in the flat fading case, for the DD–RAKE relative to the DA–RAKE.

Figure 4.12 shows the above described effect in the frequency range. The BER performance of a $Q = 2$ tap DD and DA–RAKE receiver are plotted against the normalised Doppler frequency $\nu_{max}T_s$ on a

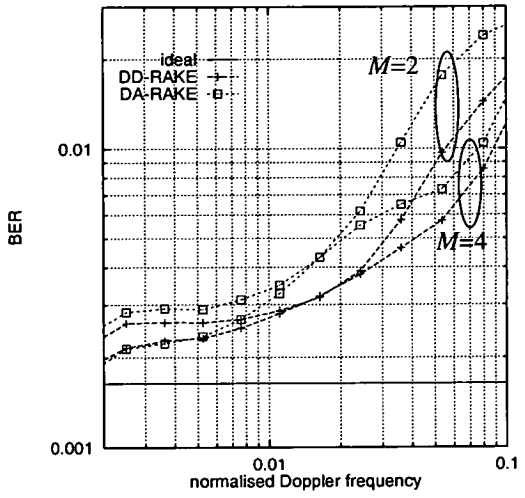


Figure 4.12: *BER vs ν'_{max} performance of a DD and DA–RAKE with two diversity taps; $Q=2$, $\bar{\gamma} = 15$ dB.*

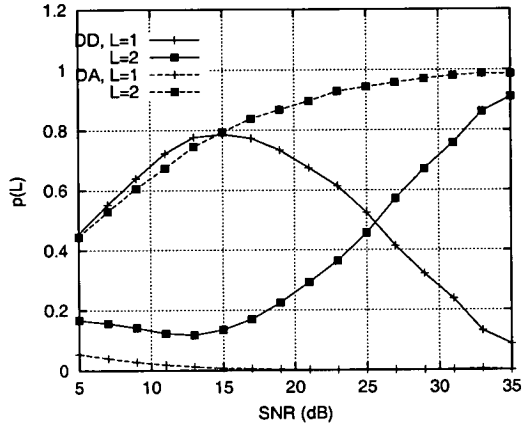


Figure 4.13: *Probability of an error burst of length $L_b = 1,\ 2$ vs SNR for a $M = 2$ predictor.*

SNR of $\bar{\gamma} = 15$ dB. It can be observed that the DD–RAKE performs always better than the DA–RAKE, for arbitrary Doppler frequencies. The difference starts to become significant for a Doppler spread of $\nu'_{max} > 0.02$ $(M = 2)$ and $\nu'_{max} > 0.01$ $(M = 4)$. Furthermore, for modest Doppler $(\nu'_{max} < 0.02)$, the DD–RAKE performance with $M = 2$ and $M = 4$ merge.

Error propagation analysis The fact that decision feedback effects improve the system performance is an interesting and unexpected result. In the literature, where the performance of similar receivers was derived analytically, the assumption of no feedback errors was referred to be a lower bound [28, 35, 74], which it is obviously not, according to these results. Note that this only applies to binary differential modulation. It should also be noted, that if there's a phase reference available, i.e. a pilot, coherent detection for the DA–RAKE can be employed with superior performance. The reason why a coherent receiver has not been implemented here is to assess the effects of error propagation.

Figure 4.13 illustrates the statistical analysis of the phase slip effect. There the probability of an error

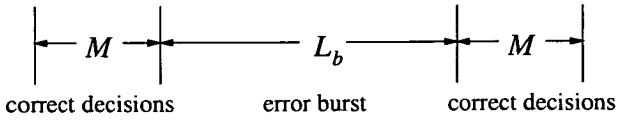


Figure 4.14: *Definition of an error burst.*

burst of length L_b is drawn against the SNR, for a $M = 2$ predictor. The number of bits L_b between two consecutive error-free regions, that are at least M bits in length, are defined as an error burst, shown in Figure 4.14. The probability that an observed error burst has the length L_b, is defined by

$$p(L_b) = \text{Prob}\{L_i = L_b\} \tag{4.19}$$

where $\{L_i\}$ is a set of randomly distributed error bursts obtained by Monte Carlo simulations. This calculation of $p(L_b)$ was repeated for a number of SNRs.

By examining Figure 4.13 it is seen that a single error burst is far more likely for the DD than for the DA system for a SNR somewhat smaller than 25 dB. In particular for SNRs $\bar{\gamma} \approx 15$ dB, the probability that an error burst is a single error reaches 80%, while the corresponding probability of the DA–RAKE is negligible. For high SNR values virtually all errors are 2 consecutive errors for both systems. The probability for an error burst larger than 2 is for both systems the same, being small for low and negligible for high SNRs.

4.3.2 Iterative channel estimation

According to the pilot aided (PA) receiver, an improved approximation of the CIR of the DD–RAKE can be made, if a smoother rather than a linear predictor is employed for estimation the CIR [73]. The phase lag, and therefore the mean squared error (MSE) of the channel estimate becomes smaller if future samples in addition to the past samples are used. The DD–RAKE however, needs a causal estimation filter, so future samples cannot be used straight away, since a decision upon a data symbol needs to be done prior to estimation. The data symbol is needed in the feedback loop to demodulate the received signal. The solution is provided by post-processing of the received signal via iterative channel estimation [39]. With a 2–stage receiver, data decisions concerning the future symbols of the 2nd stage are provided by the output of the 1st stage. In the 1st stage the linear predictor discussed in the previous section is employed, to make tentative decisions on the symbols, using for each symbol only the past received signals. The estimated CIR for the 2nd stage, using $M/2$ tentative decisions $\{\widehat{d}^{(1)}(k)\}$ of the 1st and the final decisions $\{\widehat{d}^{(2)}(k)\}$ of the 2nd stage, respectively, is given by (4.16). After an added decision delay of $M/2$ symbols, differential decoding of the decision variable in (4.17) yields the decision of the desired information symbol. A block diagram of the resulting receiver structure is depicted in Figure 4.15.

For the reference system in the form of the DA receiver, no 2–stage processing is necessary, as the unmodulated received signal, $y'_q(k)$, is assumed to be available. For comparison purposes the receiver output is again DEPSK modulated, in accordance to the linear predictor DA–RAKE.

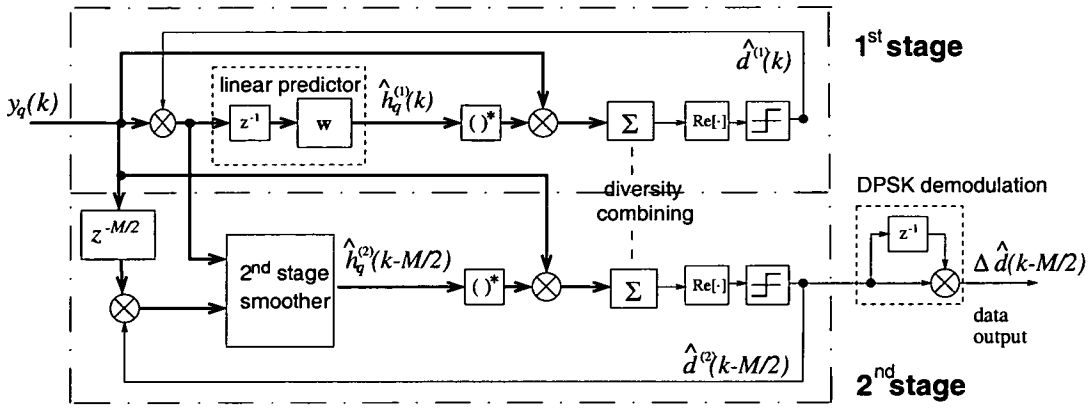


Figure 4.15: *Decision directed 2–stage receiver structure of the q^{th} diversity tap.*

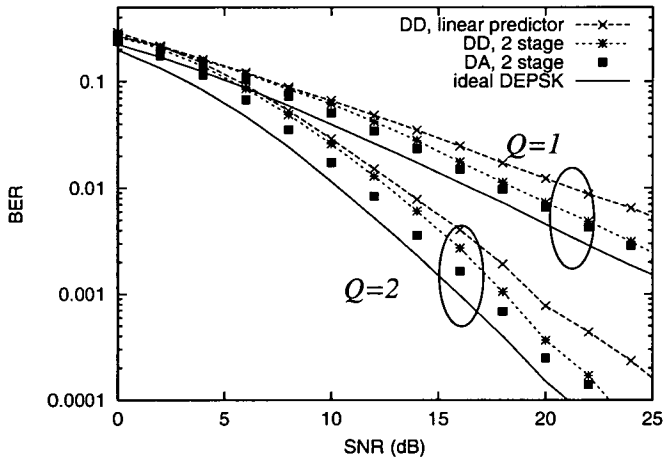


Figure 4.16: *BER vs SNR performance of a 2–stage DD–RAKE and the reference DA receiver with a smoothing type estimator, with one and two diversity taps; $M = 4$.*

Simulation Results For simulation work the same system model and parameters as in the previous section were used, defined by section 3.2.6 and Table 4.2. The BER versus mean SNR $\bar{\gamma}$ for the 2–stage receiver is shown in Figure 4.16. If the SNR is somewhat larger than 10 dB, processing $y_q(k)$ in 2–stages becomes worthwhile. For instance, the difference towards the conventional linear predictor with the same number of coefficients exceeds 2 dB for BER $= 10^{-3}$ and $Q = 2$ taps. For high SNR values, the difference between the linear predictive and the 2–stage receiver gets larger as the phase lag of the 1st stage (the linear predictor) becomes the major source of errors. The 2–stage receiver, on the other hand, has a smaller phase lag. Moreover, error events occur mainly as single errors, leading to negligible error propagation. Hence, the BER of the 2nd stage relative to the 1st stage is significantly lowered.
Unlike for the linear predictor plotted in Figure 4.11, the receiver performance of the 2–stage DD–RAKE is not superior compared to the corresponding DA–RAKE, shown in Figure 4.16. Obviously, decision errors are no longer beneficial to the performance. As mentioned earlier, a phase slips at the output of the 1st stage result in a phase change of 180°, i.e. $\arg\{\widehat{h}(k)\} = \arg\{h(k)\} \pm \pi$, as shown in Figure 4.9. The 2nd stage filter smoothes this sudden phase change into a gradual one which may last several samples,

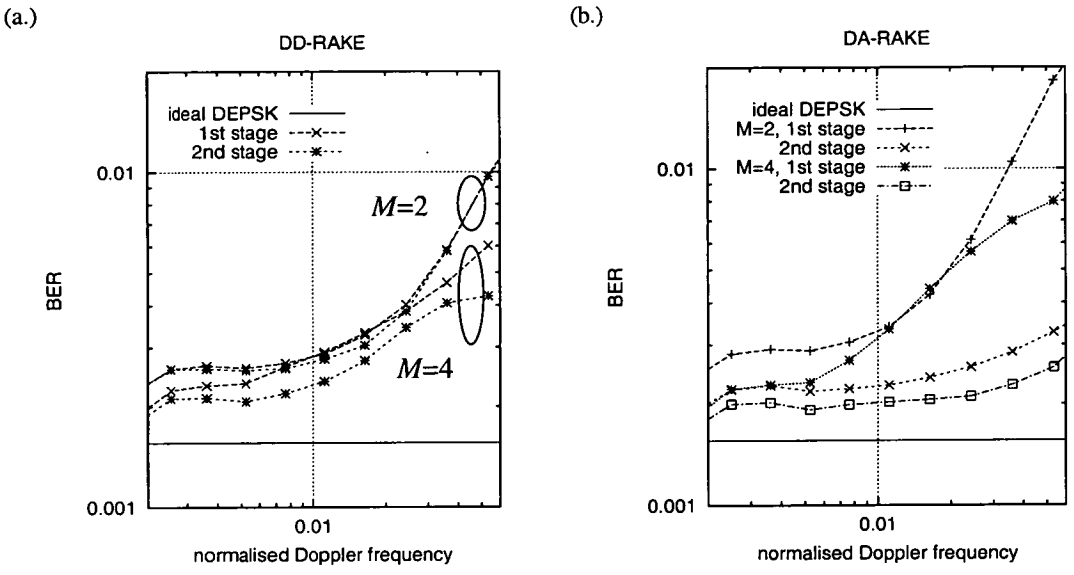


Figure 4.17: *BER vs ν'_{max} performance of a (a.) DD–RAKE and (b.) DA–RAKE with two diversity taps; $\bar{\gamma} = 15$ dB.*

resulting in additional decision errors.

Figure 4.17 shows the benefit of the 2nd stage processing as a function of ν'_{max}. It is seen that the benefit of the 2-stage processing for the DA–RAKE (Figure 4.17.b) is much higher than for the DD–RAKE (Figure 4.17.a). In particular, in Figure 4.17.a the DD–RAKE for $M = 2$ shows virtually no different between 1st and 2nd stage outputs, unless the Doppler spread is very high.

To summarise, the higher the SNR and Doppler spread, the greater the improvement in system performance by using a 2–stage receiver. However, the improvement of the 2nd stage processing is partly cancelled out due to decision feedback effects.

4.4 Summary and conclusions

Receiver structures were discussed in this chapter which detect the received signal on a symbol–by–symbol basis. Channel estimation was performed using either time multiplexed pilot symbols or in a decision directed manner. Due to the characteristics of the obtained receivers they are applicable to different fading rates: the pilot aided (PA) receiver is suitable for modest fading rates, whereas the decision directed receiver (DD) receiver can improve the performance compared to conventional DPSK significantly for high fading rates. The features of the PA and DD–RAKE are summarised below.

PA–RAKE: Only pilot symbols are used for channel estimation, so no decision feedback is employed. Therefore this receiver is robust if the fading rate ν'_{max} allows some degree of oversampling. For applications limited to fading rates which allow an oversampling factor of $\beta \geq 10$, the PA–RAKE yields significant improved performance compared to conventional DPSK, due to coherent detection. With a time delay $\Delta\kappa > 0$ a smoothing-type estimation filter can be employed which further improves the performance, at the expense of an imposed decision delay of $\Delta\kappa R$ samples.

DD–RAKE: The receiver performs decision directed channel estimation. Differential encoding of the data bits is required to ensure the receiver is robust. This means that the slow fading performance is inferior to the PA–RAKE and the performance compared to conventional DPSK can only be marginally improved. For fast fading, however, the DD–RAKE considerably lowers the error floor seen for conventional DPSK. As an unexpected result, decision feedback effects did improve the performance compared to a reference receiver, which always used correct decisions for channel prediction.

Post processing by means of iterative channel estimation (ICE) further improves the performance of both the PA and DD–RAKE, given the BER of the 1^{st} stage is reasonably low (BER $< 5 \cdot 10^{-2}$).

Chapter 5
Joint Detection & Estimation

In this chapter optimal and near optimal algorithms for data detection and parameter estimation on a Rayleigh fading channel are examined. Following previous work on detection and estimation theory, the optimum estimator–detector receiver is derived. This chapter focuses on the methodology and fundamental insights rather than on details of implementation; these are addressed in the following chapter on realizable receiver structures. Although the optimal algorithms suffer from an extremely high complexity, the ideas and strategies worked out here can be applied to the systematic development of suboptimal but realizable algorithms, which will be addressed in Chapter 6.

Even for uncoded modulation and no inter-symbol interference (ISI) present, the optimum receiver involves the entire transmitted sequence in the detection–estimation process, exploiting the high correlation of adjacent samples of the channel impulse response (CIR) to form its estimate.

Design criteria on how to optimise the receiver were discussed in section 2.2. The main part of this chapter is dedicated to maximum likelihood sequence detection (MLSD). First the estimator-correlator receiver structure studied by Kailath [31] will be reviewed in section 5.1. Then recursive realisations of optimum MLSD, and the application of a sequential decoding algorithm, known as the *Viterbi algorithm*, are addressed in section 5.2. Finally, the maximum *a posteriori* (MAP) symbol-by-symbol detector will be discussed very briefly in section 5.3, where the similarities between algorithms for MLSD and MAP symbol-by-symbol detection will be pointed out.

5.1 Optimal maximum likelihood sequence detector

The basic idea of joint detection and channel estimation is the search for the best overall fit between the model output (hypothetical data sequence transmitted over its associated hypothetical channel) and the observation (received signal), often aided by some side information on the channel dynamics. The receiver structure addressed in this section is based on the work of Kailath [31–34]. In [31–33] the optimum receiver was derived based on a discrete-time model assuming Gaussian statistics throughout; whilst in [34] a continuous-time model was adopted for the more general case of a random signal corrupted by white Gaussian noise.

Bayesian detection has been addressed in section 2.2. In this section detection of the whole sequence is considered. Maximising the likelihood function $p(\mathbf{y}|\mathbf{d}^{(o)})$ for the whole sequence $\mathbf{d}^{(o)}$ is termed maximum likelihood sequence detection (MLSD) and it was shown in 2.2 that the MAP criterion is equivalent to ML criterion, if all transmitted sequences are equally likely. The transmission of a linearly modulated

data sequence $\mathbf{d}^{(\ell)}$, perturbed by the time variant CIR $\mathbf{h}$ and noise vector $\mathbf{n}$, is considered. Define the received signal, conditioned on the hypothesis that ℓ was transmitted as

$$\mathbf{y} = \mathbf{D}^{(\ell)}\mathbf{h} + \mathbf{n} \tag{5.1}$$

The matrix $\mathbf{D}^{(\ell)}$ may contain several copies of the transmitted sequence $\mathbf{d}^{(\ell)}$ given the frequency selective nature of the fading channel (see Chapter 3). At this point the matrix $\mathbf{D}^{(\ell)}$ is not further specified, this will be done later on when more specific systems are considered. Assuming a complex Gaussian distribution for the CIR vector $\mathbf{h}$ and the noise $\mathbf{n}$, with both zero mean, i.e. $\mathrm{E}[\,\mathbf{h}\,] = 0$ and $\mathrm{E}[\,\mathbf{n}\,] = 0$; a multivariate Gaussian pdf for the received signal is obtained. The pdf of receiving $\mathbf{y}$, conditioned on $\mathbf{d}^{(\ell)}$ being transmitted, $p(\mathbf{y}|\,\mathbf{d}^{(\ell)})$, has zero mean and the data dependent covariance $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)} = \mathrm{E}\big[\mathbf{y}\mathbf{y}^H \,\big|\, \mathbf{d}^{(\ell)}\big]$ of dimension $\mathbb{C}^{N\times N}$, given by

$$p(\mathbf{y}|\,\mathbf{d}^{(\ell)}) \;=\; \pi^{-N}\det\big[\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}\big]^{-1}\exp(-\mathbf{y}^H\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}\,\mathbf{y}) \tag{5.2}$$

where N is the dimension of $\mathbf{y}$ and det denotes the matrix determinant operation. This is the pdf that describes the optimum receiver. To simplify the computations, the natural logarithm of $p(\mathbf{y}_q|\,\mathbf{d}^{(\ell)})$ can be applied, giving the log-likelihood function

$$\Lambda(\ell) \;=\; -\ln\big(p(\mathbf{y}|\,\mathbf{d}^{(\ell)})\big) \;=\; \mathbf{y}^H\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}\,\mathbf{y} - \ln\big(\pi^N\det\big[\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}\big]\big) \tag{5.3}$$

Assuming a purely phase modulated signal, the determinant of $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}$ is not data dependent, since all hypotheses have equal power, which can thus be neglected [28,31]. Hence, the ML criterion of the q^{th} diversity tap corresponds to the decision variable [31]

$$\Lambda(\ell) \;=\; \mathbf{y}^H\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}\,\mathbf{y} \tag{5.4}$$

which is essentially the negative exponent of the pdf $p(\mathbf{y}|\,\mathbf{d}^{(\ell)})$. Since $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}$ is a Hermitian matrix, the decision variable $\Lambda(\ell)$ is a real function. The maximum of $p(\mathbf{y}|\,\mathbf{d}^{(\ell)})$ over $\mathbf{d}^{(\ell)}$ is equivalent to finding the sequence that minimises $\Lambda(\ell)$.

The decision variable $\Lambda(\ell)$, which is a quadratic form, determines the operations that the receiver will have to perform. For the derivation of the decision variable the following assumptions have been made:

(i.) The decision variable is optimum for linearly modulated signals where both the CIR and the noise are Gaussian. Note that the receiver can be regarded as a generalised form of maximal ratio combining (MRC) [32]. Since the structure of the matrix $\mathbf{D}^{(\ell)}$ is arbitrary, this includes channels with inter-symbol interference (ISI).

(ii.) For constant envelope signals (5.4) applies, otherwise (5.3). For more general modulation schemes the second term in (5.3) is much smaller than the quadratic term $\mathbf{y}^H\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}\,\mathbf{y}$ and is thus commonly neglected anyway [28].

(iii.) Here only the case where the CIR is Rayleigh distributed is discussed. For the more general case of a Rician distribution, where the CIR has the mean $\bar{\mathbf{h}} = \mathrm{E}[\mathbf{h}]$, the CIR can be partitioned into a random part $\mathbf{h}$ and a deterministic part $\bar{\mathbf{h}}$, given by $\mathbf{h}_r = \bar{\mathbf{h}} + \mathbf{h}$. The receiver structure for that case is addressed in [31–33]. This also generalises the receiver to the case where the channel is deterministic, $\mathbf{h}_r = \bar{\mathbf{h}}$, which is not covered in (5.4).

(iv.) Although the receiver is optimum if the noise $\mathbf{n}$ is coloured, it is desirable to pre-process $\mathbf{y}$ with a *whitening filter* such that $\mathbf{n}'$ is additive white Gaussian noise (AWGN) with covariance $\mathbf{\Phi_{nn}} = N_0\mathbf{I}$ [31]. Thus, in the following the noise term $\mathbf{n}$ is assumed to be AWGN, since the case where the noise is coloured can always be transformed to a signal where the noise is white.

The decision of the optimum MLSD receiver is done at the end of the sequence by finding the decision variable with the minimum distance metric:

$$\widehat{\Lambda} = \min_{\ell \in \mathcal{A}_K} \Lambda(\ell)$$

For equally likely sequences this is equivalent to the MAP criterion (see section 2.2.1). Close studies of a detector utilising (5.4) for short distances are carried out in [75]. Obviously, the receiver minimising $\Lambda(\ell)$ is not implementable for long sequences, since $\mathcal{A}_K$ grows exponentially with the sequence length K. Also, for continuous transmission, decoding at the end of the sequence is clearly not feasible, because of the induced decision delay for real time applications. For long sequences the computation of the quadratic form $\mathbf{y}^H \mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1} \mathbf{y}$ has a vast computational complexity. A realization of (5.4) for a short observation interval, applied to DPSK, was developed by Dam [76].

5.1.1 Estimator–correlator structure of the receiver

In this section the estimator–correlator structure of the decision variable is established, which is yet not apparent from (5.4). However, by rewriting (5.4) using the MMSE estimator discussed previously in section 2.1, the estimator–correlator structure of the receiver becomes obvious. Note, the underlying linear model in (5.1) conditioned on hypothesis ℓ is equivalent to the linear model in (2.9). In analogy to (2.10) in section 2.1.2, the covariance matrix of $\mathbf{y}$ can be expressed as: $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)} = \mathbf{D}^{(\ell)}\,\mathbf{\Phi_{hh}}\,\mathbf{D}^{(\ell)H} + \mathbf{\Phi_{nn}}$ and $\mathbf{\Phi}_{\mathbf{hy}}^{(\ell)} = \mathbf{\Phi_{hh}}\,\mathbf{D}^{(\ell)H}$. Recall the notation for the Wiener smoothing matrix in (2.14), given by

$$\mathbf{W}^{(\ell)} = \mathbf{\Phi_{hh}}\,\mathbf{D}^{(\ell)H}\,\left[\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}\right]^{-1} = \mathbf{I} - \mathbf{\Phi_{nn}}\,\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}$$

With this notation the decision variable in (5.4) can be rewritten as [31]

$$\Lambda(\ell) = \mathbf{y}^H \mathbf{\Phi}_{\mathbf{nn}}^{-1} \mathbf{y} + \mathbf{y}^H \mathbf{\Phi}_{\mathbf{nn}}^{-1} \mathbf{W}^{(\ell)} \mathbf{y}$$

Considering AWGN the covariance of the noise term is $\mathbf{\Phi_{nn}} = N_0\mathbf{I}$. Furthermore, since the first term of the right-hand side of the above equation is independent of ℓ, only the second term needs to be taken

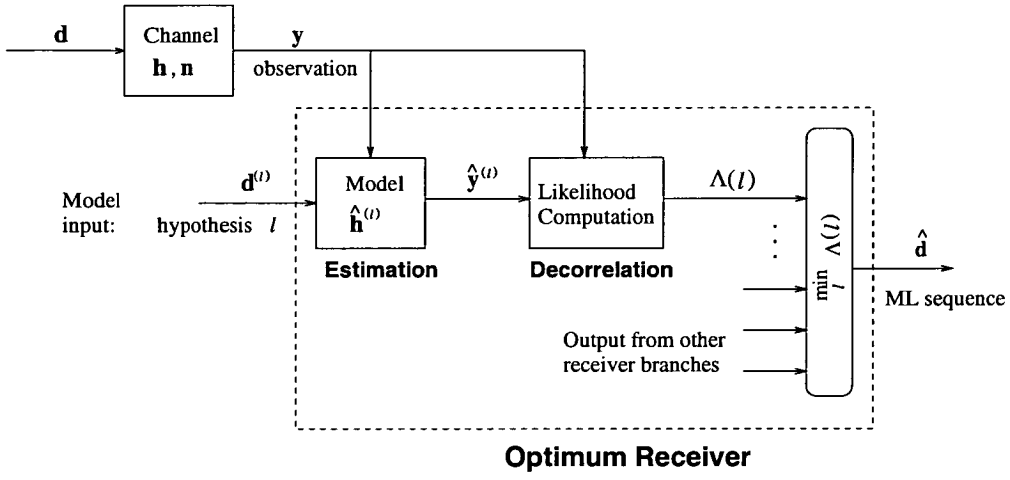


Figure 5.1: *Block diagram of the optimum estimator-correlator receiver.*

into account:

$$\begin{aligned}\Lambda(\ell) &= \mathbf{y}^H \mathbf{W}^{(\ell)} \mathbf{y} \\ &= \mathbf{y}^H \hat{\mathbf{y}}^{(\ell)} = \mathbf{y}^H \mathbf{D}^{(\ell)} \widehat{\mathbf{h}}^{(\ell)}\end{aligned} \tag{5.5}$$

where the constant noise power N_0 was also neglected.

Hence, a Wiener filter minimises the probability of error in the optimum detector. The resulting receiver structure follows Kailath's separation theorem [31]. That is the receiver consists of an estimator that delivers the MMSE estimates of the fading distortion and a detector that utilises these estimates, referred to as estimator–correlator receiver. Figure 5.1 shows a block diagram of the optimum estimator–correlator receiver, for transmission of a data sequence perturbed by the time variant CIR $\mathbf{h}$ and noise vector $\mathbf{n}$. The receiver can be divided into an estimation unit, a correlator and detector to decide upon the most likely transmitted sequence. The estimator models the channel, which is for a Rayleigh fading channel described by its covariance matrix, conditioned on hypothesis ℓ of the transmitted sequence. The model output is the estimated received signal $\hat{\mathbf{y}}$, which needs to be computed for every hypothesis ℓ. The correlator subsequently measures the "similarity", i.e. the probability that hypothesis ℓ is the transmitted sequence, $\mathbf{d}$, between model output $\hat{\mathbf{y}}$ and observation $\mathbf{y}$. The scalar output of the correlator is termed the decision variable $\Lambda(\ell)$, being a real number. The sequence which maximises this probability is then chosen by the detector, yielding the receiver output $\widehat{\mathbf{d}}$.

5.1.2 Nuisance parameters

It was seen by the estimator-correlator interpretation of the MLSD that the estimation of $\mathbf{h}$ is carried out inherently in calculating the quadratic form $\mathbf{y}^H \boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)\,-1} \mathbf{y}$, without explicitly estimating $\widehat{\mathbf{h}}^{(\ell)}$. The CIR $\mathbf{h}$ may be regarded as nuisance parameter [17], since the prime interest is in detecting $\mathbf{d}$ rather than estimating $\mathbf{h}$. Mathematically speaking, let $p(\mathbf{y}, \mathbf{h}|\mathbf{d}^{(\ell)})$ be the joint likelihood pdf for $\mathbf{h}$ and $\mathbf{y}$, the

likelihood pdf for $\mathbf{y}$ is determined by

$$\begin{aligned} p(\mathbf{y}|\mathbf{d}^{(\ell)}) &= \int_{-\infty}^{\infty} p(\mathbf{y}, \mathbf{h} \mid \mathbf{d}^{(\ell)})\, d\mathbf{h} \\ &= \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{h}, \mathbf{d}^{(\ell)})\, p(\mathbf{h})\, d\mathbf{h} \end{aligned}$$

This equation implies that nuisance parameters can be "integrated out" [16]. On the other hand, this expression may be used to design receivers which explicitly estimate $\mathbf{h}$ prior to detection of $\mathbf{d}$. The derivation for the optimum receiver applying this integral was used in [32, 33] yielding the decision variable in (5.5). On the other hand, an explicit joint detection & estimation receiver, based on maximising the pdf $p(\mathbf{y}, \mathbf{h}|\mathbf{d}^{(\ell)})$, was described in [28, Chapter 12].

In [77] an alternative solution was described. Choosing to estimate $\mathbf{h}$ prior to detection, the quadratic form may be rewritten as

$$\begin{aligned} \mathbf{y}^H \, \boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)\,-1} \, \mathbf{y} &= \left[\mathbf{y} - \widehat{\mathbf{y}}^{(\ell)}\right]^H \boldsymbol{\Phi}_{\mathbf{nn}}^{-1} \left[\mathbf{y} - \widehat{\mathbf{y}}^{(\ell)}\right] \\ &= \mathbf{y}^H \left[\mathbf{I} - \mathbf{V}^{(\ell)}\right]^H \boldsymbol{\Phi}_{\mathbf{nn}}^{-1} \left[\mathbf{I} - \mathbf{V}^{(\ell)}\right] \mathbf{y} \end{aligned} \tag{5.6}$$

where $\widehat{\mathbf{y}}^{(\ell)} = \mathbf{V}^{(\ell)}\mathbf{y}$ is the estimate of $\mathbf{y}$. An evaluation of $\mathbf{V}^{(\ell)}$ is given in [77], such that the above equation holds. It should be noted that $\mathbf{V}^{(\ell)}$ is not a MMSE estimator, thus its MSE is higher than the MSE of the Wiener smoothing matrix $\mathbf{W}^{(\ell)}$; the detector, however, is optimal in respect of detecting $\mathbf{d}$. However, the derivation of the receivers described in [28, 77] are rather tedious compared to the estimator-correlator interpretation of the MLSD. Instead, the focus will be on receivers having a recursive joint detection & estimation architecture in section 5.2, being far more practical.

5.1.3 Diversity reception

Even if a receiver that matches the ML criterion could be provided, the output sequence remains dependent upon the fading via the time varying SNR per symbol, which may deviate strongly from its average. As described in section 1.2.3 diversity can significantly improve the performance by virtue of its averaging effects, and hence approaches compliance to a more general optimum receiver design. An optimum receiver for diversity reception was derived and analysed by Turin in [78] and [79]. Kailath [80] showed how a diversity receiver can be derived as a special case of a multi-link channel, based on the results of a single link receiver as in [31]. In the special case of a spread spectrum communication system, implying transmission through a frequency selective channel, diversity can be provided by means of a Rake receiver [56]. According to [33], the optimum receiver can be interpreted as a generalisation of the Rake receiver due to Price & Green [56].

For the further discussion in this chapter the system model from section 3.3.1 is adopted, which is a rather general order Q diversity system. The received signal sequence of the q^{th} diversity tap of duration K can be expressed as

$$\mathbf{y}_q = \mathbf{D}\,\mathbf{h}_q + \mathbf{n}$$

where $\mathbf{D} = \text{diag}[d(1), \cdots, d(K)]$ is a diagonal matrix, containing the transmitted sequence. The matrix notation for an arbitrary signal $x(k)$ is defined by the column vector $\mathbf{x} \triangleq [x(1), \cdots, x(K)]^T$. The diagonal structure of $\mathbf{D}$ implies that there is no ISI present. For a brief discussion of signals corrupted by ISI and some references are given in section 5.2.5. Let $\mathbf{y} \in \mathbb{C}^N$ of dimension $N = QK$, be the received signal sequence containing the received sequences of the Q diversity taps, each of length K, defined by

$$\mathbf{y} \triangleq [\mathbf{y}_1^T, \cdots, \mathbf{y}_Q^T]^T = [y_1(1) \cdots y_1(K), \cdots, y_Q(1) \cdots y_Q(K)]^T$$

The Q taps are assumed to be mutually uncorrelated, hence the likelihood function is statistically independent with respect to the diversity taps. Accordingly, the likelihood function can be expressed as

$$p(\mathbf{y} | \mathbf{d}^{(\ell)}) = p(\mathbf{y}_1, \cdots, \mathbf{y}_Q | \mathbf{d}^{(\ell)}) = \prod_{q=1}^{Q} p(\mathbf{y}_q \mid \mathbf{d}^{(\ell)}) \tag{5.7}$$

Since the Q taps are mutually uncorrelated, the covariance matrix of $\mathbf{y}$ can be written in the block diagonal form $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)} = \text{diag}[\mathbf{\Phi}_{\mathbf{yy},1}^{(\ell)}, \cdots, \mathbf{\Phi}_{\mathbf{yy},Q}^{(\ell)}]$. As a result the decision variable in (5.4) breaks down to a sum of Q decision variables, given by

$$\Lambda(\ell) = \sum_{q=1}^{Q} \Lambda_q(\ell) = \sum_{q=1}^{Q} \mathbf{y}_q^H \mathbf{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1} \mathbf{y}_q \tag{5.8}$$

where $\Lambda_q(\ell) = \mathbf{y}_q^H \mathbf{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1} \mathbf{y}_q$ represents the decision variable of tap q.

5.1.4 Invariance over pre-multiplication

The receiver complexity can be grossly simplified if, for each hypothesis ℓ, the received signal is multiplied by the complex conjugate of the assumed transmitted signal, defined by [28, 75]

$$\mathbf{y}'_q(\ell) = \mathbf{D}^{(\ell)H} \mathbf{y}_q \tag{5.9}$$

where $\mathbf{D}^{(\ell)} = \text{diag}[d^{(\ell)}(1) \cdots d^{(\ell)}(K)]$ is a diagonal matrix containing the assumed transmitted sequence $\mathbf{d}^{(\ell)}$. The operation can be viewed as removing the assumed modulation from the received signal. Assuming a diagonal data matrix $\mathbf{D}^{(\ell)}$, it is shown in Appendix A.1 that $\Lambda(\ell)$ is invariant over pre-multiplication[1]

$$\Lambda(\ell) = \sum_{q=1}^{Q} \mathbf{y}_q'^H(\ell) \underbrace{[\mathbf{\Phi}_{\mathbf{hh},q} + N_0 \mathbf{I}]^{-1}}_{\mathbf{\Phi}_{\mathbf{y}'\mathbf{y}',q}} \mathbf{y}'_q(\ell) \tag{5.10}$$

and this notation for $\Lambda(\ell)$ is identical to (5.8). It is seen that the data dependence of the covariance matrix has been conveniently factored out. Thus, $\mathbf{\Phi}_{\mathbf{y}'\mathbf{y}',q}$ can be pre-computed and stored, if the channel statistics are known, unlike computing and inverting a covariance matrix for every hypothesis ℓ. The

[1]The constraint of the diagonal structure of $\mathbf{D}^{(\ell)}$ can be relaxed such that the product $\mathbf{D}^{(\ell)H}\mathbf{D}^{(\ell)}$ is constant, independent of ℓ. Unfortunately this is not the case for channels with ISI.

same is true for the Wiener smoothing matrix, which now becomes $\mathbf{W}_q = \mathbf{\Phi}_{\mathbf{hh},q}\,\mathbf{\Phi}^{-1}_{\mathbf{y'y'},q}$, for the q^{th} diversity tap. In accordance to section 5.1.1 the decision variable is of an estimator–correlator structure

$$\Lambda(\ell) = \sum_{q=1}^{Q} \mathbf{y}'_q(\ell)\,\mathbf{W}_q\,\mathbf{y}'^{H}_q(\ell) = \sum_{q=1}^{Q} \mathbf{y}'_q(\ell)\,\widehat{\mathbf{h}}^{(\ell)}_q \tag{5.11}$$

The MMSE estimate of sample k and tap q, $\widehat{h}^{(\ell)}_q(k)$, is determined by the k^{th} entry of $\widehat{\mathbf{h}}^{(\ell)}_q$.

Suppose the receiver can estimate the CIR aided by some side information, such that $\widehat{h}^{(\ell)}_q(k)$ becomes independent of ℓ. Then the decision variable of sample k becomes $\Lambda(\ell,k) = \sum_q y'_q(\ell,k)\,\widehat{h}_q(k)$. Now, the estimation and detection tasks are clearly separated, consequently detection can be carried out on a symbol–by–symbol basis. This bridges the gap to one-shot receiver structures studied in Chapter 4.

5.1.5 Performance analysis

To analyse the performance of a maximum likelihood sequence detector, first the decision variable $\Lambda(\ell)$, is cast into a quadratic form $\mathbf{y}^H\mathbf{Q}\,\mathbf{y}$, where $\mathbf{Q}$ will be defined below. Then the probability of a bit error can be calculated utilising Barrett's formula [60]. Only binary antipodal modulation, in the form of BPSK will be considered here, i.e. $A_m = 2$.

Let the actual transmitted sequence, $\mathbf{d}$, be the all-one sequence denoted by $\mathbf{d}^{(0)}$, and $\mathbf{d}^{(\ell)}$ be the sequence which differs from $\mathbf{d}$ in $N_e(\mathcal{E})$ symbols. Define the error signal by $e^{(\ell)}(k) = |d^{(\ell)}(k) - d^{(0)}(k)|/2$, which is $e^{(\ell)}(k) = 1$ for an error and $e^{(\ell)}(k) = 0$ for a correct symbol. Then an *error event* $\mathcal{E}$ is said to extend from time k_1 to k_2 if $e^{(\ell)}(k)$ is equal to the correct sequence outside the interval $k = \{k_1, \cdots, k_2\}$ and $e^{(\ell)}(k) = 1$ for $k = k_1, k_2$. With this definition the length of the error event becomes $L_b = k_2 - k_1$, where $L_b \geq N_e(\mathcal{E})$, with equality if all symbols within the burst are detected as errors.

Suppose the receiver has to decide between two hypothesis $\mathbf{d}^{(\ell)}$ and $\mathbf{d}^{(0)}$. Then the ML decision rule corresponds to the *likelihood ratio test*, $\text{LRT} = p(\mathbf{y}|\,\mathbf{d}^{(\ell)})/p(\mathbf{y}|\,\mathbf{d}^{(0)})$, as described by Van Trees [17]. The decision is made by choosing $\widehat{\mathbf{d}} = \mathbf{d}^{(0)}$ if $\text{LRT} < 0$, and $\mathbf{d}^{(\ell)}$ otherwise. By taking the logarithm and inverting the sign a log-likelihood function is obtained, and the LRT is given by the difference of the decision variables

$$\Delta\Lambda(\mathcal{E}) \;\triangleq\; \Lambda^{(\ell)} - \Lambda^{(0)} \tag{5.12}$$

and choosing $\mathbf{d}^{(0)}$ if $\Delta\Lambda$ is positive and $\mathbf{d}^{(\ell)}$ otherwise. According to the assumption that $\Lambda^{(0)}$ corresponds to the true sequence, a negative $\Delta\Lambda(\mathcal{E})$ equals a decision error. In case that the receiver has to choose only between two sequences $\mathbf{d}^{(0)}$ and $\mathbf{d}^{(\ell)}$ to make the decision $\widehat{\mathbf{d}}$, the pairwise probability of that error event becomes

$$P_e(\mathcal{E}) \;=\; P\{\Delta\Lambda(\mathcal{E}) < 0\} \;=\; \int_{-\infty}^{0} p\big(\Delta\Lambda(\mathcal{E})\big)\; d(\Delta\Lambda(\mathcal{E})) \tag{5.13}$$

The decision variable $\Delta\Lambda(\mathcal{E})$ can be expressed in terms of a Gaussian quadratic form, defined by

$$\Delta\Lambda(\mathcal{E}) = \mathbf{y}^H \left(\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1} - \mathbf{\Phi}_{\mathbf{yy}}^{(0)\,-1} \right) \mathbf{y} = \mathbf{y}^H \mathbf{Q}(\mathcal{E})\, \mathbf{y} \tag{5.14}$$

where $\mathbf{Q}(\mathcal{E}) = \left(\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)\,-1} - \mathbf{\Phi}_{\mathbf{yy}}^{(0)\,-1} \right)$ denotes the filter matrix. Note one decision error has an impact on the estimation/detection of the whole sequence. Since the covariance matrix $\mathbf{\Phi}_{\mathbf{yy}}^{(\ell)}$ is Hermitian, $\mathbf{Q}(\mathcal{E})$ is also Hermitian with dimension $N \times N$ and $N = KQ$.

In [81],[7, Appendix C], $P_e(\mathcal{E})$ was evaluated for pilot signal based channel estimation, by solving the integral in (5.13). Alternatively the bit error probability can be determined by using its characteristic function. The characteristic function of a complex normal multivariate pdf was derived by Turin [82]. Several solutions for the bit error rate are given in literature. Turin [79] derived the error probability for a binary hypothesis check in diversity reception. A general error formula for the binary error rate for a random signal disturbed by Gaussian noise was given by Barrett [60]. The analysis is based on extracting the eigenvalues of the quadratic form $\mathbf{y}^H \mathbf{Q}(\mathcal{E})\, \mathbf{y}$. Mämmelä [75] analysed the performance of the optimum receiver using Barrett's formula [60]. An analysis of the impact of estimation errors for a decision-feedback equaliser was presented by Stojanovic *et al* [83], which also was obtained by extracting the eigenvalues of a quadratic form. Barrett's formula can also be deployed for the performance analysis of a coherent DPSK receiver, as described by Dam [76]. Sub-optimum realisations based on Barrett's formula were developed by Mämmelä and Kaasila [68, 74, 75].
In [76] it is shown that a statistically equivalent quadratic form to (5.14) is given by

$$\Delta\Lambda(\mathcal{E}) = \mathbf{v}^H \mathbf{\Sigma}(\mathcal{E})\, \mathbf{v} \tag{5.15}$$

with $\mathbf{\Sigma}(\mathcal{E}) = \mathrm{diag}[\lambda_1, \cdots, \lambda_N]$ being a diagonal matrix of the eigenvalues of $\mathbf{\Phi}_{\mathbf{yy}}^{(0)} \mathbf{Q}(\mathcal{E})$ consisting of positive and negative real numbers. The components of $\mathbf{v}$ are independent unit-variance complex Gaussian random variables. The decision variable is thus reduced to a weighted sum of independent χ^2–distributed [7] random variables [7].

This receiver for binary signalling was analysed by Barrett [60] who derived an algebraic expression for the receiver's error performance through the use of residues. The pairwise probability of error which is obtained by evaluating (5.13), is given by[2] [60]

$$P_e(\mathcal{E}) = \sum_{\substack{n=1 \\ \lambda_n < 0}}^{N} \prod_{\substack{\nu=1 \\ \nu \neq n}}^{N} \frac{1}{1 - \lambda_\nu / \lambda_n} \tag{5.16}$$

where the set $\{\lambda_n,\ n = 1, \cdots, N\}$ are the eigenvalues of the matrix $\mathbf{\Phi}_{\mathbf{yy}}^{(0)} \mathbf{Q}(\mathcal{E})$. The matrix $\mathbf{\Phi}_{\mathbf{yy}}^{(0)} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H \middle| \mathbf{d}^{(0)} \right]$ is the covariance matrix of $\mathbf{y}$, given that hypotheses $\ell = 0$ was transmitted. It is assumed that the sequences are equally likely, so the error probability is the same for all hypotheses. The dimension of the matrices is $N \times N$, with $N = QK$. It should also be noted, that $\mathbf{\Phi}_{\mathbf{yy}}^{(0)} \mathbf{Q}(\mathcal{E})$ is in general *not* a Hermitian matrix. The eigenvalue decomposition of a non-Hermitian matrix is described in

[2] The eigenvalues $\{\lambda_n\}$ need to be mutually different.

Appendix A.2, using an algorithm given in [69].

In order to evaluate the probability of an error event over all transmitted sequences, the minimum distance needs to be found with respect to all possible sequences, not only the transmitted one as in (5.13). This results in the joint probability

$$P'_e(\mathcal{E}) = P\left\{ \bigcap_{\substack{i \in \Sigma\mathcal{E} \\ i \notin \mathcal{E}}} \hat{\Lambda} - \Lambda(i) < 0 \right\} \tag{5.17}$$

This means for the calculation of $P_e(\mathcal{E})$, the decision variable needs not only to be larger than the actual transmitted sequence $\mathbf{d}^{(0)}$, but also larger than the decision variable for all other hypotheses. After averaging over all possible error events, $\Sigma\mathcal{E}$, the average error probability is obtained

$$\bar{P}_e = \sum_{\mathcal{E} \in \Sigma\mathcal{E}} P'_e(\mathcal{E})\, P(\mathbf{d}^{(i)})\, N_e(\mathcal{E})$$

where $P(\mathbf{d}^{(i)})$ is the *a priori* probability that $\hat{\mathbf{d}}$ was transmitted and $N_e(\mathcal{E})$ are the number of bit errors associated with the error event $\mathcal{E}$. The Evaluation of the exact error probability is very complicated, however the MLSD receiver can be upper and lower bounded [84]. For a channel with ISI and known channel response these bounds are tight [85]. For the Rayleigh fading channel however, the upper bound is very loose [86, 87].

In the literature the average probability of a bit error can be lower bounded by only taking the most significant error event into account, that is $\bar{P}_e \geq P(\mathcal{E}_{max})$. This is commonly a single error, as its Euclidean distance is most likely the closest to the actual transmitted sequence. Let $\hat{\mathbf{d}}$ denote the sequence which differs from $\mathbf{d}$ in only one bit, then the lower bound becomes $\bar{P}_e \geq P_e(1)$.

An upper bound is obtained by an union bound, given by adding up the error probabilities of all possible error events

$$\bar{P}_e \leq \sum_{\mathcal{E} \in \Sigma\mathcal{E}} P_e(\mathcal{E})\, P(\mathbf{d}^{(i)})\, N_e(\mathcal{E}) \tag{5.18}$$

where $P_e(\mathcal{E})$ is the pair-wise error probability of the error event $\mathcal{E}$, given by (5.16). Strictly speaking, the set of all error events $\Sigma\mathcal{E}$ is unbounded for continuous transmission, but it can be truncated by a finite number, since the probability for an error burst with infinite length is zero. Even for a truncated upper bound there may be still too many sequences which must be calculated for (5.18). A further simplification is to take only the most significant error event $\mathcal{E}_{max}$ for a certain length L_e into account. Thus, only one pairwise error probability $P_e(\mathcal{E}_{max})$ is evaluated for an error burst of length L_b, out of 2^{L_b} possible error events. The dominant error event $\mathcal{E}_{max}$ is deemed the error sequence with the maximum number of errors, which is L_b, the length of the error burst. Hence, $N_e(\mathcal{E}_{max}) = L_b$ and $P(\mathbf{d}^{(i)}) = 2^{-L_b}$ assuming all sequences are equally likely, which yields the following approximation of the upper bound

$$\bar{P}_e \leq \sum_{\mathcal{E}_{max} \in \Sigma\mathcal{E}} P_e(\mathcal{E}_{max})\, 2^{-L_b}\, L_b \tag{5.19}$$

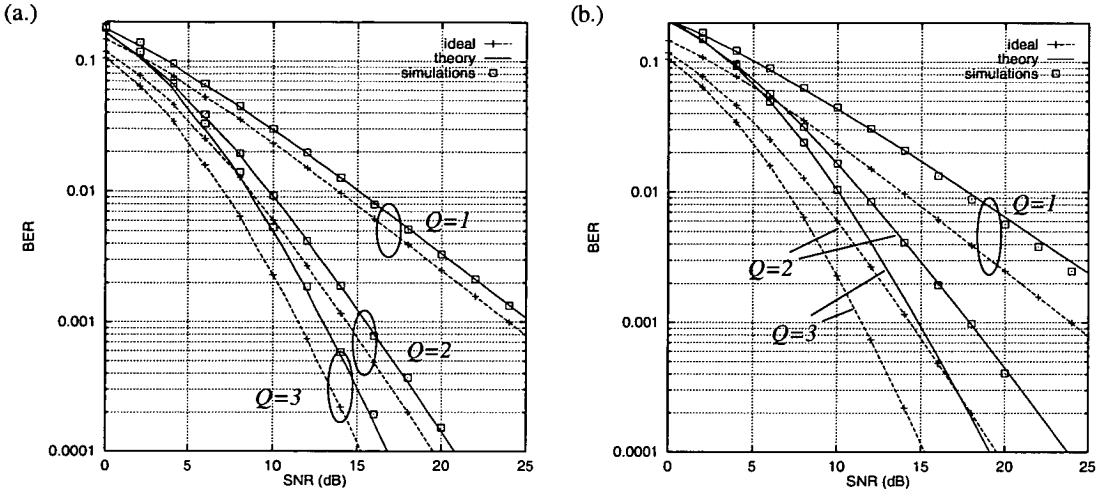


Figure 5.2: *BER vs SNR of the optimal receiver for different numbers of diversity taps Q, (a.) smoother and (b.) linear predictor; $K = 8$, $\nu'_{max} = 0.05$.*

It turns out, however, that this upper bound may still be loose for low SNR because the error events are not disjoint [86]. An example for the upper bound is given in section 5.2.4.

Numerical results This section concentrates on the evaluation of the lower bound, which corresponds to single error events with $L_b = 1$. The resulting log-likelihood test is the difference of the decision variables $\Delta\Lambda = \Lambda(1) - \Lambda(0)$. For continuous transmission ($K \to \infty$), the performance of (5.16) approaches the performance of a RAKE receiver with *a priori* knowledge of the CIR. Clearly for long sequences the computation of (5.14) is not feasible. Some results of the optimum receiver performance for a limited observation interval K, are presented. For a more thorough analysis of the optimum receiver performance the interested reader is referred to [75]. The sequence length K may be regarded as a short transmission burst, e.g. in a TDMA time slot, or a sliding window of the subsequence, in which case the receiver is of course not optimum.

The assumption that only single errors are present has the interpretation that *no* decision feedback effects are present. This means that channel estimation becomes independent of the data hypotheses and detection can be performed on a symbol–by–symbol basis, as discussed in section 5.1.4. Such an idealised receiver can be implemented, by using the actual transmitted sequence to generate $\mathbf{y}'_q = \mathbf{D}^H \mathbf{y}_q$, independent of ℓ. Monte Carlo simulations can be used to verify the lower bound $P_e(1)$ from (5.16). The lower bound is expected to match the true error probability for high SNR.

Results are presented for detection of the bit in the middle and on either end of the sequence, respectively. The smoother estimates the bit in the middle of the test sequence $k_1 = k - K/2$, while the linear predictor estimates the first or last bit of $\mathbf{y}_q$, i.e. $k_1 = k$ or $k_1 = k - K$. The simulation results shown in Figure 5.2 match the theoretically predicted results very closely. Since the linear predictor in Figure 5.2.b estimates the unknown bit at the tail of the sequence, the performance is significantly poorer.

The dependency of the error probability on the position of the unknown bit in the sequence $\mathbf{y}_q$ is shown in Figure 5.3. Unless the unknown bit is near the tails of the sequence, its influence is negligible.

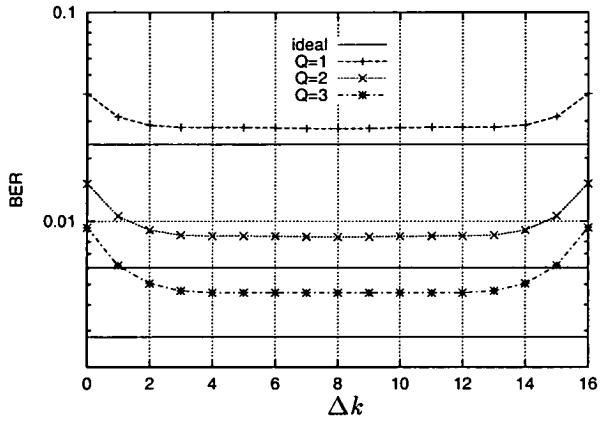


Figure 5.3: *BER dependent on the position Δk of the unknown bit in the sequence, of the optimal receiver for different numbers of diversity taps Q; $K=16$, $\nu'_{max}=0.05$.*

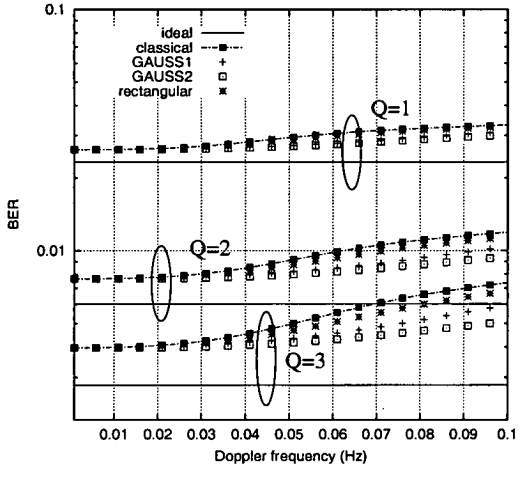


Figure 5.4: *BER vs normalised Doppler frequency ν'_{max} of the smoother for different channel models; $\bar{\gamma}=10\,dB$, $K=8$.*

However, the dependence of Δk on the BER increases with the number of diversity taps Q.

The effect of the choice of the channel model on the performance of the smoother is shown in Figure 5.4. The distributions of the Doppler spread for the channel models are described in section 3.2.5. The difference between the channel models becomes more significant as ν'_{max} and the diversity Q increases. It is seen that the classical Doppler model can be used as a worst case for the system performance. In the remainder of this thesis only the classical Doppler model will be considered further on.

5.2 Recursive MLSD receiver

The receiver structures (5.10) and (5.5) require future symbols to decide upon the k^{th} received signal sample. We now look at receiver structures employing present and past samples for estimation and detection only, as studied in [35–37,88]. The first recursive MLSD receiver structure, applicable to slow

and fast fading channels was presented by Morley and Snyder [88].[3] The receiver consisted of a bank of parallel time-continuous linear filters. These filters were obtained by evaluating integral equations. Häb and Meyr [35] along with Lodge and Moher [36] derived a recursive formulation of the MLSD, independently of each other, for the flat fading channel using a discrete-time signal system. Compared to [88] it has the usual advantages of digital signal processing as compared to analog signal processing. Later Yu and Pasupathy [37] derived an equivalent MLSD receiver by applying the innovations approach. They also generalised the results of [35,36] to the inter-symbol interference channel. Another interpretation of the recursive MLSD was presented by Makrakis *et al.* [90]. There the branch metric computation for phase modulated signals is performed through *multiple differential detection.* A comparison of this technique with the receiver in [36] was given in [91], showing the similarities of the two approaches. The recursive MLSD decision rule provides a powerful tool to develop close to optimum or sub-optimum but realisable receiver structures. This has been made possible by applying the Viterbi algorithm (VA) to the recursively updated decision variable, first developed in [88] and further investigated in [36,37]. Later the approach of employing the VA to the problem of MLSD with unknown parameter estimation has been coined the principle of *per-survivor processing* (PSP) [92]. In that paper no claim concerning the optimality of the algorithm was made. The theoretical foundation of PSP was provided by Chugg [89,93], including the development of the receiver front-end providing sufficient statistics for the discrete-time received signal [93]. PSP is a general approach to joint detection and estimation by using a different channel estimation filter for every decoder state in the trellis. PSP furthermore employs the VA to search the trellis, as an in general sub-optimum decoding algorithm. In other words, it involves processing a separate channel estimate for each survivor in the trellis.
After deriving the recursive receiver structure, the application of the VA to the problem according to the PSP principle, will be discussed in section 5.2.3.

Let $\mathbf{d}^{(\ell)}(k) = [d^{(\ell)}(1), \cdots, d^{(\ell)}(k)]^T$ and $\mathbf{y}(k) = [\mathbf{y}_1^T(k), \cdots, \mathbf{y}_Q^T(k)]^T \in \mathbb{C}^{kQ}$ denote the data and received sequences of hypotheses ℓ, up to the k^{th} sample, respectively, the likelihood function from (5.7) reads

$$p\big(\mathbf{y}(k) \mid \mathbf{d}^{(\ell)}(k)\big) = \prod_{q=1}^{Q} p\big(\mathbf{y}_q(k) \mid \mathbf{d}^{(\ell)}(k)\big) \tag{5.20}$$

The set $\mathcal{A}_K$ contains A_m^K possible sequences, and the sequence which maximises $p\big(\mathbf{y}(k) \mid \mathbf{d}^{(\ell)}(k)\big)$ is the most likely transmitted sequence $\widehat{\mathbf{d}}$. Let $\boldsymbol{\xi}(\ell,k) \triangleq \{(y_q(1), \cdots, y_q(k)), \mathbf{d}^{(\ell)}\}$ represent the observation up to time k and hypotheses ℓ, and $\mathbf{y}_q(k) = [\mathbf{y}_q(k-1), y_q(k)]^T$. Applying the definition of a conditional pdf, $p(A|B)p(B) = p(A,B)$, to $p\big(\mathbf{y}_q(k) \mid \mathbf{d}^{(\ell)}(k)\big)$, the following is obtained [36]

$$p\big(\mathbf{y}_q(k) \mid \mathbf{d}^{(\ell)}(k)\big) \;=\; p\big(y_q(k) \mid \boldsymbol{\xi}(\ell,k)\big)\; p\big(\mathbf{y}_q(k-1) \mid \mathbf{d}^{(\ell)}(k)\big) \tag{5.21}$$

The term $p\big(y_q(k) \mid \boldsymbol{\xi}(\ell,k)\big)$ is the pdf pertaining to the *one-step prediction* of the received sample $y_q(k)$, given the past received signal vector $\mathbf{y}_q(k-1)$ and the data hypothesis $\mathbf{d}^{(\ell)}(k)$ up to the present.

[3] In [88] were some misleading claims about the optimality of the algorithm, these claims were corrected by Chugg [89].

The observation of the q^{th} tap, $\mathbf{y}_q(k)$, is a complex Gaussian i.i.d. random variable, hence the pdf $p(y_q(k) \mid \boldsymbol{\xi}(\ell, k-1))$ is Gaussian distributed with conditional mean $\widehat{y}_q(\ell, k)$ and variance $\sigma^2_{y,q}(k)$, given by

$$p(y_q(k) \mid \boldsymbol{\xi}(\ell, k)) = \frac{1}{\pi\sigma_{y,q}(k)} \exp\left(-\frac{\left|y_q(k) - \widehat{y}_q(\ell, k)\right|^2}{\sigma^2_{y,q}(k)}\right) \tag{5.22}$$

where $\sigma^2_{y,q}(k) = \mathrm{E}\left[\left|y_q(k) - \widehat{y}_q(\ell, k)\right|^2\right]$ denotes the variance between the received signal $y_q(k)$ and its estimate $\widehat{y}_q(\ell, k)$. With the presumption that the sequence $\mathbf{d}^{(\ell)}(k)$ has been transmitted, the mean $\widehat{y}_q(\ell, k)$ can be expressed as

$$\begin{aligned} \widehat{y}_q(\ell, k) &= \mathrm{E}\left[y_q(k) \mid \boldsymbol{\xi}(\ell, k-1)\right] \\ &= d^{(\ell)}(k) \underbrace{\mathrm{E}\left[h_q(k) \mid \boldsymbol{\xi}(\ell, k-1)\right]}_{\widehat{h}_q(\ell,k)} + \underbrace{\mathrm{E}[n(k)]}_{=0} \end{aligned}$$

where $\widehat{h}_q(\ell, k) = \mathrm{E}\left[h_q(k) \mid \boldsymbol{\xi}(\ell, k-1)\right]$ is the optimal linear prediction estimate of $h_q(k)$.

Recursive metric computation Following the discussion for diversity reception in 5.1.3 the pdf of the entire observation $\mathbf{y}(k)$ can be obtained by substituting (5.22) into (5.20). Taking the logarithm and inverting the sign, $\Lambda(\ell, k) = -\ln\left[p(\mathbf{y}(k) \mid \mathbf{d}^{(\ell)}(k))\right]$, the recursion for the decision metric is obtained

$$\Lambda(\ell, k) = \Lambda(\ell, k-1) + \sum_{q=1}^{Q} \frac{\left|y_q(k) - d^{(\ell)}(k)\,\widehat{h}_q(\ell, k)\right|^2}{\sigma^2_{y,q}} + \ln\left[\pi\sigma_{y,q}\right]$$

Terms which are independent of the hypotheses ℓ can again be neglected, so the above simplifies to

$$\Lambda(\ell, k) = \Lambda(\ell, k-1) + \sum_{q=1}^{Q} \underbrace{\left|y_q(k) - d^{(\ell)}(k)\,\widehat{h}_q(\ell, k)\right|^2}_{\Delta(\ell,k)} \tag{5.23}$$

where $\Delta(\ell, k)$ denotes the Euclidean distance between the received signal and the CIR estimate, conditioned on hypothesis ℓ. Repeated application of (5.23) starting from the end of the sequence $k = K$ yields the decision variable, which is to be minimised

$$\widehat{\Lambda} = \min_{\ell \in \mathcal{A}_K} \Lambda(\ell, K) = \min_{\ell \in \mathcal{A}_K} \sum_{k=1}^{K} \sum_{q=1}^{Q} \left|y_q(k) - d^{(\ell)}(k)\,\widehat{h}_q(\ell, k)\right|^2 \tag{5.24}$$

where the factors which are constant and independent of ℓ have been removed. The optimum receiver performs a tree search with respect to k, with a complexity growing exponentially with the sequence length K. The sequence $\{d^{(\ell)}(k)\,\widehat{h}_q(\ell, k)\}$ can be interpreted as the "model output" of the receiver, and it is to be compared with the observation $\{y_q(k)\}$ for all hypotheses ℓ. According to (5.24), the model output with the minimum Euclidean distance is the most likely transmitted sequence $\widehat{\mathbf{d}}$. A block diagram of the receiver structure with $A_m L = A_m^k$ parallel branches is shown in Figure 5.5.

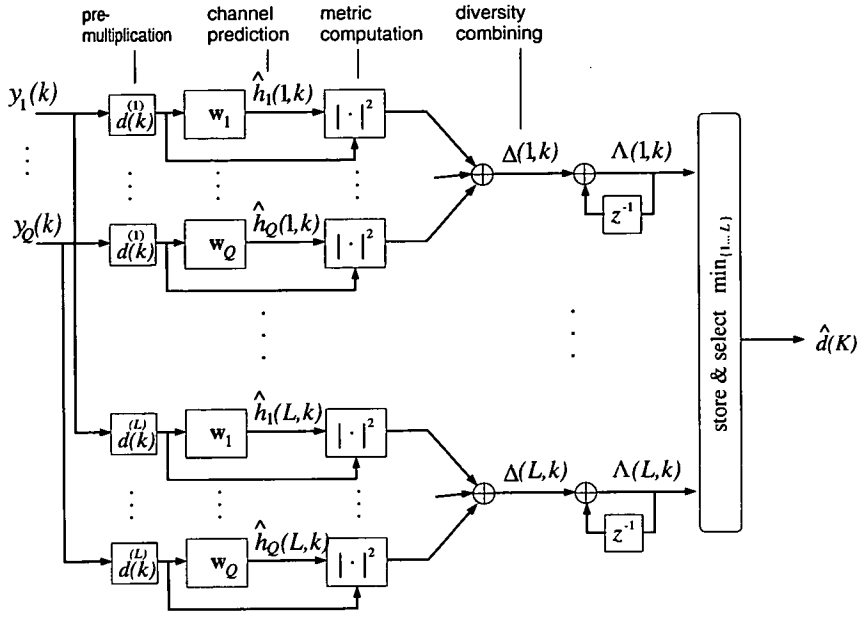


Figure 5.5: *Block diagram of the recursive MLSD receiver.*

Note the linear predictor estimate $\widehat{h}_q(\ell, k)$ of (5.24) is clearly suboptimal with respect to the smoothed estimates of (5.5). That is because $\widehat{h}_q(\ell, k)$ from the pdf $p\big(y_q(k) \mid \mathbf{y}_q(k-1),\, \mathbf{d}^{(\ell)}(k)\big)$ is conditioned on the past received signals and the hypotheses $\mathbf{d}^{(\ell)}(k)$ up to the present only; unlike to the smoother in (5.5), which utilises the whole sequence to form an estimate. Nevertheless, the metric $\Lambda(\ell, K)$ at the end of the message remains optimal for detection. Therefore, the MSE as a performance measure for the optimum receiver is inadequate, since even though the two detectors optimise the performance in the sense that they minimise the probability of error, their corresponding estimators do have a different MSE.

Innovation based MLSD An equivalent derivation for the recursive MLSD is given by the innovations approach [37]. The basic idea is to transform the decision variable for the optimum receiver $\mathbf{y}^H {\boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)}}^{-1} \mathbf{y}$ from (5.4) to a quadratic form such that its entries are mutually statistically independent. Recall from section 2.1.2 that the innovations approach provides exactly that, since it "whitens" $\mathbf{y}$ such that its entries are mutually uncorrelated. The innovations process of $\mathbf{y}$ is denoted by $\boldsymbol{\epsilon}^{(\ell)}$ which is a white noise process, given by the linear transformation $\boldsymbol{\epsilon}^{(\ell)} = {\mathbf{L}^{(\ell)}}^{-1}\mathbf{y}$. The transformation matrix $\mathbf{L}^{(\ell)}$ is a lower triangular matrix which is given through factorisation of $\boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)} = \mathbf{L}^{(\ell)}\boldsymbol{\Sigma}{\mathbf{L}^{(\ell)}}^H$. The rows of ${\mathbf{L}^{(\ell)}}^{-1}$ are the coefficients of a 1-step predictor for orders 0 through $K-1$ and the elements of $\boldsymbol{\Sigma}$ are the corresponding error covariances. With this factorisation the decision variable becomes

$$\begin{aligned} \Lambda(\ell, K) = \mathbf{y}^H {\boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)}}^{-1} \mathbf{y} &= {\boldsymbol{\epsilon}^{(\ell)}}^H \boldsymbol{\Sigma}^{-1}\, \boldsymbol{\epsilon}^{(\ell)} \\ &= \sum_{k=1}^{K}\sum_{q=1}^{Q} \frac{|\epsilon_q(\ell, k)|^2}{\sigma_{y,q}^2(k)} \end{aligned} \tag{5.25}$$

Note that $\epsilon_q(\ell, k) = y_q(k) - d^{(\ell)}(k)\,\widehat{h}_q(\ell, k)$ denotes the estimation error of a one-step linear predictor. Furthermore, the error covariance $\sigma_{y,q}^2(k)$ from (5.22) is independent of the data and may therefore be

neglected. Thus, the equality of the innovations based approach to (5.23) follows readily. The equivalence of the innovations based MLSD in (5.25) and the linear predictive MLSD was pointed out in [94]. An extension of the innovations based MLSD [37] to a receiver employing Kalman filters was presented in [95].

5.2.1 Channel estimation

The one-step prediction estimate of the CIR for a wide-sense stationary channel is given by a Wiener filter. For the more general case of a non-stationary channel, optimal one-step prediction is performed by a Kalman filter. The application of a Kalman filter to MLSD was studied for flat fading in [35] and for the frequency selective ISI channel in [96, 97]. The Wiener filter for one-step prediction can be expressed in terms of the auto and cross-correlation matrices

$$\widehat{h}_q(\ell, k+1) = \underbrace{\mathrm{E}\big[\, h_q(k+1)\, \mathbf{y}_q^H(k) \;\big|\; \boldsymbol{\xi}(\ell,k)\,\big] \cdot \mathrm{E}\big[\, \mathbf{y}_q(k)\mathbf{y}_q^H(k) \;\big|\; \boldsymbol{\xi}(\ell,k)\,\big]^{-1}}_{\text{Wiener filter}} \cdot \underbrace{\mathbf{y}_q(k)}_{\text{signal}}$$

where $\mathrm{E}\big[\, h_q(k+1)\, \mathbf{y}_q^H(k) \;\big|\; \boldsymbol{\xi}(\ell,k)\,\big] = \mathbf{D}^{(\alpha)}(k+1)\, \mathrm{E}\Big[\, h_q(k+1)\, \mathbf{y}_q'^H(\ell,k)\,\Big]$ denotes the cross correlation vector between the CIR, $h_q(k+1)$, and the observation $\mathbf{y}_q(k)$, conditioned on hypothesis ℓ. The term $\mathrm{E}\big[\, \mathbf{y}_q(k)\mathbf{y}_q^H(k) \;\big|\; \boldsymbol{\xi}(\ell,k)\,\big]$ denotes the auto-correlation matrix of the observation up to the k^{th} sample, $\mathbf{y}_q(k)$. According to section 5.1.3, the pre-multiplied received signal $\mathbf{y}_q'(\ell,k)$ may again be applied to obtain the estimate

$$\begin{aligned} \widehat{h}_q(\ell, k+1) &= \underbrace{\mathrm{E}\Big[\, h_q(k+1)\, \mathbf{y}_q'^H(\ell,k)\,\Big] \cdot \mathrm{E}\Big[\, \mathbf{y}_q'(\ell,k)\, \mathbf{y}_q'^H(\ell,k)\,\Big]^{-1}}_{\mathbf{w}^{(k)H}} \mathbf{y}_q'(\ell,k) \qquad (5.26) \\ &= \mathbf{w}^{(k)H}\, \mathbf{y}_q'(\ell,k) = \sum_{m=1}^{k-1} w_m^{(k)*}\, y_q'(\ell, k-m+1) \end{aligned}$$

which follows from (A.1) in Appendix A.1. The filter $\mathbf{w}^{(k)}$ is a k^{th} order one-step linear prediction filter, with coefficients $\{w_m^{(k)};\ m = 1, \cdots k\}$. The filter $\mathbf{w}^{(k)}$ is obtained from solving the Wiener-Hopf equation. The channel prediction from the pre-multiplied observation $\mathbf{y}_q'(\ell,k)$ have become entirely independent of the data $\mathbf{d}^{(\alpha)}(k)$.

FIR channel estimation The filter $\mathbf{w}_q^{(k)}$ can be pre-computed and used for all ℓ hypotheses, however there is a separate filter needed for every sample. In analogy to section 2.1.3, $\mathbf{w}_q^{(k)}$ can be truncated by a time independent moving average (MA) predictor, $\mathbf{w}_q^{(M)}$, of order M. For the sake of simplicity the superscript $^{(M)}$, to indicate the order of the filter will be dropped in the following. In analogy to (2.17) the prediction of the future sample $k+1$ then becomes

$$\widehat{h}_q(\ell, k+1) = \sum_{m=1}^{M} w_{qm}^*\, y_q'(\ell, k-m+1) = \mathbf{w}_q^H \mathbf{y}_q'(\ell,k) \qquad (5.27)$$

where the vector $\mathbf{y}'_q(\ell,k)$ is assumed to have the appropriate dimension $\mathbb{C}^M$. Note, the second order statistics, such as the cross-correlation or the auto-correlation matrix become also time independent, thus the index k will be dropped in the following. According to section 2.1.3 a notation for the Wiener-Hopf equation from (2.18) is obtained by defining: $\mathbf{\Phi}_{\mathbf{y'y'},q} = \mathrm{E}\big[\,\mathbf{y}'_q(\ell,k)\,\mathbf{y}'^H_q(\ell,k)\,\big]$ with dimension $M \times M$ for the auto-correlation, and $\boldsymbol{\phi}_{h\mathbf{y}',q} = \mathrm{E}\big[\,h_q(k+1)\,\mathbf{y}'^H_q(\ell,k)\,\big]$ with dimension M for the cross-correlation function; the MMSE filter can be determined

$$\mathbf{w}_q \;=\; \mathbf{\Phi}^{-1}_{\mathbf{y'y}',q}\cdot\boldsymbol{\phi}_{h\mathbf{y}',q}\;;\quad \in\mathbb{C}^M \tag{5.28}$$

MLSD using linear predictive FIR filtering was studied for a flat fading channel in [36,94,98] and was extended to ISI channels in [37]. Clearly, this requires knowledge of the second order statistics of the CIR. Ways to estimate the predictor $\mathbf{w}_q$ are described in section 2.1.4. Another, sub-optimal approach is to choose $\mathbf{w}_q$ such that it does not require any knowledge about the channel statistics. The predictor $\mathbf{w}_q$ is chosen *a priori* and does therefore not need to estimate the 2nd order statistics of CIR. This may be achieved by modelling the fading process as a polynomial in time [99].

Next the variance σ_y^2 is determined and related to the MMSE, V_{min}^2. The variance of the q^{th} diversity tap from (5.22) is: $\sigma_{yq}^2 = \mathrm{E}\big[\,\big|y'_q(\ell,k)-\widehat{h}_q(\ell,k)\big|^2\,\big]$. Furthermore, $\sigma_y^2 = \sigma_{y'}^2 = \sum_q \sigma_{y,q}^2$ is equivalent to the MMSE plus the variance of AWGN. Therefore, (2.19) can directly be applied for one-step prediction of diversity tap q. For the variance of the Q^{th} order diversity system, the following is obtained

$$\sigma_y^2 = V_{min}^2 + N_0 \;=\; 1 + N_0 \;-\; \sum_{q=1}^{Q} \boldsymbol{\phi}^H_{h\mathbf{y}',q}\,\mathbf{\Phi}^{-1}_{\mathbf{y'y}',q}\,\boldsymbol{\phi}^H_{h\mathbf{y}',q} \tag{5.29}$$

For constant envelope signals $\boldsymbol{\phi}_{h\mathbf{y}',q}$ can directly be replaced with $\boldsymbol{\phi}_{y'\mathbf{y}',q}$ in (5.29), since $\boldsymbol{\phi}_{y'\mathbf{y}',q} = d^{(*)}(k)\,\boldsymbol{\phi}_{h\mathbf{y}',q}$. Note that the error variance σ_y^2 is independent of ℓ.

IIR estimation filter Another possibility of performing an one-step prediction is to use an infinite impulse response (IIR) filter instead of a FIR filter, discussed in 2.1.5. According to (2.21) ,the recursive channel estimator for hypothesis ℓ can be expressed as

$$\widehat{h}_q(\ell,k+1) \;=\; (1-\alpha)\,y'_q(\ell,k) + \alpha\,\widehat{h}_q(\ell,k)\;;\qquad 0\le\alpha<1\,. \tag{5.30}$$

The filter has the form of a low-pass filter. Thus, it reduces the effects of noise at the expense of some imposed pass-band distortions. The channel estimate needs to be generated for each tap separately, and is subsequently processed in the same way as the FIR filter estimate to obtain the decision variable $\Lambda(\ell,k)$ in (5.23). This type of channel estimation may be viewed as a sub-optimum approximation of a Kalman filter, as suggested by Iltis *et al* [97]. 1st order IIR filtering, also referred to as LMS filtering applied to MLSD, has been investigated further for the ISI channel in [93,100] and for flat fading in [30,101].

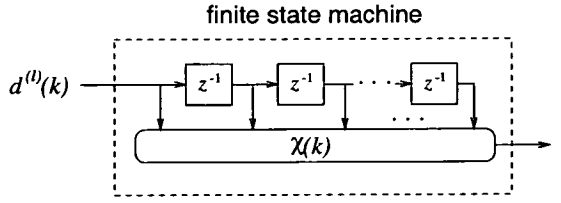


Figure 5.6: *Shift register model of the transmitted sequence.*

5.2.2 Finite state channel model

Suppose the CIR estimated is approximated by a M^{th} order moving average (MA) filter. With this approximation the receiver can be realised by a trellis search. This sub-optimum approximation offers a solution for (5.23) with a complexity independent of the sequence length K but exponentially dependent on the predictor order M. Thus the Euclidean distance in $\left|y_q(k) - d^{(i)}(k)\,\widehat{h}_q(\ell,k)\right|^2$ of (5.23) depends only on $M+1$ samples; the present one plus M past samples used for channel estimation. Assuming an auto-regressive (AR) model for the channel, as described in section 2.1.3, the CIR can be described by a finite state Markov process and is graphically represented by a trellis diagram. The following definitions are used to describe the trellis diagram:

State: A state at time k is defined by

$$\chi(k) \triangleq \left\{a^{(i)}(k), \cdots, a^{(i)}(k-M+1)\right\}\,; \quad i \in \mathcal{S} \tag{5.31}$$

where $a^{(i)}(\cdot) \in \mathcal{D}$ is the information symbol of state $\chi(k) = i$. There is a one-to-one correspondence between $a^{(i)}(k)$ and the transmitted signal hypothesis, that is for MPSK $d^{(i)}(k) = \exp\left(j2\pi\, a(k)^{(i)}/A_m\right)$. Assuming a A_m-ary symbol alphabet $\mathcal{D}$, the set of states is denoted by $\mathcal{S} = \{\chi(k) : i = 0, \cdots, A_m^M - 1\}$. There are $L = A_m^M$ states $\chi(k) \in \mathcal{S}$ per time instant, according to the memory of M time samples of one state.

Transition: A transition or **branch** between the states $\chi(k-1) = i$ and $\chi(k) = i'$ is defined by

$$\begin{aligned}\left(\chi(k-1), \chi(k)\right) = (i, i') &\triangleq \left\{a^{(i)}(k), \cdots, a^{(i)}(k-M)\right\} \\ &= \left\{a^{(i)}(k),\, \chi(k)\right\}\end{aligned} \tag{5.32}$$

There are A_m transitions per state, which sums up to $A_m L = A_m^{M+1}$ transitions per sample in total. There is obviously a one-to-one correspondence between state sequences $\{\chi(k)\}$ and transition sequences, given by a *finite state machine* or a *shift register process* [84], since it can be modelled by a shift register of length M, depicted in Figure 5.6. By observing the trellis, the analogy of the trellis structure to equalisation of channels with ISI [85] becomes apparent, however, the branch metrics are defined differently. The metric of the branch is the Euclidean distance between the estimated and received signal, given by

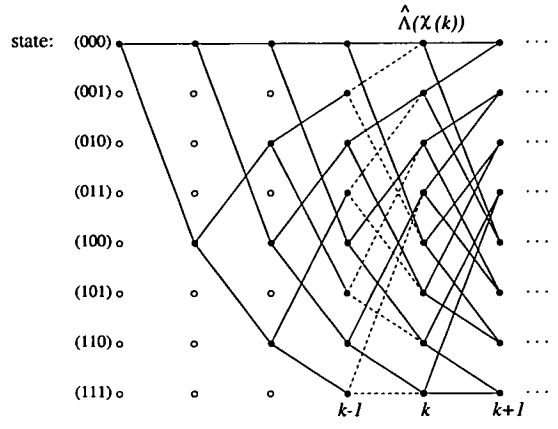


Figure 5.7: *Distance metric illustrated in a trellis for the recursive MLSD, $M=2$.*

$$\begin{aligned}\Delta(i,i') &= \sum_{q=1}^{Q} \left| y'_q(i',k+1) - \widehat{h}_q(i,k+1) \right|^2 \\ &= \sum_{q=1}^{Q} \left| \sum_{m=0}^{M} w^*_{qm}\, y'_q(j,k-m+1) \right|^2 \quad ; \qquad \begin{aligned} w_{q0} &\triangleq -1 \\ j &\in (i,i') \end{aligned}\end{aligned} \tag{5.33}$$

where $y'_q(j,k+1) = d^{(i')*}(k+1)\, y_q(k+1)$ denotes the pre-multiplied received signal of state $j = \{i,i'\}$. Associated to state $\chi(k-1)=i$ is the one step channel prediction $\widehat{h}_q(i,k+1)$. According to section 5.2.1 the CIR estimate is obtained by either an FIR or IIR estimation filter.

Path: A path at time k is defined by

$$\begin{aligned}\mathbf{a}^{(\imath)}(k) &\triangleq \{\chi(k),\cdots,\chi(1)\} \\ &= \{a^{(\ell)}(k),\cdots,a^{(\ell)}(2),\chi(1)\}\ ; \quad \ell \in \tilde{\mathcal{A}}_k\end{aligned} \tag{5.34}$$

It is assumed that transmission starts at time $k=1$ with a known initial state. With the defined transition metrics in (5.33), the metric of a path entering at state $\chi(k)$ is equivalent with the decision variable

$$\begin{aligned}\Lambda(\chi(k)) &= \Lambda(\chi(k-1)) + \Delta(\chi(k-1),\chi(k)) \\ &= \sum_{m=2}^{k} \Delta(\chi(k-m-1),\chi(k-m))\end{aligned} \tag{5.35}$$

which represents the decision variable in (5.23).

This results in a trellis structure, illustrated in Figure 5.7 for a finite state machine of order $M=2$.

5.2.3 Application of the Viterbi algorithm

It was stated earlier that number of paths $\mathbf{a}^{(\imath)}(k)$ grow exponentially with time k. The prohibitive high complexity of ML sequence detection, can be reduced by employing the Viterbi Algorithm (VA) [84], originally proposed for the decoding of convolutional codes by Viterbi [102]. The VA is an asymptotically

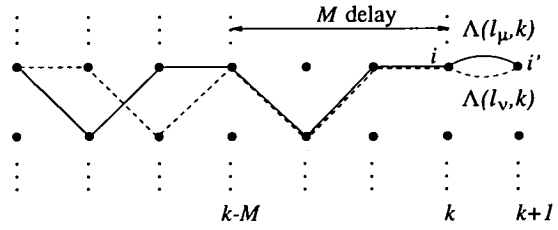


Figure 5.8: *Distance metric illustrated in a trellis for the recursive MLSD ($A_m = 2$).*

optimum decoding algorithm, based on Bellman's optimality principle of dynamic programming [103]. Later the algorithm was applied for optimum equalisation of channel with inter symbol interference (ISI) by Forney [85], multi user detection in multiple access channels by Verdú [104], speech processing, and many more applications. The VA offers an efficient solution where maximum likelihood detection of a whole sequence is required.

Optimality considerations Conditions for the applicability of the VA to MLSD with unknown parameter estimation were analysed by Chugg [89, 93]. It was argued that the VA is the optimal decoding algorithm if the transition metric is fully described by (5.33). Otherwise the VA is the optimum decoding algorithm for an incomplete and thus sub-optimum finite state representation of the received signal. Consequently, MLSD using the VA is a sub-optimum approximation to the optimal estimator-correlator receiver described in [31].

Generally speaking, the VA represents an optimum solution of MLSD for a finite state, discrete time Markov process observed in memoryless noise. The VA is only optimal as a decoding algorithm if the process is Markovian.[4] That is the probability, $P\big(\chi(k+1)|\chi(1),\cdots,\chi(k)\big)$, of being in state $\chi(k+1)$ depends only on the state $\chi(k)$ [84]:

$$P\big(\chi(k+1)\,|\,\chi(1),\cdots,\chi(k)\big) = P\big(\chi(k+1)\,|\,\chi(k)\big)\ .$$

To apply this criterion to our problem consider Figure 5.8. Suppose the sequences say $\ell = \ell_\nu$ and $\ell = \ell_\mu$ are the same for at least M samples and they are different for at least one symbol with delay larger than M. The VA is optimum if and only if $\Delta(i,i')$ from (5.33) is identical for ℓ_ν and ℓ_μ. If that is not the case and the VA was applied at $\chi(k) = i$ and either of the candidates was discarded, the corresponding survivor may not be the ML path in the end of the sequence and the VA would therefore be sub-optimum. Note, with a M^{th} order linear predictor this condition is met such that the VA is optimum. The receiver itself, however, may be sub-optimum because of the truncation of the predictor from (5.26) to (5.27).

In the following the application of the VA using the definitions in (5.31)–(5.35) will be discussed, according to the principle of per-survivor processing (PSP) [92]. This certain type of receiver will be referred to as VA-MLSD. Let $\widehat{\Lambda}(i,k)$ denote the survivor path of state $\chi(k) = i$. That is the metric with

[4] A 1^{st} order Markov process is described as Markovian.

the minimum distance entering this state, which is obtained by minimising (5.24), given by

$$\widehat{\Lambda}(i,k) = \min_{\ell \in i}\{\Lambda(\ell,k)\} \tag{5.36}$$

Associated with each survivor is the path

$$\tilde{\mathbf{a}}^{(i)}(k) \;=\; \left\{\tilde{a}^{(i)}(k), \cdots, \tilde{a}^{(i)}(2),\, \chi(1)\right\}\;; \quad i \in \tilde{\mathcal{S}} \tag{5.37}$$

where the past history $\{\tilde{a}^{(i)}(k)\}$ denotes the tentative decisions of state i. The Viterbi algorithm can now be described as follows:

Storage: Survivor terminating in $\chi(k)$: $\widehat{\Lambda}(i,k) \quad 0 \le i < L$
associated metric: $\tilde{\mathbf{a}}^{(i)}(k) \quad 0 \le i < L$

Initialisation:
$$\begin{aligned} k &= 1 \\ \chi(1) &= 0 \;\rightarrow\; \tilde{\mathbf{a}}^{(i)}(k) = 0 \\ \widehat{\Lambda}(i,1) &= \begin{cases} 0\,; & i=0 \\ \infty\,; & i\neq 0 \end{cases} \end{aligned}$$

Recursion: To extend the survivors of (5.36) to sample $k{+}1$, the metric from state $\chi(k)=i$ to $\chi(k{+}1)=i'$ is computed, giving the trellis update $\Lambda(i',k{+}1) = \widehat{\Lambda}(i,k) + \Delta(i,i')$.
The computation of $\Delta(i,i')$ involves the evaluation of the one-step channel prediction, $\widehat{h}(i,k{+}1)$, associated to state $\chi(k)=i$. From the branches entering at state i', the ones with the larger metric are discarded, giving the survivor at time $k{+}1$:

$$\widehat{\Lambda}(i',k{+}1) \;=\; \widehat{\Lambda}(i,k) + \min_{(i,i')} \Delta(i,i')\;. \tag{5.38}$$

This is illustrated in Figure 5.7, where the VA calls for choosing between two transitions, $\Delta(i,i')$, printed as dashed and solid lines.

Complexity issues: Given the trellis is in its equilibrium, i.e. every state has A_m entering and A_m leaving branches per time step for MPSK, one survivor out of all the branches entering a particular state is to be determined. Note, for every state A_m new transitions need to be calculated. The trellis has A_m^M number of states, with A_m branches entering and leaving each state per sample, resulting in A_m^{M+1} transitions in total for the whole trellis. For each transition, Q CIRs must be estimated (one per diversity tap). The computational cost of a transition $\big(\chi(k-1),\chi(k)\big)$ is mostly dependent on the choice of the estimation filter, its order M, and the number of diversity taps. The complexity in terms of the number of survivors which are to be processed, which is here equivalent to the number of states in the trellis, can be classified by the order $O(A_m^M)$. The complexity order $O(A_m^M)$ states that the computational cost of VA–MLSD grows exponentially with M.

Although the ML decision rule states that the final decision is taken at the end of the sequence, little degradation is expected if the definite decision on the most likely path is made after a delay of only a few samples.

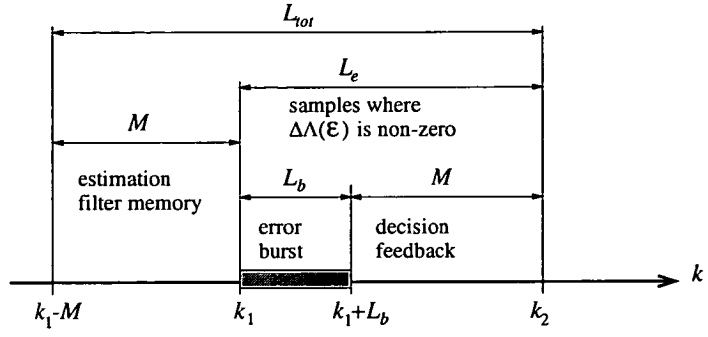


Figure 5.9: *Effect of an error burst to the differential decision variable* $\Delta\Lambda_q(\mathcal{E})$.

5.2.4 Performance analysis

The performance of a maximum likelihood sequence detector is analysed, following the steps in section 5.1.5, where the error probability of the non-causal receiver was evaluated. Similar expressions for the analysis of the recursive MLSD in a fading environment are discussed in [86, 87, 105]. An analysis of the performance for MLSD was given for a flat fading [86, 87] and for a frequency selective fading channel [86, 105]. For the recursive MLSD receiver, the derivation of a quadratic form is not as straight forward as in the previous case. Now the decision variable $\Lambda(\ell)$ from (5.24)[5] is a sum of length K, approaching infinity for continuous transmission. We shall see, however, that the LRT, $\Delta\Lambda(\mathcal{E}) = \Lambda(\ell) - \Lambda(0)$, has only $L_e \ll K$ non-zero entries, since the estimation filter has finite memory M. For the derivation of the error performance of the recursive MLSD the LRT is cast into a quadratic form:

$$\Delta\Lambda(\mathcal{E}) = \Lambda(\ell) - \Lambda(0) \;=\; \mathbf{u}^H \mathbf{Q}(\mathcal{E})\,\mathbf{u} \tag{5.39}$$

then the probability of error can be upper and lower bounded following the discussion from section 5.1.5.

First the effective length L_e of an error event $\mathcal{E}$ is determined. Define $e^{(\ell)}(k) = |d^{(\ell)}(k) - d^{(0)}(k)|/2$ and the corresponding error state sequence $\chi_e(k) = \{e^{(\ell)}(k), \cdots, e^{(\ell)}(k-M)\}$ with memory M. Assuming that the transmitted sequence is the all one sequence, $d(k) = d^{(0)}(k)$, the error state χ_e equals the state definition of χ in (5.31). An *error event* $\mathcal{E}$ is said to extend from time k_1 to k_2 if $\chi_e(k)$ is equal to the correct state sequence outside the interval $\{k_1, \cdots, k_2\}$ and nowhere in between. Thus the length of the error event becomes $L_e = k_2 - k_1 - 1$. Denoting the length of an error burst by L_b we have $L_e = L_b + M$. The decision variable is affected by the burst $\mathcal{E}$ for $L_{tot} = L_b + 2M$ symbols. That is the duration in which an error burst of length L_b has an impact of the estimation – detection procedure. This is illustrated in Figure 5.9. The $M-1$ samples *prior* the occurrence of k_1 are non-zero due to the estimation filter $\mathbf{w}$. The effect of the error burst may persist for another M samples *past* k_1, due to decision feedback. However, no subsequent errors are assumed to be induced after time $k_1 + L_b$. Figure 5.10 shows possible error events in a $M = 3$ state trellis. The shortest error event with $L_b = 1$ persists for $M+1$ samples before it again merges with the path for the transmitted sequence. This single error event will be used to

[5] For simplicity the time index of the decision variable from (5.24) is omitted, $\Lambda(\ell) = \Lambda(\ell, K)$ as long as it is clear from the context.

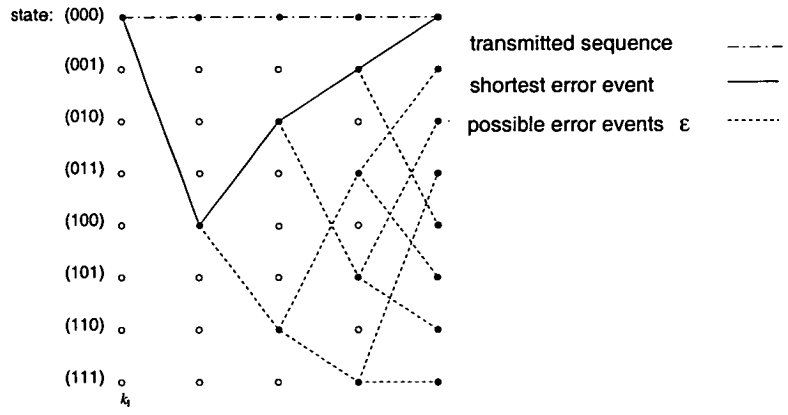


Figure 5.10: *Effect of an error burst to the differential decision variable* $\Delta\Lambda_q(\mathcal{E})$.

evaluate the lower bound, $P_e(1)$. In the following it is assumed that this is also the case for IIR channel estimation. Due to the exponential decline of the terms contributing to $\widehat{h}^{(\mathrm{IIR})}(\ell, k)$ from (5.30), this gives a reasonable approximation of the true lower bound for sufficiently large M.

The test vector $\mathbf{u} = [\mathbf{u}_1, \cdots, \mathbf{u}_Q]^T$ of dimension $\mathbb{C}^{L_{\mathrm{tot}}Q}$, containing the entries of the Q diversity taps, represents the subsequences of the symbols being affected by $\mathcal{E}$. Its entry for the q^{th} diversity tap is defined by

$$\mathbf{u}_q = \left[y_q(k_1 - M), \cdots, y_q(k_1 + M + L_b)\right]^T \qquad \in \mathbb{C}^{L_{\mathrm{tot}}Q} \tag{5.40}$$

To evaluate quadratic form $\mathbf{u}^H \mathbf{Q}(\mathcal{E})\,\mathbf{u}$ from (5.39) an expression for the matrix $\mathbf{Q}(\mathcal{E})$ is required. An expression for $\mathbf{Q}(\mathcal{E})$ is derived in Appendix A.3.2.

For the pairwise error probability, $P_e(\mathcal{E})$, equation (5.16) needs to be evaluated

$$P_e(\mathcal{E}) = \sum_{\substack{n=1 \\ \lambda_n < 0}}^{N} \prod_{\substack{\nu=1 \\ \nu \neq n}}^{N} \frac{1}{1 - \lambda_\nu / \lambda_n} \tag{5.41}$$

where the set $\{\lambda_n,\ n = 1, \cdots, N\}$ are the eigenvalues of the matrix $\mathbf{\Phi}^{(0)}_{\mathbf{uu}}\mathbf{Q}(\mathcal{E})$. The matrix $\mathbf{\Phi}^{(0)}_{\mathbf{uu}} = \mathrm{E}\left[\mathbf{u}\mathbf{u}^H \middle| \mathbf{d}^{(0)}\right]$ is the covariance matrix of $\mathbf{u}$, given that hypotheses $\ell = 0$ was transmitted. Some results are presented for MLSD with FIR and IIR filter respectively. The theoretical performance of (5.41) will be compared to simulation results for implementations of the recursive MLSD in Chapter 6.

Numerical results Generally, curves labelled "ideal" identify the case where the CIR is known *a priori*. The general assumptions for the results presented for the recursive MLSD are the same as in section 5.1.5. For results of the lower bound, $P_e(1)$ in (5.41), was evaluated for the single error event according to Figure 5.10. For the calculation of the upper bound (5.19) error bursts with length $L_b \leq 8$ were considered. Figure 5.11 shows the BER performance against the SNR for the recursive MLSD with a FIR-type predictor of order $M = 4$. A fading rate of $\nu'_{max} = 0.05$ is assumed. It is seen that the bounds are tight for high SNR ($\bar{\gamma} \geq 20$ dB), while they are rather loose for low SNR values. Given that the upper bound in

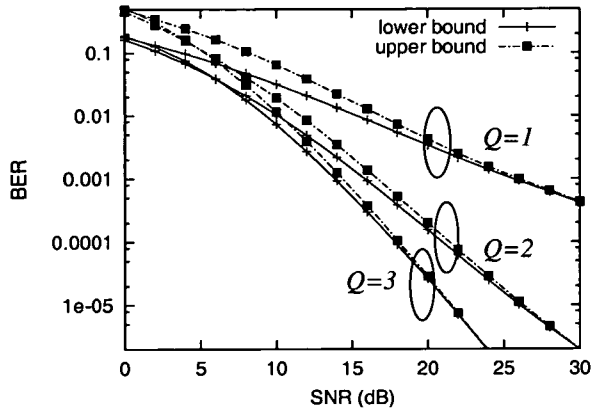


Figure 5.11: *BER vs SNR for the recursive MLSD with a FIR-type predictor of order $M = 4$. A diversity signal with Q taps is considered; $\nu'_{max} = 0.05$.*

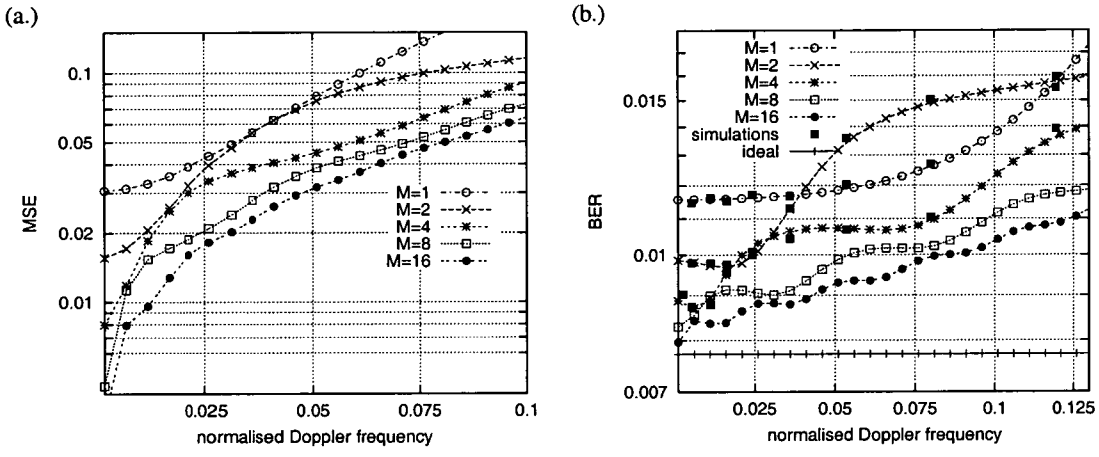


Figure 5.12: *(a.) MSE vs ν'_{max} for various M.*
(b.) BER vs ν'_{max} of the lower bound for the recursive MLSD; $\bar{\gamma} = 15$ dB, $Q = 1$.

(5.19) is not a true upper bound (see the discussion in section 5.1.5), the lower bound will be mainly used in the following. Moreover, the lower bound will prove to be more useful for comparison purposes with receiver structures implemented for simulation work in Chapter 6.

Figure 5.12 shows the MSE, V^2_{min} from (5.29) against the normalised Doppler frequency ν'_{max}, of a linear predictor (part a.) and the BER of the lower bound (part b.), for various filter orders M. In Figure 5.12.b simulations were carried out[6] in order to confirm the evaluation of the pairwise error probability, $P_e(1)$ in (5.41), which serves as lower bound. It can be observed in Figure 5.12 that the MSE of the receiver is not a reliable indicator for its error performance, although there is a rough relation between MSE and BER. For some combinations of M and ν'_{max} the receiver performance employing a M^{th} order predictor is poorer than the corresponding receiver with a $M-1$ order predictor. This is most obvious for $M = 2$ at fading rates in the range $\nu'_{max} = [0.04, 0.12]$, where a $M = 1$ yields better results

[6]For the simulations the VA–MLSD described by (5.36)–(5.38) was implemented, for details see section 6.4. In order to simulate the lower bound, which only accounts for single errors ($L_b = 1$), one unknown bit was followed by $2M$ known bits, to produce the required test sequences, $\mathbf{u}_q$ in (5.40). Due to the vast complexity of VA–MLSD for large M, simulation results for $M \geq 4$ are not included in the graph.

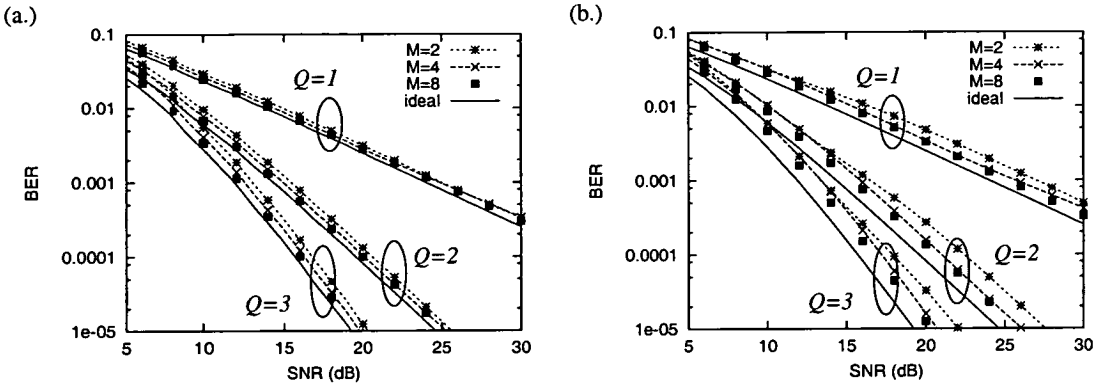


Figure 5.13: *BER vs SNR for MLSD with FIR filtering with different filter orders M and diversity taps Q.*
(a.) slow fading: $\nu'_{max}=0.005$; *(b.) fast fading: $\nu'_{max}=0.05$.*

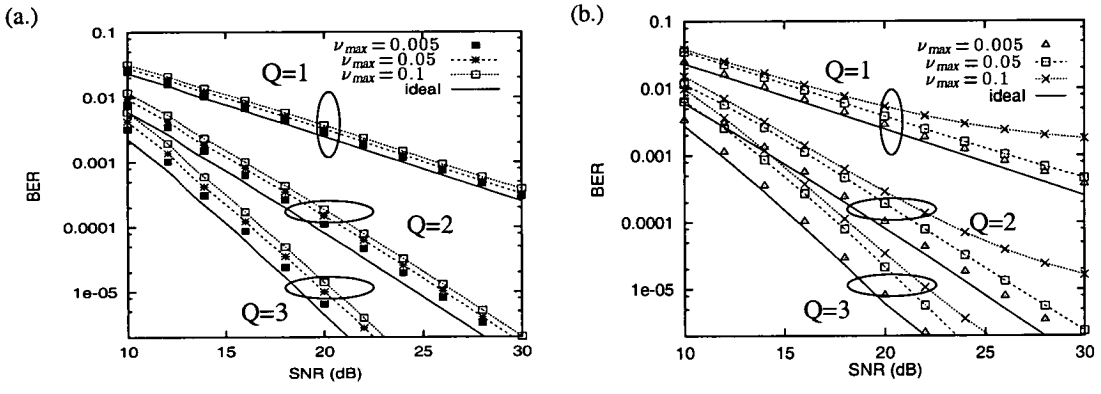


Figure 5.14: *BER vs SNR for the recursive MLSD with some ν'_{max} and Q.*
(a.) FIR filtering ($M = 8$); *(b.) IIR filtering ($\alpha = \alpha_{opt}$).*

than $M = 2$ in terms of the BER, whereas the corresponding MSE for $M = 2$ is never higher than for $M = 1$. Furthermore the BER is not a monotonic function for ν'_{max}, whereas the MSE is.

To investigate further the recursive MLSD with a FIR-type predictor, in Figure 5.13 the lower bound of the bit error probability against the SNR is shown for different numbers of filter orders M and diversity taps Q. For slow fading (part a.) there is little difference in BER by varying M between 2 and 8. For flat fading the lower bound is very close to the case were the receiver has perfect knowledge of the CIR. By introducing diversity, the difference between the lower bound and the receiver with known CIR becomes slightly larger. That is to be expected since the mean SNR per tap, $\bar{\gamma}_q$, decreases by increasing the number of taps Q, since the average overall SNR is normalised, such that $\bar{\gamma} = E_s/N_0 = \sum_{q=1}^{Q} \bar{\gamma}_q$. A degradation of $\bar{\gamma}_q$ increases the MSE, therefore the receiver performance becomes slightly poorer. For fast fading (part b.) and long filters ($M \geq 4$), there is little difference in performance.

In Figure 5.14 the lower bound of the bit error probability is shown for FIR-type and 1st order IIR-type filtering in part (a.) and (b.) respectively. Considering part (b.), the filter constant $\alpha = \alpha_{opt}$ was chosen according to (2.22). For low Doppler the performance for both FIR and IIR-type filtering is very close to the case when the CIR is known *a priori*. If the Doppler frequency increases FIR filtering outperforms IIR-type filtering, which is an expected result, due to the simplicity of the 1st order IIR filter. Especially for high SNR the plots flatten out for IIR filtering, due to an irreducible BER (IBER).

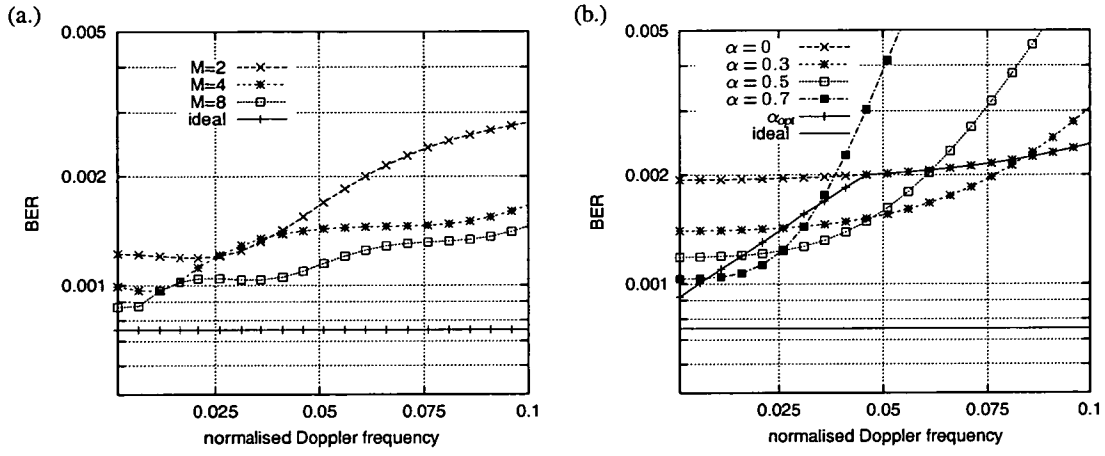


Figure 5.15: *BER vs ν'_{max} for the recursive MLSD; $\bar{\gamma} = 15$ dB, $Q = 2$.*
(a.) FIR filtering and different M; *(b.) IIR filtering and different α.*

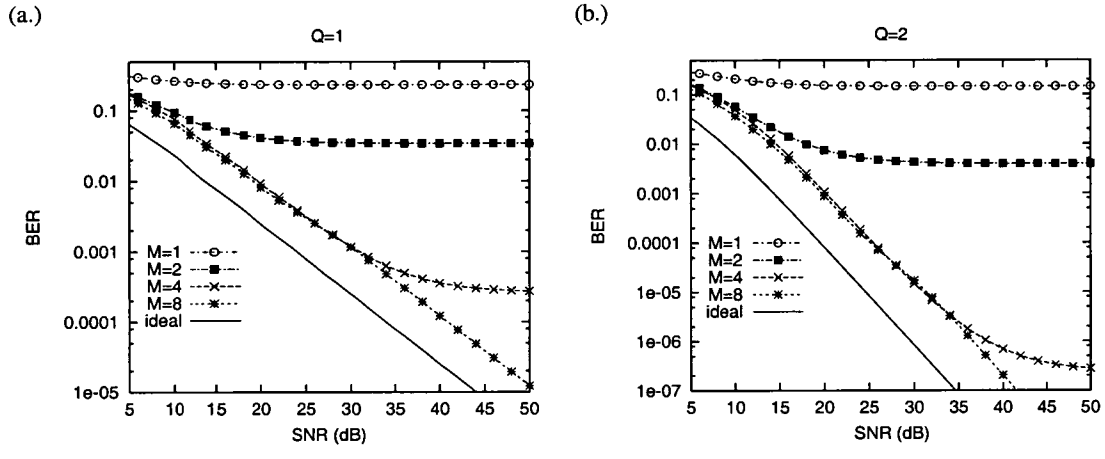


Figure 5.16: *BER vs SNR for MLSD with FIR filtering with different filter orders M for very fast fading $\nu'_{max} = 0.3$.*
(a.) flat fading: $Q = 1$; *(b.) double diversity: $Q = 2$.*

Figure 5.15 shows the BER against the maximum normalised Doppler frequency ν'_{max}, for double diversity $Q = 2$. Again FIR-type and 1st order IIR-type filtering are compared in part (a.) and (b.) respectively. In part (a.) different filter orders are shown, while part (b.) shows the dependence of the receiver on the filter constant α. For comparison purposes a plot of a receiver with $\alpha = \alpha_{opt}$ is also included in part (b.). It is observed that $\alpha = \alpha_{opt}$ is not the best possible solution to minimise the error probability, as seen in Figure 5.15.b. This conforms with the discussion in section 2.1.5 as the approximations used to derive α_{opt} from (2.22) where only valid for slow fading and low SNR. On the other hand, MLSD using a FIR filter with $M \geq 4$ is superior for fast fading, while its performance is approximately the same for slow fading.

Note that setting $\alpha = 0$ is equivalent to $M = 1$ and both cases correspond to performing MLSD without estimating the CIR at all. The CIR estimate is simply given by the pre-multiplied received signal of the previous sample, $\widehat{h}_q^{(\alpha)}(k) = y'_q(\ell, k-1)$, similar to a conventional DPSK detector. It can be observed from Figure 5.15.b that for $\nu'_{max} > 0.08$ the BER of a receiver with $\alpha = 0$ is superior to the corresponding BER for the receiver with $\alpha > 0$. It can therefore be concluded that 1st order IIR filtering is not appropriate for fast fading.

Error floor analysis: Figure 5.16 shows the performance for very fast fading, $\nu'_{max} = 0.3$, to investigate the error floors for FIR-type prediction, dependent on the predictor order. As expected the the error floors in form of an IBER decrease by employing a higher order predictor. The case $M = 1$, which is equivalent with $\alpha = 0$ for IIR filtering, is the worst case for a linear predictor. This conforms with [36], where it was shown that MLSD employing a linear predictor experiences no IBER, if the predictor order M approaches infinity. The IBER can be considerably lowered by employing diversity, as shown for a two tap diversity receiver in part (b.) of Figure 5.16. As a result, for realistic receivers with FIR prediction filters of the size $M \geq 4$, the existence of an IBER does not affect the BER performance in ranges of interest, particularly if diversity is employed. A typical raw BER on a mobile radio link is in the range $[10^{-2}, 10^{-3}]$ dependent on the application. The IBER can be lowered if a higher order Kalman filter is used [95].

5.2.5 Extension to channels with inter-symbol interference

The application of the VA to detection of signals transmitted through a multipath fading channel was originally derived for channels with inter-symbol interference (ISI) with perfectly known CIR. Optimal detection may be realized according to various receiver structures. Forney [85] showed that the receiver may be divided into two distinct components: a front-end processor, called *whitening matched filter* (WMF); and a non-linear post-processor based on the VA. In an alternative solution proposed by Ungerböck [106], the front-end processor is reduced to a matched filter, requiring a modified metric computation.

Moreover, a large part of the literature concerning MLSD with unknown parameter estimation is devoted to channels corrupted by ISI. Although ISI channels are not explicitly considered for simulation work, the generalisation of MLSD for ISI channels is briefly discussed in the following, because it is closely related to the work carried out in this thesis. The work of Lodge and Moher [36] was extended to the frequency selective fading channel in [37, 96, 97]. Applications on Equaliser structures in wireless communications are summarised in [8, 107]. The design of the receiver front end, in case the CIR is unknown was addressed in [93, 107]. Consider the received signal at time k, corrupted with ISI after sampling and matched filtering

$$y(k) = \sum_{q=1}^{Q_{ISI}} d(k-q+1)\, h_q(k) + n(k) \tag{5.42}$$

where the CIR $h_q(k)$ is equivalent to the previous discussion in this thesis. It is assumed that one received sample is affected by Q_{ISI} information symbols and the memory of the channel is M. In matrix notation the received sequence can be expressed as

$$\mathbf{y}(k) \;=\; \mathbf{D}(k)\,\mathbf{h}(k) + \mathbf{n}(k)$$

where $\mathbf{y}(k)$ and $\mathbf{n}(k)$ are of dimension M. The CIR vector is an $M + Q_{ISI}$ dimensional column vector $\mathbf{h}(k) = [\mathbf{h}_1^T(k), \cdots, \mathbf{h}_{Q_{ISI}}^T(k)]^T$, with the independent fading taps $\mathbf{h}_q(k)$ of dimension M. The data

matrix is of the block diagonal form

$$\mathbf{D}(k) = \mathrm{diag}\big[\mathbf{d}^T(k), \ldots, \mathbf{d}^T(k - Q_{ISI})\big]$$

with the data vectors $\mathbf{d}(k)$ also of dimension M.

Note, with this definition the received signal is still linear, hence the Gauss-Markov theorem still applies [16] (see section 2.1). Thus, Kailath's estimator-correlator receiver [31] with the decision variable $\mathbf{y}^H \boldsymbol{\Phi}_{\mathbf{yy}}^{(\ell)\,-1}\mathbf{y}$ in (5.4) remains the optimal receiver for (5.42). However, pre-multiplication of the received signal (5.9) does not apply in this case. Accordingly, for the recursive MLSD, the recursion in (5.21) holds, but pre-multiplication does again not apply. Following [37], the decision variable for an ISI corrupted signal is

$$\Lambda(\ell, k) = \Lambda(\ell, k-1) + \left| y(k) - \sum_{q=1}^{Q_{ISI}} d^{(\ell)}(k-q+1)\, \widehat{h}_q(\ell, k) \right|^2 \tag{5.43}$$

If the signal is corrupted by a known CIR, the estimate $\widehat{h}_q(\ell, k)$ is replaced by the true CIR $h_q(k)$. Normally, the receiver has no prior knowledge of the CIR, so it needs to be estimated. In section 5.2, the conditional mean of the pdf pertaining to the one-step prediction of the received sample $y(k)$, was identified to be $\widehat{h}_q(\ell, k) = \mathrm{E}\big[\, h_q(k) \;\big|\; \mathbf{y}(k-1),\, \mathbf{d}^{(\ell)}(k) \,\big]$. With the Gauss-Markov theorem from (2.10) the MMSE estimate of $\widehat{h}_q(\ell, k)$ is obtained by applying the Wiener-Hopf equation

$$\begin{aligned} \widehat{h}_q(\ell, k) &= \mathrm{E}\big[\, h_q(k)\, \mathbf{y}^H(k-1) \;\big|\; \mathbf{d}^{(\ell)}(k) \,\big] \cdot \mathrm{E}\big[\, \mathbf{y}(k)\, \mathbf{y}^H(k) \;\big|\; \mathbf{d}^{(\ell)}(k) \,\big]^{-1} \cdot \mathbf{y}(k-1) \\ &= \mathrm{E}\big[\, h_q(k)\, \mathbf{h}^H(k-1) \,\big]\, \big[\mathbf{D}^{(\ell)}(k)\big]^H \\ &\quad \cdot \Big(\mathbf{D}^{(\ell)}(k)\, \mathrm{E}\big[\, \mathbf{h}(k)\, \mathbf{h}^H(k) \,\big]\, \big[\mathbf{D}^{(\ell)}(k)\big]^H + N_0 \mathbf{I} \Big)^{-1} \cdot \mathbf{y}(k-1) \end{aligned} \tag{5.44}$$

With equations (5.43) and (5.44) the Viterbi algorithm can be employed as sub-optimum decoding algorithm [89, 93], similar to section 5.2.3. However, due to ISI the number of different channel estimates at any one time has increased to $A_m^{M+Q_{ISI}}$, compared to the $A_m^{Q_{ISI}}$ required for the ISI free case. Moreover and equally important, the estimation filter

$$\mathbf{w}^{(\ell)} = \big[\mathbf{D}^{(\ell)}(k)\big]^H \cdot \Big(\mathbf{D}^{(\ell)}(k)\, \mathrm{E}\big[\, \mathbf{h}(k)\, \mathbf{h}^H(k) \,\big]\, \big[\mathbf{D}^{(\ell)}(k)\big]^H + N_0 \mathbf{I} \Big)^{-1}$$

has become data dependent. However, $\mathbf{w}^{(\ell)}$ still is time-invariant. Therefore, in most practical receivers the separation principle is used, that is the CIR is obtained by a training sequence and subsequently used as if it was known *a priori*, termed trained MLSD [107, 108]. With this assumption the trellis reduces to $A_m^{Q_{ISI}}$ states and the metric computation is also grossly simplified. For instance, a trained MLSD is used for the GSM standard. With this approach the receiver complexity is grossly reduced at the expense of some loss in spectrum efficiency. Furthermore, such a receiver is only applicable to a slow fading channel. The receiver developed by Fechtel and Meyr [109] extends Cavers work [11] to ISI channels (see also [28]). That receiver utilises time multiplexed pilots and thus extends the pilot aided receiver structures studied in section 4.2 to channels with ISI.

Considering blind equalisation, adaptive MLSD algorithms were reported in [106, 110, 111]. This terminology implies that a single channel estimate is maintained and updated based on tentative decisions fed back from the VA. This also leads to a trellis with $A_m^{Q_{ISI}}$ states, since only one channel estimate is kept a certain time. The VA decoding depth for providing tentative decisions to the channel estimator is a critical parameter for that type of adaptive MLSD. The design trade-off is that a large delay is required for reliable decision feedback, while a small delay is desirable in order to track channel dynamics.

The per-survivor processing (PSP) technique eliminates this trade-off by providing parallel, zero-delay decision feedback for per-path channel estimation [92]. As a result, PSP provides superior performance and robustness at the expense of increased computational complexity, requiring $A_m^{Q_{ISI}+M}$ decoder states. PSP may be viewed as a generalisation of [37, 96, 112], although they were developed independently of [92]. The innovations based MLSD [37] was extended to Rician fading in [105], including the estimation of the second order channel statistics. Dai and Shwedyk [96] employed Kalman filters while [37] used a linear predictive receiver. A similar receiver to [37, 96] was developed independently by Kubo *et al.* [112]. In that paper channel estimation was performed using the LMS algorithm. In [112] a comparison between the PSP based receiver and conventional adaptive MLSD is given, in terms of performance complexity and implementation of the algorithms. Later Chugg and Polydoros [100] analysed PSP using LMS based channel estimators. There the channel estimation performance was analysed with regard to different receiver front-end approximations.

Another approach to blind equalisation algorithm was proposed by Seshadri [113], which operates on a number of parallel trellises, where each correspond to a hypothesised estimate of the CIR. There however, equalisation on time-invariant channels was studied, while we are mainly concerned with an improvement of the tracking capability on fast time-varying channels.

In general, the complexity of these algorithms grow exponentially with the length of the CIR. Thus, for many applications, in particular when the number of channel taps Q_{ISI} is large, these algorithms are not feasible. Therefore, *reduced complexity* variants of the VA have been devised. The number of states required due to ISI, Q_{ISI}, can be reduced if prior knowledge of the basic structure of the RF radio channel is used. This prior knowledge is used as side information, in terms of independently fading propagation paths or sky waves [114]. Suppose the impulse response $\{h_q(k)\}_{q=1}^{Q_{ISI}}$ with a large number of fading taps, Q_{ISI}, is a linear combination of Q_0 orthogonal basis vectors, i.e. $\text{rank}\big(\text{E}[\,\boldsymbol{h}(k)\boldsymbol{h}^H(k)\,]\big) = Q_0$, where $\boldsymbol{h}(k) = [h_1(k), \cdots h_{Q_{ISI}}(k)]^T$. The estimator then predicts the Q_0 independent fading paths rather than the Q_{ISI} taps of the CIR $\{h_q(k)\}_{q=1}^{Q_{ISI}}$. This approach requires $Q_0 \ll Q_{ISI}$ in order to considerably reduce the complexity. In [115] a recursive MLSD is presented where the CIR is expressed as a linear combination of polynomial basis vectors.

Another approach is to reduce the number of decoder states in the trellis [116–119]. One such technique is the M-algorithm [116]. Hashimoto [117] combined the M-algorithm with the VA, leading to a *list type reduced-constraint generalisation of the VA* (LVA). The proposed algorithm contains the Viterbi and the M-algorithm as a special case. Another approximation of MLSD is *reduced state sequence estimation* (RSSE) developed by Eyuboğlu and Qureshi [118, 119]. There subsets of states were merged to super-states and the number of transitions of these super-states were reduced with *state dependent*

decision feedback. The receiver is a combination of the traditional MLSD and decision feedback equalisation (DFE) [7]. Iltis *et al.* [97] implemented the recursive MLSD employing a bank of parallel Kalman filters by using RSSE. PSP is also a form of reduced complexity solution for MLSD. RSSE and PSP are very similar approaches, since they both exploit state dependent decision feedback. Thus, the proposed receiver in [97] may also be viewed as a PSP based receiver.

5.3 Soft output detection

Following the discussion in section 1.2.2, supplying hard decisions for subsequent processing contradicts Shannon's information theory, since information about the received signal is discarded prematurely. Improved performance can be achieved by applying *soft* decisions to the channel decoder. These soft decisions may be used as a form of reliability information in subsequent receiver stages. With optimum sequence detector from (5.24) soft decisions may be provided with a channel state information (CSI), that is the CIR estimate of the ML sequence $\widehat{h}_q(k)$. An improved CSI can be obtained by a 2^{nd} stage estimation filter [120], which is a smoothing type filter. This is similar to the iterative channel estimation from section 4.2.2 and 4.3.2. A system using MLSD to obtain tentative decision including a CSI, for decoding of Trellis modulated signals was investigated in [98, 120].

However, a more powerful approach is to compute the *a posteriori* probability $p\big(d^{(\ell)}(k)\big|\mathbf{y}\big)$ from (2.25) which is provided by the MAP symbol-by-symbol detector (MAP–SbSD). Let a state at time k be denoted by $\chi(k) = \{d^{(\ell)}(k), \cdots, d^{(\ell)}(k-L)\}$, where L is the combined memory of the channel and modulator. By applying Bayes' theorem, the MAP probabilities can be expressed as

$$p\big(d^{(\ell)}(k) \mid \mathbf{y}\big) = \sum_{i' \in \mathcal{S}} p\big(\chi(k-1) = i', \chi(k) = i \mid \mathbf{y}\big) \tag{5.45}$$

where $\mathcal{S}$ is the set of states. The posterior probability of the state transition from state $\chi(k-1) = i'$ to $\chi(k) = i$ is in the form

$$p\big(\chi(k-1) = i', \chi(k) = i \mid \mathbf{y}\big) = \frac{\sum_{\ell \in \mathcal{A}_K(i',i)} p\big(\mathbf{d}^{(\ell)} \mid \mathbf{y}\big)}{\sum_{\ell \in \mathcal{A}_K} p\big(\mathbf{d}^{(\ell)} \mid \mathbf{y}\big)} \tag{5.46}$$

where $\mathcal{A}_K$ denotes set of all hypothesis and $\mathcal{A}_K(i', i)$ is the subset of hypothesis that traverse the trellis branch between states $\chi(k-1) = i'$ and $\chi(k) = i$. Note, $p\big(\mathbf{d}^{(\ell)}\big|\mathbf{y}\big)$ represents the *a posteriori* probability for the sequence $\mathbf{d}^{(\ell)}$, which is the sequence detector from the previous section. The above equation states that, for each branch in the trellis the following ratio needs to be computed

$$\frac{\text{Sum of MAP sequence probabilities which pass through a given branch}}{\text{Sum of MAP sequence probabilities of all paths through the trellis}}$$

It is observed that the MAP–SbSD and the MLSD are closely related through (5.46) [121]. Since the MAP–SbSD consists of the sum of MAP sequence probabilities, the resulting receiver structure requires a much higher computational complexity than a corresponding sequence detector.

A solution including forward and backward recursion was given by Bahl *et al.* [38]. This algorithm

was applied to a flat fading channel with unknown channel response in [122]. Optimum MAP–SbSD under the constraint of a fixed delay was developed by Abend and Fritchman [123]. This approach seems to be more suited for real time applications because it only requires forward processing. Unfortunately, the number of variables that need to be stored and updated grows exponentially. Recently, [124] reformulated the Abend & Fritchman algorithm [123], with the result that the complexity grows only linearly with the sequence length. Their algorithm was extended to the case of unknown channel parameters by utilising PSP [125].

The development of *iterative* decoding techniques, can yield vast improvements in system performance compared to non-iterative systems [122]. There detection and decoding is performed iteratively, where soft decisions between decoder and detector are exchanged. This requires some sort of soft-in soft-out detection and decoding, performing an update of the *a posteriori* probabilities of both information and coded symbols [126]. In particular the development of "turbo codes" [127, 128] has brought the code performance much closer to Shannon's channel capacity. This is achieved at the expense of increased complexity and coding delay, which may be unacceptable for real time applications, such as wireless telephony.

Optimum MAP–SbSD may be too difficult to implement in practice, basically because of the occurrence of non-linear functions and and mixed multiplications and additions of these values, which make it difficult to transform them into the log-domain. Faced with the high computational complexity of MAP–SbSD, sub-optimum approximations operating in the log-domain are commonly used. The Max–Log–MAP [129, 130] imposes the maximum rule to the Euclidean distance computation occurring in the log-domain: when paths merge only the one with the best metric contributes to the new state probability, such that the resulting recursion is exclusively additive. Techniques to reduce the complexity even further are [131]: state reduction by *state dependent decision feedback*; and reduced backward recursion.

There has also been some research activity in extracting soft-outputs from MLSD with the VA, by producing so-called reliability informations. This has led to the concept of list-type Viterbi algorithm (LVA) [132], and the soft-output Viterbi algorithm (SOVA) [133]. For MLSD with unknown parameter estimation PSP techniques can be employed for both the LVA and SOVA.

Algorithms providing soft-decisions and operating in the log-domain are compared and analysed in [134–136]. In [135] an adaption of the SOVA was derived which made it equivalent to the Max–Log–MAP, while [134] modified the Max–Log–MAP such that it was equivalent to optimum MAP–SbSD but without its computational burden.

5.4 Summary and conclusions

In this chapter receiver architectures were reviewed which are optimal for linear modulated signals, transmitted through a random time-variant channel described by a Gaussian pdf. The optimum ML sequence detector from section 5.1 follows Kailath's separation theorem [31]. That is the receiver consists of an estimator that delivers the MMSE estimates of the fading distortion and a detector that utilises these estimates, by decorrelating the received signal. It was shown that if the channel estimate can be determined

independently of the data sequence, the estimation and detection tasks are independent of each other. This has established a link between the estimator–correlator receiver and the one-shot receivers studied in Chapter 4.

Unfortunately, the optimum receiver may be too complex for most practical application, since its complexity grows exponentially with the sequence length. The recursive formulation for MLSD studied in section 5.2 provided a powerful tool to reduce drastically the complexity by application of the Viterbi algorithm (VA) to the problem. Although, the resulting receiver termed VA–MLSD is generally sub-optimum, it is the best possible solution when the channel estimator is truncated by a M^{th} order moving average linear predictor. Alternatively, an IIR-type channel estimation filter can be applied to the VA–MLSD. The application to diversity systems and the extension to signals with ISI and MAI was also addressed. The performance of the VA–MLSD was lower and upper bounded by an analytical expression to evaluate the pairwise error probability of an error event. Unfortunately, the bounds are only tight for high SNR. Thus, in the following chapter, the VA–MLSD and reduced complexity derivatives based on per-survivor processing (PSP) will be extensively analysed by means of Monte Carlo simulations to examine the utility of the theoretical bounds.

Finally, in section 5.3 an overview of algorithms deemed to implement the MAP symbol–by–symbol detector was given and similarities to MLSD were pointed out.

Chapter 6
Implementation of MLSD with Unknown Parameter Estimation

The techniques studied in this chapter are based on the recursive ML sequence detector [35–37] from section 5.2. The PSP based receiver applied to ISI free channels is examined, implemented according to section 5.2.3. PSP was used for detection of MPSK modulated signals transmitted over a flat fading channel in [98, 120], using a predictive FIR filter. In [137] Kam *et al* extended their PSP based receiver in [101] to the Rayleigh fading channel, using a 1^{st} order IIR filter to estimate the channel response. Simulation results for both the FIR and IIR type channel estimation filter will be presented.

Following section 5.2.3, the computational cost of VA-MLSD (the Viterbi algorithm applied to recursive MLSD) is of order $O(A_m^M)$, which corresponds to the number of states in the trellis. In section 6.1 state reduction techniques to reduce the complexity of the receivers will be discussed, involving decision directed techniques. This compromises the optimality of the receiver but drastically reduces its complexity. The performance of PSP based on a two state trellis will be assessed through simulations in section 6.3. The robustness of the receiver will become an issue due to decision feedback effects. Conditions which can cause stability problems due to decision feedback are addressed in section 6.3.3. A performance bound employing a Markov model is derived which matches the simulation results more closely than the lower bound from section 5.2.4. In section 6.4 simulation results for other state reduction techniques are presented. Receiver structures considered in that section offer generally better performance, are more robust, but are more complex than the receivers from section 6.3. Then hybrid receiver structures are proposed in section 6.5, which are less complex than the algorithms in section 6.4 and are more robust than the algorithms in section 6.3. The major results of the simulations and analysis are summarised and discussed in section 6.7.

6.1 State reduction techniques

The number of decoder states of VA-MLSD (the Viterbi algorithm applied to recursive MLSD described in section 5.2.3) grows exponentially with the estimation filter memory M. Thus the computational cost of processing A_m^M states in the trellis may still be prohibitively large. From the reduced complexity variants of the VA, originally proposed for the ISI corrupted channel reviewed in section 5.2.5, there are two techniques which will be the focus of further investigation:

(i.) *State dependent decision feedback* (SDDF) proposed by Eyuboğlu and Qureshi [118]. Only the most significant states are taken into account, the remaining states are determined with decision feedback. PSP can be interpreted as an application of this technique.

(ii.) *List type Viterbi algorithm* (LVA) developed by Hashimoto [117]. States are merged to super-states. Then the M-algorithm [116] is performed on these super-states.

6.1.1 State dependent decision feedback

The complexity is reduced by means of state reduction [118]. Consider the cross correlation between the CIR and its estimate $\phi_{\widehat{h}h}(0) = \mathrm{E}[\,\widehat{h}_q(k)\, h_q^*(k)\,]$, which is a sum of correlation coefficients:

$$\phi_{\widehat{h}h}(0) = \sum_{m=1}^{M} w_{qm}^*\, \phi_{hh}(m) = \sum_{m=1}^{M} \psi_m \tag{6.1}$$

where $\psi_m = w_{qm}^*\, \phi_{hh}(m)$ denotes the correlation coefficient of filter tap m. The number of states in the trellis can be reduced if only the most significant taps ψ_m which contribute in the calculation of the Euclidean distance, $\Delta(i, i')$ from (5.33), are represented in the state description. All other conditional entries of the estimated CIR are determined by state dependent, tentative decisions, $\tilde{d}^{(i)}(k)$. These tentative decisions are given by the path history $\tilde{\mathbf{a}}^{(i)}(k)$ of state $\chi(k) = i$. Decision feedback of $\tilde{d}^{(i)}(k)$ is then used to determine the decision variables of subsequent samples.

The significance of a decoder state is given by the magnitude of the corresponding filter tap ψ_m. For slow fading ψ_m decreases monotonically with increasing m. Thus, the first D taps $\{\psi_0, \cdots, \psi_D\}$ are the most significant. For fast fading or long filters with a large M, there may be zeros in ψ_m and therefore the most significant D taps will no longer be the first D taps. This particular form of state reduction was implemented by Mehlan *et al* [131] for soft output equalisation. For the sake of simplicity the case of relatively short M will be considered, where ψ_m is a monotonic function of m. With these assumptions a state description at time instant k in (5.31) changes to

$$\chi(k) \triangleq \{a^{(i)}(k),\, a^{(i)}(k-1), \cdots,\, a^{(i)}(k-D+1)\} \tag{6.2}$$

and the corresponding set of states changes accordingly to $\mathcal{S} = \{\chi(k) : \; i = 0, \cdots, L-1\}$, where $L = A_m^D$, $D \leq M$. The computational cost is of the order $O(A_m^D)$, which has become independent on the filter order M. With the definition of (6.2) the CIR estimate from $\chi(k) = i$ to $\chi(k+1) = i'$ is obtained

$$\widehat{h}_q(i, k+1) = \sum_{m=0}^{D-1} w_{q,m+1}^*\, d^{(i)*}(k-m)\, y_q(k-m) + \sum_{m=D}^{M-1} w_{q,m+1}^*\, \tilde{d}^{(i)*}(k-m)\, y_q(k-m) \tag{6.3}$$

where $d^{(i)}(k)$ and $\tilde{d}^{(i)}(k)$ are the state dependent hypothesis and the tentative decision of the transmitted signal, respectively. The first term in (6.3) represent the filter taps of the finite state machine in the trellis; the second term represents taps determined through the path history by decision feedback, specified by the survivor path $\tilde{\mathbf{a}}^{(i)}(k)$ of that state. The survivor for state $\chi(k+1) = i'$ is obtained by extending the metric and applying the VA given by (5.36)–(5.38). Note that setting $D = M$ gives the VA-MLSD described in 5.2.3. An interesting special case is $D = 1$, resulting in a $L = A_m$ state trellis. This technique has attracted considerable research interest recently [30, 101, 120] and will be analysed in section 6.3.2. Setting $D = 0$ results in a entirely decision directed receiver. Thus, all filter taps which contribute in the

calculation of $\widehat{h}_q(i,k)$ and $\Delta(i,i')$, are obtained through decision feedback. Note, PSP with $D=0$ is closely related to the decision directed two stage receiver studied in section 4.3.2.

The introduction of state dependent decision feedback provides an additional degree of freedom for the design of a PSP based receiver, since D and M can be chosen independently to optimise the receiver design. So higher order predictors may now be employed for a certain D, without significantly increasing the complexity. In the following discussion in this chapter, this more general description of the VA applied to MLSD will be assumed.

6.1.2 List type Viterbi algorithm

With some modifications state reduction techniques, previously applied to channels with ISI and known CIR, can be applied MLSD for an unknown channel. These techniques are based on the fact, that of the A_m^D survivors, only a few have a sufficiently small metric to contribute to the overall best survivor path, and therefore the most likely sequence. The vast majority of the survivors can be discarded before the final decision is made, without compromising the system performance. The purpose of this algorithm is to keep the decoding complexity less than that of the VA while avoiding error propagation due to reduced state decoding.

A large class of such algorithms are constituted by the list type Viterbi algorithm (LVA) [117, 132]. The LVA in [117], contains the VA and the M-algorithm [116], a breadth first sequential decoding algorithm, as a special case. Every state retains a list of J best candidates and the M-algorithm is performed for each state to update the path metrics. The application of the M-algorithm for MLSD with unknown parameter estimation was proposed by Auer *et al* [138] and will be described in the following.

Applying the LVA to MLSD involves building subsets $\mathcal{S}^\star$ of the full set of states $\mathcal{S}$, termed super-states $\chi^\star(k)$. Let all states $\chi(k)$ which are the same for the first $D_{\mathcal{S}^\star}$ symbols, belong to a certain subset $\mathcal{S}^\star$, with $D_{\mathcal{S}^\star} \leq D \leq M$. A super-state $\chi^\star(k)$ then consists of all states within a certain subset $\mathcal{S}^\star$. With this definition there exist $A_m^{D-D_{\mathcal{S}^\star}+1}$ of such subsets, each having $A_m^{D_{\mathcal{S}^\star}}$ candidates. The LVA only keeps a certain number of candidates per super-state, $J \leq A_m^{D_{\mathcal{S}^\star}}$; while discarding all other survivor paths. The algorithm works as follows:

Path extension: In order to extend the list to sample $k+1$, $A_m J$ candidates per state, i.e. A_m per list element j need to be computed. That involves calculating the transitions from $\chi^\star(k)$ to $\chi^\star(k+1)$ including the CIR estimate for that transition, to obtain $A_m J$ decision variables. Then the J best paths with the smallest metric are stored for each state. The remaining $(A_m-1)J$ paths are discarded. The principle of list type Viterbi processing is shown in Figure 6.1. It illustrates the process of updating the list, with the survivors and discarded paths drawn with solid and dashed lines, respectively. The M-algorithm is performed for the J candidates of each state. Thus, the LVA can be regarded as a state dependent M-algorithm.

In order to find the J best paths of the $A_m J$ candidates, sorting is necessary. However, since J should be a small number the particular choice of the algorithm and impact on the overall complexity should

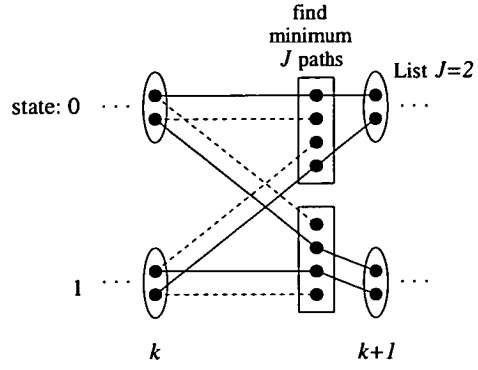


Figure 6.1: *List type survivor processing (J-SP) illustrated in a trellis with two states.*

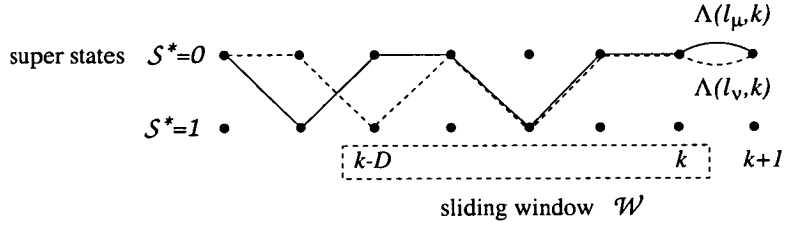


Figure 6.2: *The ambiguity check for List type survivor processing (J-SP).*

be negligible. The list of J survivor sequences per super state $\chi^{\star}(k)$ defines the J best candidates up to sample k. The j^{th} entry of the list is in the form

$$\widehat{\Lambda}^{(j)}(k) = \min_{\substack{\ell \in \hat{A}_k \\ \ell \neq \{1,\cdots,j-1\}}} \Lambda(\ell, k)\ ; \qquad j = \{1, \cdots, J\}\ . \tag{6.4}$$

The first list element, being the one with the minimum metric is computed first, then the second down to the J^{th} element. Note, the actual ML path may not be among the J candidates, a situation which becomes more likely with decreasing J.

Ambiguity check: Note, for the transition from $\chi^{\star}(k)$ to $\chi^{\star}(k+1)$, only previous samples with a time delay equal to or less than M, are of interest. Therefore, a further condition for a path to be within J is that all subsequences with time delay $D \leq M$ are mutually different. An ambiguity check prevents the accidental storage of two paths with the same symbols [116]. Let the subset of paths with a maximum time delay D, where $D \leq M$, be defined by $\mathcal{W}$. If two path maps differ once (as they do initially), they may become identical when the differing symbols are dropped, considering only time delays smaller than D, i.e. they are no longer with the sliding window defined by $\mathcal{W}$. The basic idea of the ambiguity check is illustrated in Figure 6.2. If the sliding window $\mathcal{W}$ moves from time step k to $k+1$, the path associated with either the dashed or solid line will be discarded by the ambiguity check, regardless all other survivor sequences being in the list. In other words, paths which are the same within $\mathcal{W}$ are deleted by the ambiguity check, as the path with the larger metric can never be the ML path. To justify this constraint consider two subsequences having the same path history for at least $D = M$ samples.

The Euclidean distance of these two paths, $\left|\sum_{m=0}^{M} w_m^* y_q'(\ell, k-m)\right|^2$, would be identical, hence no information is lost if the path with the larger metric is discarded. The LVA in [132] searches for the globally J best paths, however, here the local J best path are of interest.

Computational cost: Denoting the number of super states by $L^\star = A_m^{D-D_{\mathcal{S}^\star}+1}$, the overall number of paths at a given time is $JL^\star$. Thus, the computational cost of the LVA is of the order $O(L^\star J) \leq O(A_m^D)$, where $O(A_m^D)$ denotes the computational cost of PSP with $L = A_m^D$ states. Updating the trellis involves calculating $A_m J L^\star \leq A_m^{D+1}$ transitions. The corresponding path extensions of these transitions are then sorted and updated finding the minimum $A_m J$ paths. The cost of a transition is mostly dependent on the estimation filter type and its order.
There are a number of interesting special cases:

(i.) $D_{\mathcal{S}^\star} = D$, with $\mathcal{S}^\star = \mathcal{S}$ yielding a single super-state. The resulting algorithm is equivalent to the original M-algorithm [116]. In the simplest case when $J = 1$, the M-algorithm essentially reduces to the operation of a decision feedback equaliser (DFE) [7, 139].

(ii.) $D_{\mathcal{S}^\star} = D-1$, then A_m super-states per time step are obtained. The computational cost of the algorithm is of the order $O(A_m J)$. The super-states perform a generalised type of PSP, where a list of the J best paths is kept instead of a single survivor for the original PSP, which will be called J-survivor processing (J-SP) [138]. This case was investigated in [138], and accordingly simulation work in this thesis will be limited to this case.

(iii.) $D_{\mathcal{S}^\star} = 0$, then the super-states become the same as the original states and $J = 1$, which is PSP with $L = A_m^D$ states. Finally, if $M = D$ the original VA-MLSD without state reduction is obtained.

6.2 Reference phase tracking

Per-survivor processing (PSP) can be applied in a straightforward manner if orthogonal waveforms, or non-coherent modulation such as differential encoding is employed [140]. Differential encoding of the data bits, however, potentially imposes a significant degradation in system performance. On the other hand, in order to perform coherent detection, a form of phase reference for the receiver must be provided. The idea of reference symbol phase tracking was introduced in [10, 11]. These receivers only used the pilot symbols multiplexed in the data stream for channel estimation. Such a receiver was studied in section 4.2. The performance can be improved, especially for fast fading, if pilot symbols as well as data symbols are used for channel estimation, developed by Irvine and McLane [141]. The idea of pilot-aided plus decision-directed channel estimation was further developed in [142]. Their receiver combined decision feedback and adaptive linear prediction. This essentially requires a more sophisticated receiver design and joint estimation and detection of the entire transmitted sequence becomes attractive. The PSP based receiver discussed in section 5.2.3 can be employed for the pilot symbol insertion (PSI) approach. PSP was applied to a pilot aided system in a flat fading channel by [98, 120]. For MLSD on a pilot symbol aided system, the data aided information is embedded in the VA. Since channel estimation and

detection are performed jointly, the combination of PSP and reference symbol phase tracking appears to be a particularly suitable concept. First of all, the spacing of the pilots can potentially be extended, since it is no longer primarily dependent on the fading rate. Hence, the spectral efficiency of the receiver may be improved, due to a reduction in system overhead, because less pilot symbols are required. Second, the recursive MLSD is derived as the optimum receiver, thus its performance is ultimately superior to the empirical approaches in [141, 142].

6.2.1 Differential modulation

According to the discussion in section 4.1, differential modulation can be applied to PSP in a straightforward way, by means of differentially encoded PSK (DEPSK). In this case detection is performed in analogy to coherent PSK and the receiver output $\widehat{d}(k)$ of (4.1) is then differentially decoded. The differential decoding is done after the final decisions of the Viterbi processor, i.e. $\Delta\widehat{d}(k) = \widehat{d}(k)\,\widehat{d}^*(k-1)$. The VA-MLSD itself remains unchanged and can be employed according to (5.36)–(5.38).

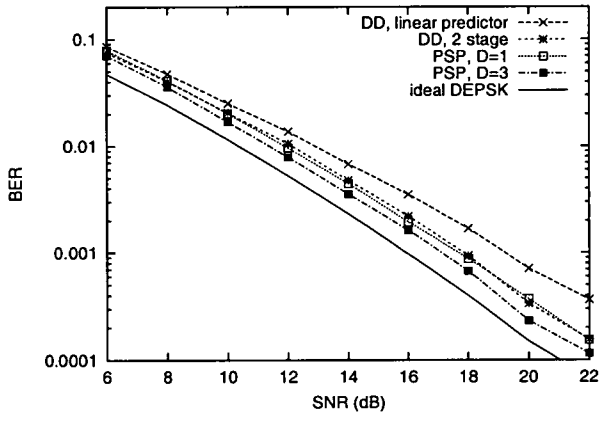


Figure 6.3: *BER vs SNR for the 2–stage decision directed receiver (DD–RAKE), compared with the linear predictive DD–RAKE and the PSP based receiver with* $M = 4$. $Q = 2$, $\nu'_{max} = 0.05$.

The receiver performance of the receiver is plotted in Figure 6.3, together with the decision directed (DD) receivers from section 4.3, these are the decision directed linear predictor and the 2–stage receiver from section 4.3.2. The curve labelled "ideal DEPSK" shows the receiver performance when the CIR is known *a priori* to the receiver, which serves as a lower bound. It is seen that for DEPSK, the performance of PSP with $D = 1$ is identical to the two stage receiver, discussed previously, while its complexity is approximately doubled. Moreover, simulation results suggest that some improvement is observed for PSP with $D = 3$, particularly for high SNR. For DPSK, the use of PSP cannot be justified because decision directed receivers with less complexity perform almost as well. So, PSP with differential modulation will not be pursued further.

6.2.2 Pilot aided channel estimation

The necessary phase reference is incorporated in the form of time multiplexed pilot symbols, named pilot symbol insertion (PSI). The multiplexing rate is R; one known symbol is followed by $R-1$ data

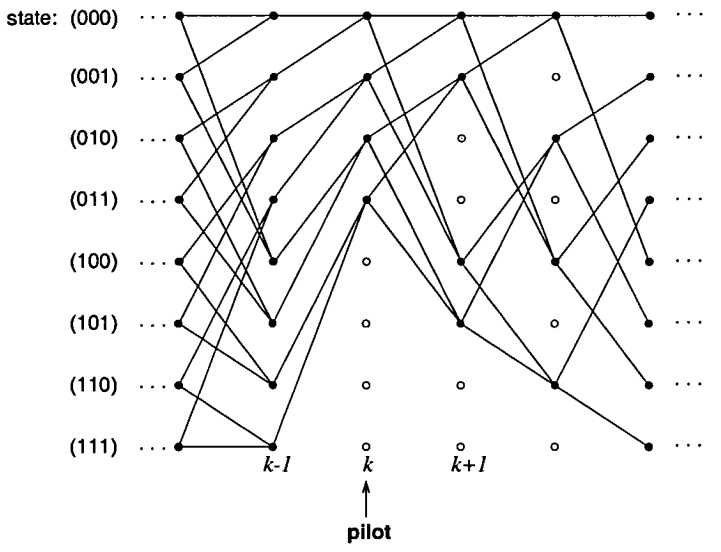


Figure 6.4: *Allowed transitions in the trellis when a pilot is being detected, $D = 2$.*

symbols. If a pilot is being transmitted, $d^{(\iota)}(k)$, is defined to be 1 and therefore states $\chi(k)$, which converge to $d^{(\iota)}(k) = -1$, are not allowed. The multiplexed pilots can be incorporated in (5.38), (5.27) and (5.30) by re-defining the pre-multiplied received signal

$$y_q'^{(\iota)}(k) = \begin{cases} h_q(k) + n(k) \; ; & k \mod R = 0 \\ d^{(\iota)*}(k)\, y_q(k) \; ; & \text{elsewhere.} \end{cases} \tag{6.5}$$

The following M samples the pilot is shifted through the state machine, leaving half of the possible states in the trellis for $M+1$ samples. This is illustrated in Figure 6.4, where a pilot is assumed to be received at sample k. In general, a state is *not* allowed because of a transmitted pilot symbol at time instant $k-m$, if the following relation is true:

$$d^{(\iota)}(k-m) \neq 1 \; ; \qquad (k-m) \mod R = 0$$

Pilot aided channel estimation combined with PSP and the LVA will be analysed in the remainder of this chapter.

6.3 PSP based on a 2–state trellis

In this section the performance of PSP with two states is investigated which is particularly attractive, due to its low complexity. The effect in terms of performance degradation and decision feedback effects are extensively analysed.

In this section the decision delay inherent in the VA is truncated to $D = 1$, so a trellis with $L = 2$ states is obtained. Following the discussion in section 6.3.1, the state $\chi(k) = i$ is defined by $\chi(k) = a^{(i)}(k)$. So state $\chi(k) = i$ is identical to the information symbol $a^{(i)}(k)$, which contains $\log_2 A_m$ bits.

That is, all filter taps which contribute to the calculation of (6.3) are obtained through decision feedback, consequently the CIR estimate is obtained based entirely on the tentative decision of the transmitted signal $\tilde{d}^{(i)}(k)$, i.e. $\widehat{h}_q(i,k) = f\big(d^{(i)}(k-1), \tilde{d}^{(i)}(k-2), \cdots, \tilde{d}^{(i)}(k-M)\big)$. Apart from that, the VA applied to recursive MLSD from (5.36)–(5.38) is employed with the parameters $D = 1$ and $L = 2$. This particular receiver was studied in [120, 137].

A block diagram of the resulting receiver structure and the corresponding trellis representation is depicted in Figure 6.5.[1] The receiver structure may be separated into four parts, these are:

(i.) Pre multiply the received signal to remove the data modulation of the received signal.

(ii.) Calculate the decision variable $\Lambda(i', i, k+1) = \widehat{\Lambda}(i,k) + \Delta(i,i')$. In general, there are A_m states, leading to A_m^2 possible transitions. For the diversity receiver, combining of the Q diversity taps is performed after calculating the Euclidean distances to obtain $\Delta(i,i')$ in (5.33).

(iii.) The Viterbi processor selects A_m survivors out of the A_m^2 candidates according to (5.38). Channel estimation is performed in a per-survivor fashion, i.e. the time delayed decision $\tilde{d}^{(i)}(k)$ is fed back and applied to the received signal.

(iv.) After a delay of R symbols only one survivor is left, yielding the final decision $\widehat{d}(k-R)$. According to (6.5) it is assumed that $d(k) = d^{(0)}(k) = 1$ if a pilot is transmitted.

6.3.1 Simulation setup

Simulation work is based on a complex baseband urban channel, described in section 3.2.6. The statistics of the qth diversity tap are specified by the classical Doppler power spectra [52], having the ACF $\phi_{hh,q}(1) = J_0(2\pi\nu'_{max})$. The bit error rate (BER) was obtained by simulating the MLSD receiver designs over a large number (between 10^6 to 10^7) of Monte Carlo runs.[2] Binary modulation (BPSK with $A_m = 2$) in the form of BPSK was used for all results presented in this chapter. Simulation results are compared to conventional DPSK,[3]. Generally, for the discussion throughout this chapter, curves labelled "ideal" identify the case where the CIR is known *a priori* according to (4.5), and "theory" is the label for the lower bound of (5.41). Unless otherwise stated the results presented in this section, are based on the parameter specifications in Table 6.1.

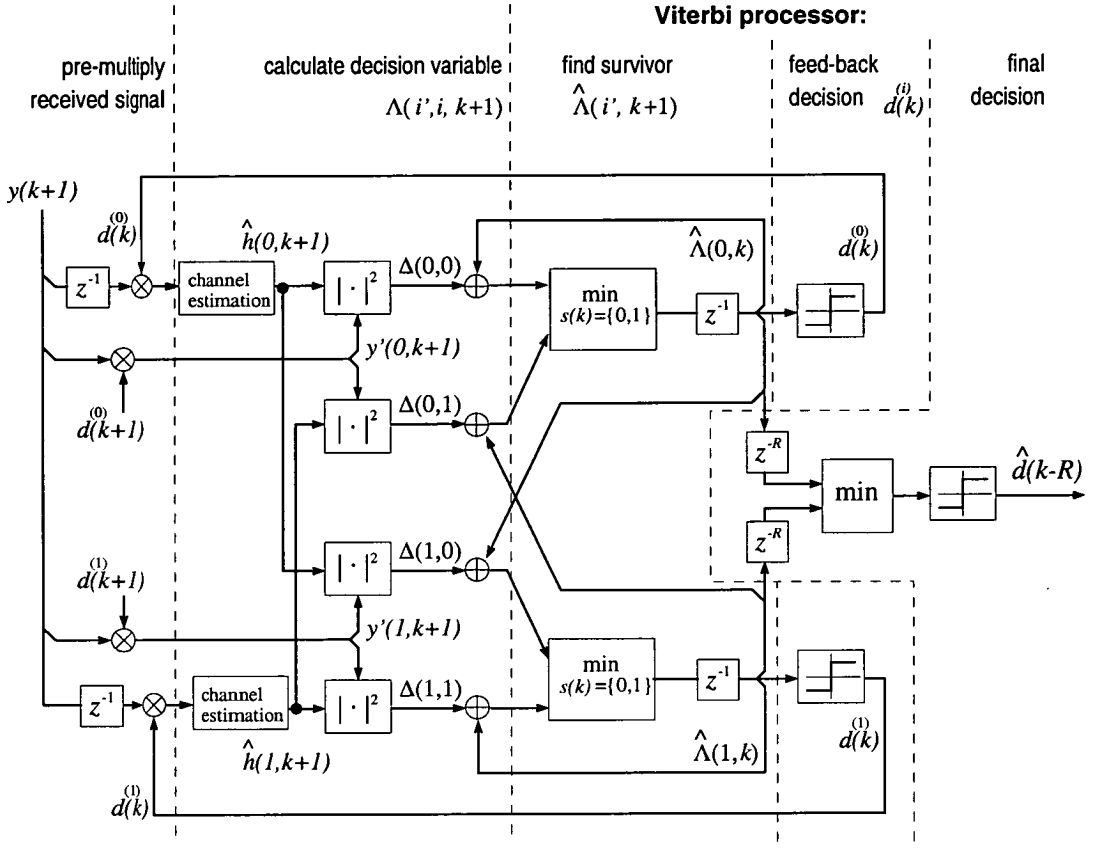
6.3.2 Results for FIR estimation filter

The PSP based receiver employing a FIR estimation filter is considered in the Figures 6.6–6.8. The BER against the average SNR, $\bar{\gamma}$, for conventional DPSK and FIR–PSP, operating in a fast fading channel

[1] For simplicity a binary non-diversity receiver is shown ($Q = 1$, $A_m = 2$). The subscripts for the q^{th} diversity tap have been dropped, since no diversity is considered in the graph, e.g. $y_1(k) = y(k)$. The generalisation to larger Q and A_m are straightforward, however, it would rather obscure the receiver principles.

[2] The number of runs necessary to achieve sufficient convergence is dependent on the particular receiver realisation and the simulation parameters, e.g. some receivers produce long error bursts which occur very occasionally, hence a larger number of runs is needed to average over a sufficient number of bursts. Or for slow fading there are longer but less deep fades, requiring more runs.

[3] Conventional DPSK, where no attempt is made to estimate the CIR, was discussed in section 4.1 and the probability of error is given in (4.6).



Corresponding trellis:

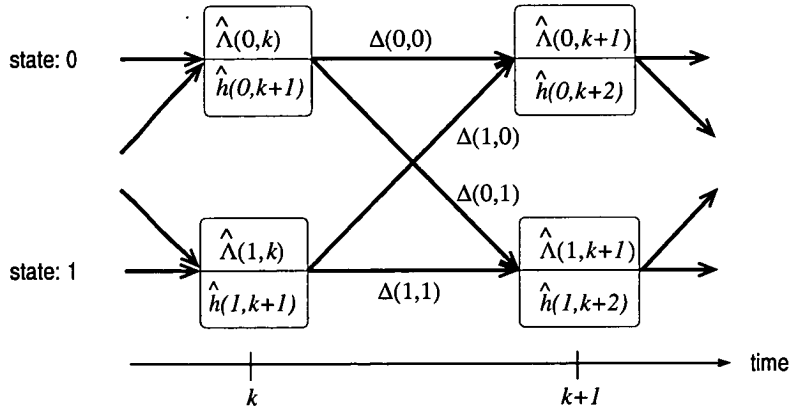


Figure 6.5: *Block diagram of the receiver structure for PSP, with $A_m = 2$ states and flat fading, $Q = 1$.*

VA decoding delay	D	1
$\rightarrow$ Number of states	L	2
Modulation		BPSK
Phase reference technique		PSI

Table 6.1: *System & simulation parameters for 2-state PSP.*

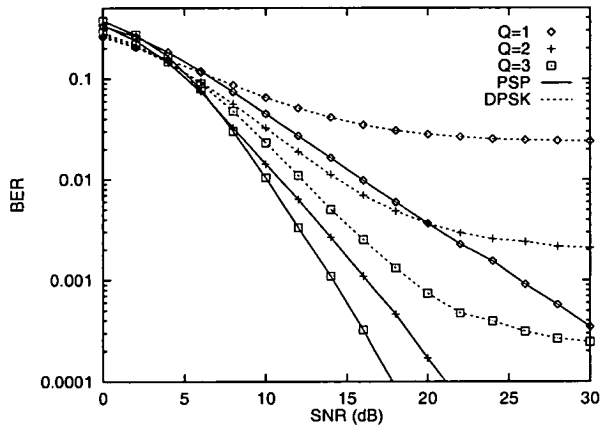


Figure 6.6: *BER vs SNR for FIR-PSP (solid lines) and conventional DPSK (dashed lines) for different numbers of diversity taps Q. $M=8$, $R=10$, $\nu'_{max}=0.05$.*

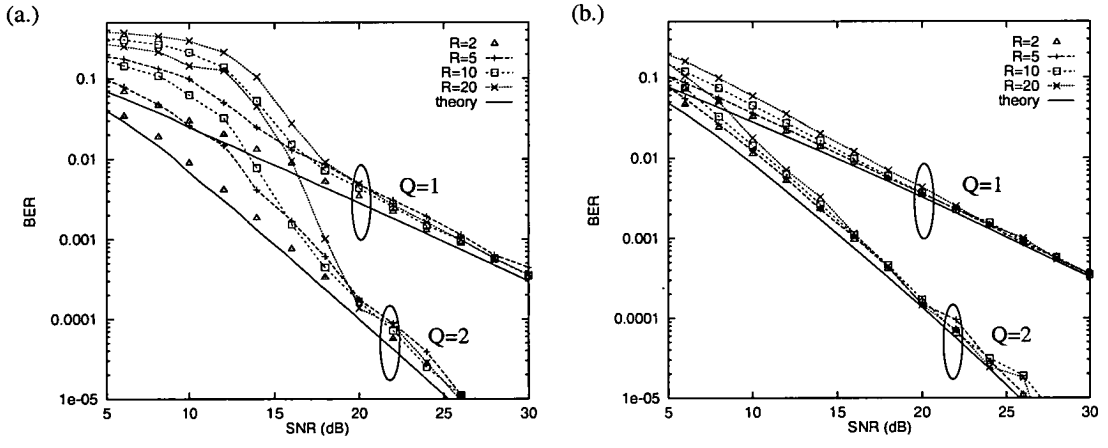


Figure 6.7: *BER vs SNR for different ratios of data to known symbols R. $M=8$,*
(a.) slow fading: $\nu'_{max}=0.005$, *(b.) fast fading: $\nu'_{max}=0.05$*

($\nu'_{max}=0.05$), is shown in Figure 6.6. The difference between the two receivers is impressive, although the difference becomes smaller when the diversity increases (for $Q=3$ and BER $=10^{-3}$ it is 5 dB). This is because the irreducible BER (IBER) which is observed on a DPSK receiver becomes less the larger Q becomes. On the other hand the channel estimation for the coherent BPSK receiver gets poorer with larger Q. This is due to the decreasing signal power per tap. Since channel estimation is performed for each tap separately, the MSE rises.

The effects of varying R on the system performance is shown in Figure 6.7. The BER is plotted against the SNR, for a receiver with $M=8$, and $Q=1$ and 2 diversity taps. For $R\geq 5$ the performance of the slow fading channel (Figure 6.7.a) is poorer than for the fast fading case (Figure 6.7.b). This is

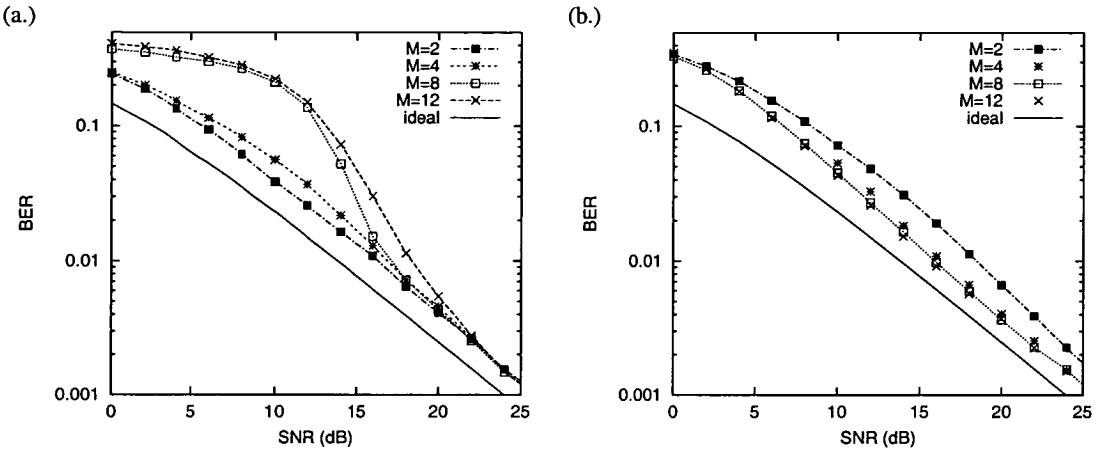


Figure 6.8: *BER vs SNR for different numbers of filter orders M; $R=10$, $Q=1$,*
(a.) slow fading: $\nu'_{max}=0.005$, *(b.) fast fading: $\nu'_{max}=0.05$*

a surprising and unexpected result, as normally channel estimation and the detection of uncoded signals transmitted over a slowly fading channel are considered to be less problematic. However, here, particularly for low SNR and/or slow fading, the receiver with $R \geq 5$ is found to be not robust. These stability problems were identified by Auer *et al* [143, 144] and will be extensively analysed in the next section. In general, for both Figure 6.7 parts (a.) and (b.), it is observed that the lower bound of (5.41) matches the simulated results for high SNR ($\bar{\gamma} > 20$ dB) almost independent of R. For low SNR however ($\bar{\gamma} < 15$ dB), long error bursts and error propagation cause the bound to become loose. Hence the single error assumption for the derivation of the lower bound, described in section 5.2.4, is not valid. For the limiting cases of $R \rightarrow \infty$ the error probability approaches $P_e = 0.5$ which is the expected asymptote, since then the receiver has no phase reference. If the pilot multiplexing rate R becomes smaller ($R < 5$), the theoretical bound becomes tight for all SNR values, as shown in Figure 6.7. This is because, a smaller R breaks up long error bursts and therefore reduces error propagation. However, it is desirable to choose R as large as possible to maximise the spectral efficiency.

The stability problems observed for slow fading are more likely to occur for large filter orders M, this is illustrated in Figure 6.8.a. It is seen that in slow fading conditions, only a receiver with $M > 4$ performs poorly. Lower filter orders, on the other hand, degrade if the Doppler frequency becomes larger, as seen in Figure 6.8.b, because of the reduced ability of the estimator to track the fading fluctuations. The results from Figure 6.8 indicate, that an optimum value M for FIR–PSP, to operate in arbitrary fading conditions does not exist. Rather is there a trade-off dependent on the Doppler frequency ν'_{max} and the SNR, from which the simulation results favour a larger or smaller M. The best compromise between stability for slow fading and phase tracking capability for fast fading appears to be a predictor with $M = 4$ coefficients.

6.3.3 Stability analysis of the 2–state PSP algorithm

In the following, the reasons for the stability problems described above are investigated. Close studies of the problem showed two main reasons. The first reason arises from decision feedback, since the decision variable of one state, $\chi(k) = i$, is dependent upon decisions concerning the previous M samples. For

(a.)

(b.)

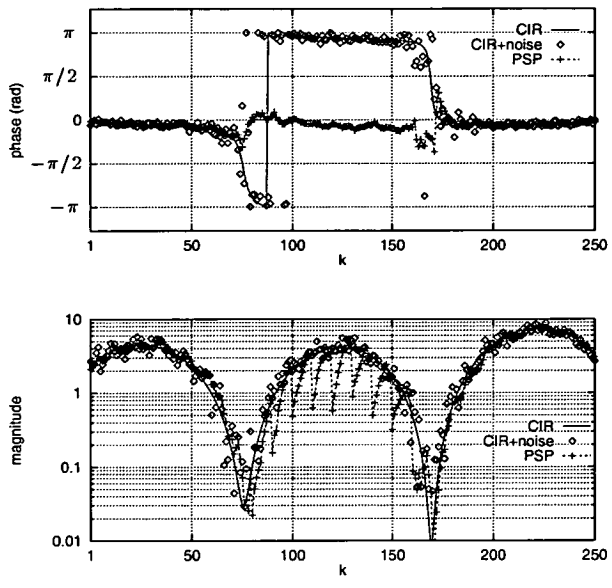


Figure 6.9: *Phase (a.) and magnitude (b.) of FIR–PSP vs time k, compared to the CIR with and without AWGN;* $\nu'_{max}=0.005$, $\bar{\gamma}=10$ *dB*, $M=8$.

$L = 2$ states, the decision upon the survivor, $\widehat{\Lambda}(i,k)$, of that state, however, is made with only one sample time delay. Thus, crucial information may be discarded by applying the VA. Since adjacent samples are more correlated in slow fading this loss of information becomes more important.

The second reason is due to the nature of the predictor coefficients $\{w_m\}$. It has already been illustrated in Figure 6.8 how the filter length M affects the performance. In the following a comparatively long filter ($M = 8$) is employed in order to further examine these stability problems. Furthermore, a flat fading channel with $Q = 1$ will be assumed, so the subscript q to identify the diversity taps can be dropped, i.e. $h(k) = h_1(k)$. Figure 6.9 shows the phase (Figure 6.9.a) and magnitude (Figure 6.9.b) of the estimated CIR, $\widehat{h}(i,k)$, for FIR–PSP in the time domain, compared to the true CIR with and without AWGN. It is seen that the receiver can be locked in a false state, where the channel estimator phase is flipped (shifted 180° relative to the CIR phase) i.e. $\arg\left[\widehat{h}(i,k)\right] = \arg\left[h(k)\right] + \varphi_\epsilon \pm \pi$, where φ_ϵ denotes the prediction error phase. This can be observed in the graph between samples $k \approx [80, 170]$. In this interval the error rate becomes virtually 100%. The receiver is entering the false state after a deep fade and stays locked until the following deep fade, as shown in the graph. During a deep fade, rapid phase changes and a signal to noise ratio up to 20 dB smaller than the average SNR $\bar{\gamma}$ may cause the channel estimator to lose track of the received signal phase. Then the prediction error due to noise becomes large compared to the CIR, $h(k)$, resulting in a poor channel estimates. Subsequently, in a good reception area, the phase estimate of the CIR remains locked in the false state; despite the fact, that the estimated magnitude (Figure 6.9.b) deviates from the actual CIR, particularly if a pilot symbol is detected, seen as negative spikes in Figure 6.9.b. It appears the pilot symbols are not capable of providing a sufficient phase reference to the receiver.

In order to understand the effects which cause these phase slips observed in Figure 6.9, consider the Euclidean distance metric $\Delta(i,i') = \left|\epsilon(i,i',k)\right|^2$ defined in (5.33), where the prediction error is defined

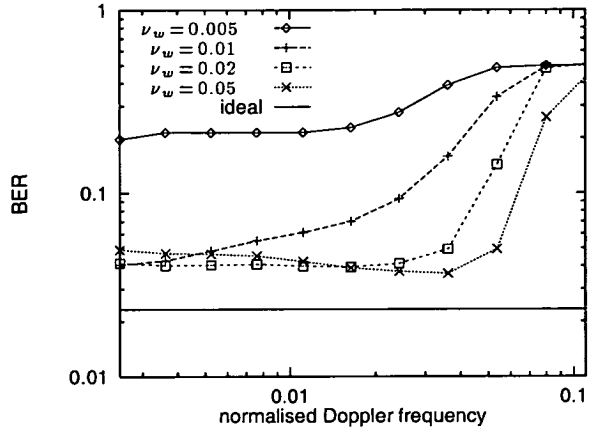


Figure 6.10: *BER vs normalised Doppler frequency ν'_{max}, for filters* **w** *matched to ν'_w and $\bar{\gamma}_w = \bar{\gamma} = 10$ dB; $M = 8$, $Q = 1$, $R = 10$.*

by $\epsilon(i, i', k) = y'(i', k) - \widehat{h}(i, k)$. The estimation error depends on $M + 1$ samples, $\epsilon(i, i', k)$, that is the current sample $d^{(i')}(k)$, sample $d^{(i)}(k-1)$ and the tentative decisions on the previous $M-1$ samples $\{\tilde{d}^{(i)}(k-m)\}_{m=2}^{M}$. Clearly, two sequences which merged at least $M + 1$ samples ago will have identical prediction errors $\epsilon(i, i', k)$ and consequently the same Euclidean distance $\Delta(i, i')$. Furthermore, due to magnitude squared operation in calculating $\Delta(i, i')$, a conjugate sequence has also the same Euclidean distance, that is

$$\begin{aligned} &\text{dist}\{d^{(i')}(k), d^{(i)}(k-1), \tilde{d}^{(i)}(k-2), \cdots, \tilde{d}^{(i)}(k-M)\} \\ &\qquad = \text{dist}\{-d^{(\bar{i}')}(k), -d^{(\bar{i})}(k-1), -\tilde{d}^{(\bar{i})}(k-2), \cdots, -\tilde{d}^{(\bar{i})}(k-M)\} \\ \Rightarrow \quad &\Delta(i, i') = \Delta(\bar{i}, \bar{i}') \end{aligned} \tag{6.6}$$

This observation is analogous to the *phase ambiguity problem* [75]. Recall the definition of the path metric by $\Lambda(\ell, k) = \sum_{(i,i') \in \ell} \Delta(i, i')$. A sequence with a path metric of $\Lambda(\ell, k)$ with $k - 1$ errors will have the same distance metric as a sequence with the path metric $\Lambda(\bar{\ell}, k)$ with a single error, where $\bar{\ell}$ denotes the conjugate sequence of ℓ. Now suppose the receiver is in a false state as depicted in Figure 6.9, then the phase reference of a pilot is rejected as noise (seen as negative spikes in the magnitude plot Figure 6.9.b), and the survivor is the sequence having $R - 1$ data errors per pilot symbol, while the true ML path has previously been discarded by the VA, before the receiver entered the bad state. However, Figure 6.9 does not indicate why this only happens for certain channel parameters, such as slow fading and low SNR.

6.3.3.1 Identifying conditions for instability

To investigate the effect of the nature of the predictor weights $\{w_m\}$ consider the following experiment. The estimation filter **w** is determined by the Wiener-Hopf equation (2.18), dependent on ν'_{max} and $\bar{\gamma}$. It is shown in the following how **w** can degrade the performance of 2–state PSP, although its coefficients satisfy the MMSE criterion. Let a channel estimation filter be matched to the parameters $\bar{\gamma}_w$ and ν'_w,

operating in a scenario with the actual channel parameters $\bar{\gamma}$ and ν'_{max}, and consider the BER performance shown in Figure 6.10. It is seen that a filter **w** matched to low Doppler, $\nu'_w = 0.005$, has very poor performance for all Doppler frequencies, while for filters with $\nu'_w \geq 0.02$, the filter shows no stability problems as long as $\nu'_{max} \leq \nu'_w$. This establishes a dependence of the stability of the 2–state PSP based receiver to the predictor weights **w** [143].

The previous example showed that the stability of the receiver depends critically on the filter weights $\{w_m\}$. The next step is to consider the correlation between filter taps and the CIR $\{\psi_m = w_m^* \phi_{hh}(m)\}$, defined in (6.1). By using $\{\psi_m\}$, a rough indicator for the stability of the receiver is established. Consider a scenario where the receiver has been locked in a false state for at least M samples. Suppose a pilot has been detected at sample $k = k_p$. Since only one state is allowed at time k_p, a sole survivor with final decisions $\{\widehat{d}(k_p - m) = -d(k_p - m)\}_{m=1}^{R-1}$ is obtained, after applying the VA. In other words, assuming that $R > M$, there are $M - 1$ errors and one correctly received pilot, which contribute to the prediction of the CIR, $\widehat{h}(i, k_p + 1)$. Simulations suggested that at reasonably high SNR of $\gamma = 10$ dB, the majority of the survivor paths merge after a time delay of just one sample, at the output of the Viterbi processor. Since the observed phase slips persist over good reception areas, as observed in Figure 6.9.b, this high SNR assumption appears to be justified. Thus, the minimum sequence out of the two candidates at time $k_p + 1$ is likely to be the output of the VA. The correlation between $\widehat{h}(i, k_p + 1)$ and the CIR for that sample $h(k_p + 1)$ from (6.1), is given by $\phi_{\widehat{h}h}(0) = \sum_{m=1}^{M} \psi_m$. For this case ψ_1 accounts for the proportion of pilot symbol on the correlation $\phi_{\widehat{h}h}(0)$, while $\sum_{m=2}^{M} \psi_m$ accounts for the proportion of the corresponding $M - 1$ data symbols. The pilot to data ratio (PDR) is defined by

$$\text{PDR} \triangleq \frac{\psi_1}{\sum_{m=2}^{M} \psi_m} = \frac{w_1^* \phi_{hh}(1)}{\sum_{m=2}^{M} w_m^* \phi_{hh}(m)} \tag{6.7}$$

With this definition, a high PDR implies a strong phase reference with respect to the $M - 1$ data symbols. Qualitatively, for low SNR and/or low fading rates the PDR is rather low, due to the high correlation of $\{\psi_m\}$. Furthermore, increasing the filter order M also results in a degradation of the PDR. Thus, the PDR can be used as an indicator for the stability of the receiver. Figure 6.11 illustrates the impact of pilot and data symbols on the correlation between $\widehat{h}(i, k_p + 1)$ and the CIR, $h(k_p + 1)$, for slow and fast fading. For slow fading, adjacent samples are more correlated, which is manifested in both $\{w_m\}$ and $\phi_{hh}(m)$. Hence ψ_m becomes less dependent on m (compared to the fast fading case, as indicated in Figure 6.11). Therefore past samples have more impact in calculating the metric $\Delta(i, i')$ and the sub-optimality of the receiver, in terms of state dependent decision feedback, becomes more significant. As a result, the ratio between received signal powers of pilot and data symbols decreases. This can cause the receiver to become unstable [143].

$\bar{\gamma}$/ dB \ $\nu'_{max} \cdot 10^3$	2.5	5	10	20
12	0.182	0.281	**0.517**	0.798
18	0.278	**0.511**	0.788	1.05
24	**0.509**	0.785	0.96	1.71

Table 6.2: *pilot to data ratio (PDR) of a FIR filter with $M = 8$ coefficients.*

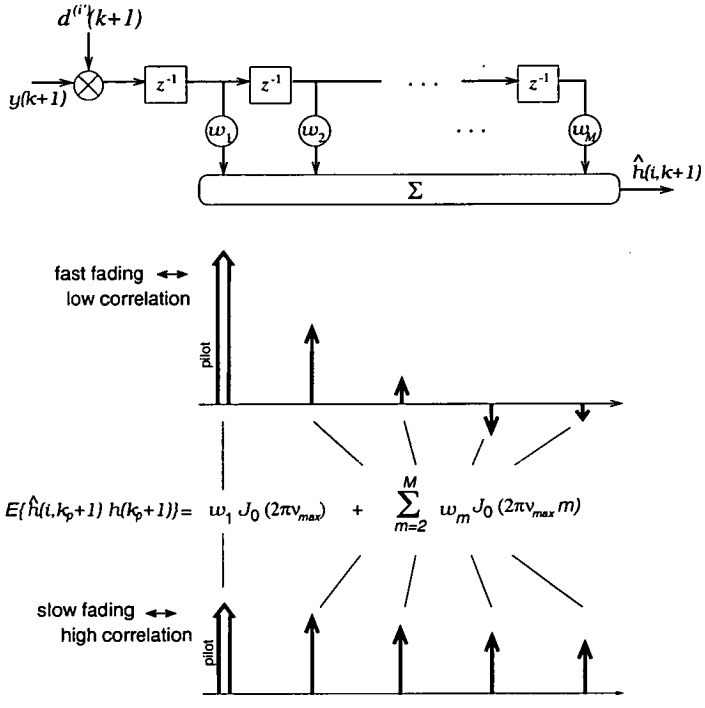


Figure 6.11: *Impact of pilot and data symbols on the correlation between $\widehat{h}(i, k_p+1)$ and the CIR, $h(k_p+1)$, for slow and fast fading.*

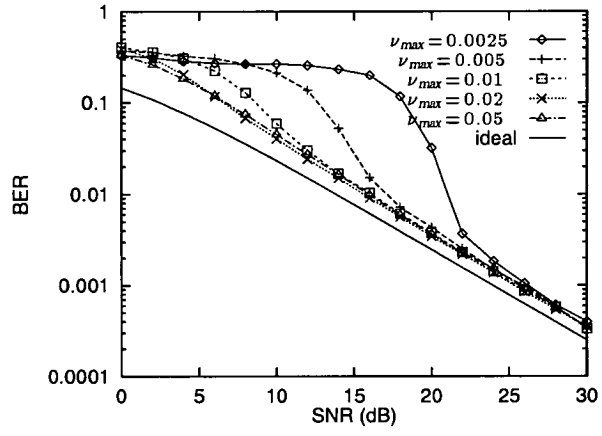


Figure 6.12: *BER vs SNR for various Doppler frequencies ν'_{max}. $M=8$; $Q=1$; $R=10$.*

Table 6.2 shows PDR values for some channel parameters ν'_{max} and $\bar{\gamma}$, for the classical Doppler fading model with ACF $\phi_{hh}(m) = J_0(2\pi\nu'_{max}\, m)$. Figure 6.12 shows the 2–state PSP receiver performance against the SNR, for some values of ν'_{max}. Note, the higher the fading rate ν'_{max}, the less SNR is required for the receiver to be robust. Comparing Table 6.2 with Figure 6.12 shows, that if the FIR-PSP performance shows no sign of instability, the PDR is over a certain threshold $\text{PDR}_{th} \approx 0.5$, for $M=8$ and $R=10$. The framed entries in Table 6.2 identify the threshold PDR_{th}. Thus, for a given filter order M and pilot multiplexing rate R, the FIR-PSP is always stable if $\text{PDR} \geq \text{PDR}_{th}$. So the lower the Doppler frequency ν'_{max} the higher the required SNR, $\bar{\gamma}$, for stability.

6.3.3.2 Analysis using the Gilbert-Elliot channel (GEC)

In this section the instability of the 2–state PSP is examined using *hidden Markov models*, so called because the model is characterised by a non-observable Markov chain [145]. In communications a hidden Markov model (HMM) can be employed to model burst errors in discrete-time channels, called finite state Markov channels. In such a channel, a binary symmetric channel (BSC)[4] [7] with a given error probability is associated with each state of the HMM. The study of finite state Markov channels emerged from early work of Gilbert [146] and Elliott [147]. They studied a 2–state Markov channel known as the Gilbert–Elliott channel (GEC). The error process is generated according to the following probabilistic mechanism: when the Markov Model is in state $\chi(k) = 0$, the good state, the channel corrupts the transmitted bit with the error probability $P_e(\chi(k) = 0) = P_{e0}$. Otherwise, when the channel is in state 1, the burst state, the channel produces an error with higher probability $P_e(\chi(k) = 1) = P_{e1}$.

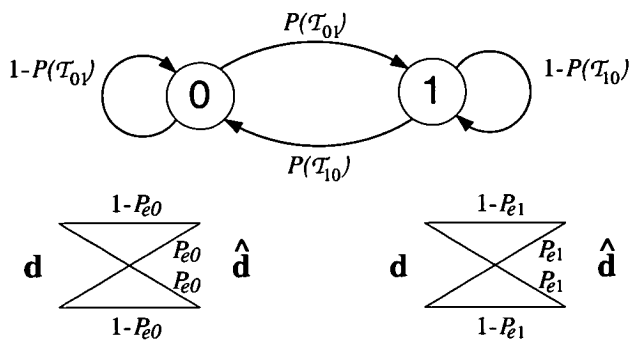


Figure 6.13: *The Gilbert–Elliott channel (GEC).*

The GEC is illustrated in Figure 6.13, there the 2–state HMM is depicted in the upper part of Figure 6.13, and the channels for the good and burst state are shown beneath. After every channel transmission, the chain makes one transition of state according to the transition probability matrix $\mathbf{P}$. Let a transition from state 0 to 1 and the corresponding transition probability be denoted by $\mathcal{T}_{01}$ and $P(\mathcal{T}_{01})$. Since a constant Markov process has the property of stationary transitions, the transitions probability is independent of the time index k. With these definitions, the transition matrix can be written as

$$\mathbf{P} \triangleq \begin{bmatrix} 1 - P(\mathcal{T}_{01}) & P(\mathcal{T}_{01}) \\ P(\mathcal{T}_{10}) & 1 - P(\mathcal{T}_{10}) \end{bmatrix} \tag{6.8}$$

Furthermore, define the probability of the HMM being in state 0 or 1, as p_0 and p_1, respectively. Note that p_i and $P(\mathcal{T}_{ij})$ must satisfy the equilibrium condition which states that for any given states $i, j \in \{0, 1\}$, the incoming flow and outcoming flow must be equal [148], cast in a set of linear equations

$$\mathbf{p}\,\mathbf{P} = \mathbf{p}\,; \quad \text{with} \quad \mathbf{p} \triangleq [p_0, p_1]$$

[4]The channel in this context is defined in a more abstract way, than the fading channel which has been used so far. The channel considered in association with Markov models is a discrete-input, discrete-output channel, similar to the information theoretic channel discussed in section 1.2.2. For binary modulation and the equal probability of error for each bit, the BSC is obtained.

The state probabilities of the HMM and the error probabilities of the binary symmetric channel are related by

$$\bar{P}_{GEC} = p_0 P_{e0} + p_1 P_{e1}$$

In the analysis it is assumed that if the probability of error in the good state is $P_{e0} = P_e$, the burst-error probability is $P_{e1} = 1 - P_e$, according to the discussion in the previous section. The GEC is then uniquely defined by $\mathbf{p}$, $\mathbf{P}$ and P_e, giving the average probability of error for the GEC

$$\bar{P}_{GEC} = \frac{P_e P(\mathcal{T}_{10}) + (1 - P_e) \cdot P(\mathcal{T}_{01})}{P(\mathcal{T}_{10}) + P(\mathcal{T}_{01})} \tag{6.9}$$

For the case $P(\mathcal{T}_{01}) = 0$, the GEC approaches the performance of the good state, thus $\bar{P}_{GEC} = P_e$. On the other hand, if $P(\mathcal{T}_{01}) = P(\mathcal{T}_{10})$, the GEC has a probability of $\bar{P}_{GEC} = 50\%$, which can easily be verified in (6.9). This is the case for a binary signal without a phase reference. Generally speaking, if $P(\mathcal{T}_{10})$ is only marginally larger than $P(\mathcal{T}_{01})$, the error probability of the GEC, $\bar{P}_{GEC}$, will be much higher than P_e. This is the case when the pilots cannot impose a sufficient phase reference on the receiver. On the other hand, the bigger the difference between $P(\mathcal{T}_{10})$ and $P(\mathcal{T}_{01})$ the more robust is the system and the better the performance. Thus, the GEC can be used to model the stability problems observed for 2–state PSP with appropriately chosen transition probabilities $P(\mathcal{T}_{10})$ and $P(\mathcal{T}_{01})$, which will be specified in the following paragraph.

The GEC described above, can be extended to a finite state Markov channel with more than two states, following e.g. [149]. This may enable modelling the PSP based receiver having $L = 2^D$ states, with a L^{th} order finite state Markov channel. The generalisation of the GEC to a N–state Markov channel is the Fritchman channel [150], having N_g good states and $N - N_g$ burst states. However, estimating the state and transition probabilities, $\mathbf{p}$ and $\mathbf{P}$, becomes more complicated. An efficient algorithm to estimate $\mathbf{p}$ and $\mathbf{P}$ using a given observation sequence was developed in [151]. This algorithm was used to model error burst in a Rician fading channel in [152]. It should be noted at this point, that modelling 2–state PSP receiver output with a 2–state HMM is not a sufficient model. The transition probabilities $P(\mathcal{T}_{ij})$ for $i, j \in \{0, 1\}$ clearly depend on the position of the detected bit relative to the nearest pilot. So, to precisely model the 2–state PSP, a higher order HMM is necessary. However, it will be apparent later on that it is possible to develop performance bounds with the GEC, which in many cases model the stability problems of the receiver well. For this reasons and for the sake of simplicity, for the analysis of 2–state PSP only the GEC will be considered further on.

To describe the 2–state PSP receiver with a HMM we follow a different approach from [151, 152] to determine $\mathbf{P}$: that is Barrett's formula to determine the pairwise probability of an error event $\mathcal{E}$ from (5.41) is used, derived for the recursive MLSD in section 5.2.4, to evaluate $P(\mathcal{T}_{ij})$ of the transition matrix $\mathbf{P}$. Furthermore, according to the discussion in section 5.1.5, the calculation of the pairwise error probability is itself only an approximation of the actual error probability. Thus, to precisely model the receiver by a HMM appears difficult using an analytically derived $P(\mathcal{T}_{ij})$. On the other hand, results obtained for the GEC were encouraging; that the stability problems of the receiver are modelled rather

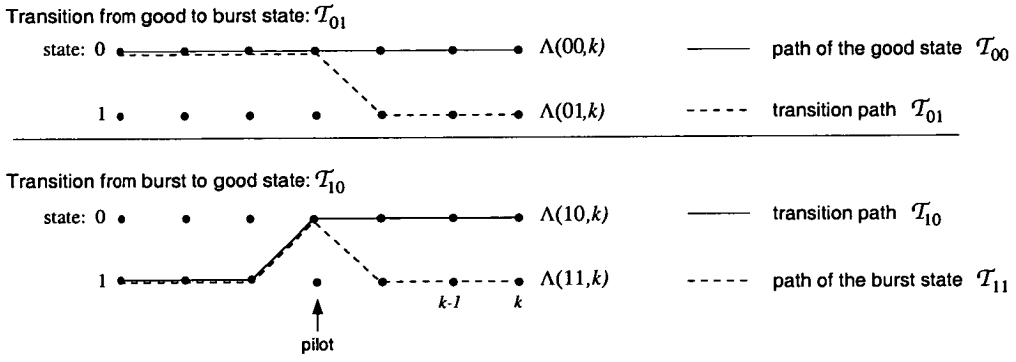


Figure 6.14: *Trellis diagram of transitions between the good and bad state of the GEC.*

well.

To employ the GEC to the given receiver, some idealised assumptions must be made. The error probability of the good state is approximated by the lower bound, $P_e = P_e(1)$. It is assumed that the transitions from the good to the burst state and vice versa, are dominated by two events, which exclusively account for the transition probabilities $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$. The corresponding transition sequences are depicted in Figure 6.14. The first event, $\mathcal{T}_{01}$, accounts for the transition from the good to the burst state, while the second, $\mathcal{T}_{10}$, accounts for the complementary transition. The difference to the definition of the error event $\mathcal{E}$ in section 5.2.4 is that the two considered sequences end up in different states. For the first event, $\mathcal{T}_{01}$ according to Figure 6.14, the sequence $\ell = 01$ associated with $\Lambda(01)$ is defined by:

$$\mathcal{T}_{01}: \quad \mathbf{d}^{(01)} = \{\cdots d^{(00)}(k-M), \cdots, d^{(00)}(k-1), -d^{(00)}(k), \cdots, -d^{(00)}(k+M)\cdots\}$$

where $\mathbf{d}^{(00)}$ denotes the all one sequence, which is assumed to be the transmitted sequence. The probability of a transition from 0 to 1, $P(\mathcal{T}_{01})$, equals the probability that $\Lambda(01) < \Lambda(00)$, or accordingly $\Delta\Lambda(\mathcal{T}_{01}) = \Lambda(01) - \Lambda(00)$ being smaller than zero. The log-likelihood ratio $\Delta\Lambda(\mathcal{T}_{01})$ is essentially non-zero within the range $[k-M,\ k+M]$, since transitions outside this interval have the same Euclidean distance. Assuming that a channel predictor of order M is employed, the branch metrics computed outside the above interval are not dependent on the transition from 0 to 1, and therefore (6.6) applies.

To specify the transition from state 1 to 0, $\mathcal{T}_{10}$ from Figure 6.14, the sequence which enters the good state from the burst state, is defined by

$$\mathcal{T}_{10}: \quad \mathbf{d}^{(10)} = \{\cdots - d^{(00)}(k_p-M), \cdots, -d^{(00)}(k_p-1), d^{(00)}(k_p), d^{(00)}(k_p+1), \cdots, d^{(00)}(k_p+M)\cdots\}$$

It is assumed that the event $\mathcal{T}_{01}$ is triggered by a pilot at time $k = k_p$. Furthermore, it is assumed that the transmitted sequence $\mathbf{d} = \mathbf{d}^{(00)}$, was already discarded earlier on. Instead, suppose the survivor at sample $k = k_p$ is the path $\mathbf{d}^{(11)}$, associated with the decision variable $\Lambda(11)$, which is the same as the transmitted sequence $\mathbf{d}^{(00)}$ if a pilot symbol is detected ($k = k_p$), but different for all other symbols. So, $\Lambda(11)$ specifies the path of transition $\mathcal{T}_{11}$, that is the path which stays in the burst state after reception of

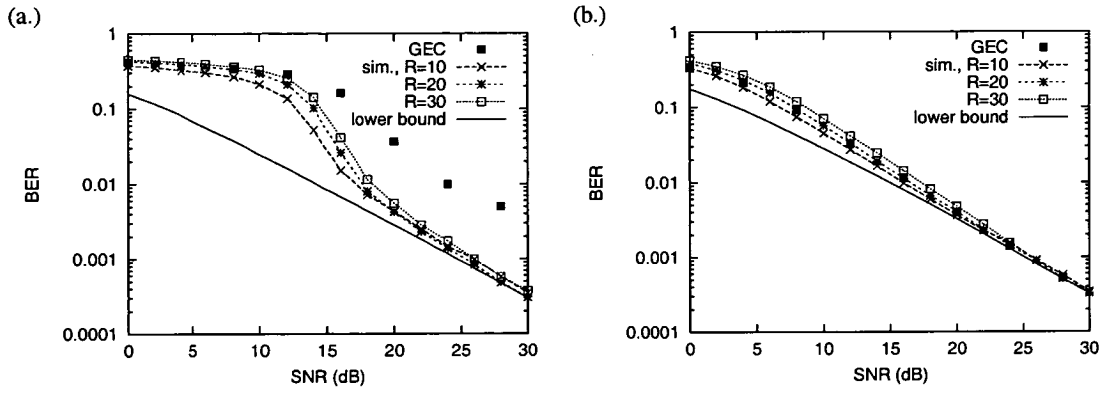


Figure 6.15: *BER vs SNR for the GEC compared to simulations of 2–state PSP with different R; $M=8$. (a.) slow fading: $\nu'_{max}=0.005$; (b.) fast fading: $\nu'_{max}=0.05$.*

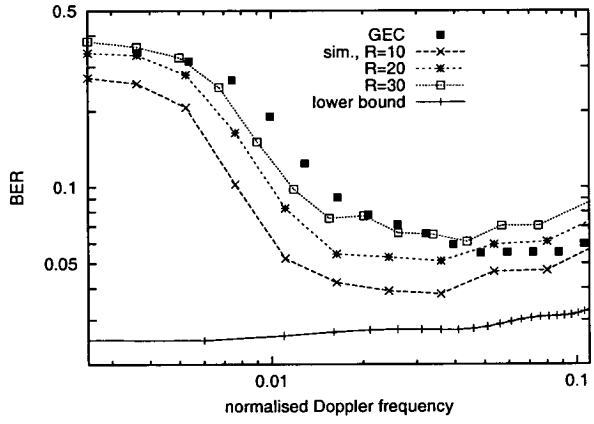


Figure 6.16: *BER vs ν'_{max} for the GEC compared to simulations of 2–state PSP with different R; $M=8$, $\bar{\gamma}=10$ dB, $Q=1$.*

the pilot, defined by

$$\mathcal{T}_{11}: \quad \mathbf{d}^{(11)} = \{\cdots - d^{(00)}(k_p-M), \cdots, -d^{(00)}(k_p-1), d^{(00)}(k_p), -d^{(00)}(k_p+1), \cdots, -d^{(00)}(k_p+M)\cdots\}$$

With the assumption that $R > M$ the log-likelihood ratio $\Delta\Lambda(\mathcal{T}_{10}) = \Lambda(10) - \Lambda(11)$ is also only non-zero within the range $[k_p - M,\ k_p + M]$. It is the definition of the path of $\Lambda(11)$ which imposes the presence of the pilot at time k_p.

The probability of the event $\mathcal{T}_{ij}$ can be approximated by the pairwise probability that $\Delta\Lambda(\mathcal{T}_{ij})$ is smaller than zero. Following the steps from section 5.2.4 the transition probabilities $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$ can be obtained by evaluating Barrett's formula in (5.41) and substituting $\mathcal{T}_{01}$ or $\mathcal{T}_{10}$ with $\mathcal{E}$. Some numerical results for $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$ are presented in Appendix B.1.2.

Now the GEC is fully specified by $P(\mathcal{T}_{01})$, $P(\mathcal{T}_{10})$ and $P_e(1)$ and the average error probability of the GEC, $\bar{P}_{GEC}$, can be calculated using (6.9). Numerical results for the GEC, compared to simulations of 2–state PSP with different R, are shown in the Figures 6.15–6.17. The graphs for the transition probabilities $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$, used to calculate the predictions shown in Figures 6.15–6.17, are shown in Figures B.4 and B.5 in Appendix B.1.2. In Figure 6.15 the (a.) slow and (b.) fast fading performance is

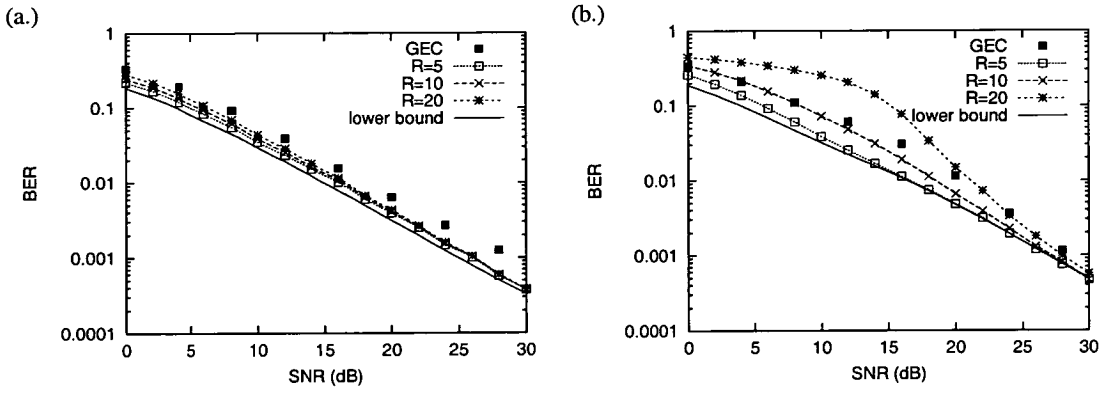


Figure 6.17: *BER vs SNR for the GEC compared to simulations of 2–state PSP with different R; $M = 2$.*
(a.) slow fading: $\nu'_{max} = 0.005$; *(b.) fast fading: $\nu'_{max} = 0.05$.*

plotted against the SNR. Considering slow fading in part (a.), the instability of 2–state PSP is accurately modelled by the GEC for low SNR ($\bar{\gamma} \leq 10\,\text{dB}$), while the GEC is too pessimistic for high SNR. For high SNR, the assumption that the transitions between the states 0 and 1 are caused by only two dominant events is less valid. Therefore, in order to model the low Doppler and high SNR scenario more precisely a higher order HMM may be beneficial. For fast fading (Figure 6.15.b) on the other hand, the GEC matches the simulation results for all SNR, and the GEC turns out to be a better approximation than the lower bound $P_e(1)$ (also shown in the graph). In Figure 6.16 where the BER is plotted against ν'_{max}, the GEC appears a good model for this receiver.

In Figure 6.17 results for the GEC with $M = 2$ are shown, all other parameters correspond to Figure 6.15. The GEC correctly predicts that a receiver with $M = 2$ is more robust for slow fading (Figure 6.17.a). Figure 6.17.b illustrates the major limitation of the GEC: if the receiver performance depends significantly on R, the GEC is not a good approximation of the receiver.

Note, no assumptions about R have been made in the derivation of the GEC parameters. This implies that the GEC may only be expected to work well for scenarios where the dependence of R on the performance is not significant. This limitation may be overcome by employing a higher order HMM, which also accounts for R. This could be achieved by constructing a HMM and then trying to approximate the transition matrix $\mathbf{P}$ from (6.8) analytically, similar of the approach for the 2–state case. On the other hand, estimating the transition probabilities based on the observation sequence $\hat{\mathbf{d}}$ as suggested in e.g. [151], combined with some extra side information about the receiver may model the PSP based receiver more accurately. Note, a higher order model may also be used to analyse PSP based on a $L = 2^D$ state trellis.

Thus, Markov models may be used to deliver tighter performance bounds for receivers operating in fading channels. However, given the simplicity of the GEC, the predicted results match Monte Carlo simulations rather well. The GEC offers further insight in understanding error propagation mechanisms, since the dominating sources which may cause instability can be identified and modelled by a HMM.

6.3.4 Results for IIR estimation filter

In the following a PSP receiver based on a 2–state trellis employing an α–tracker (IIR–PSP) is considered. Except for the different channel estimation filter, the receiver and simulation set-up specified in section 6.3.1 was used, which is equivalent to the simulation parameters of previous section. Simulation results presented in Figures 6.18–6.20 are carried out on a frequency-flat fading channel ($Q=1$).

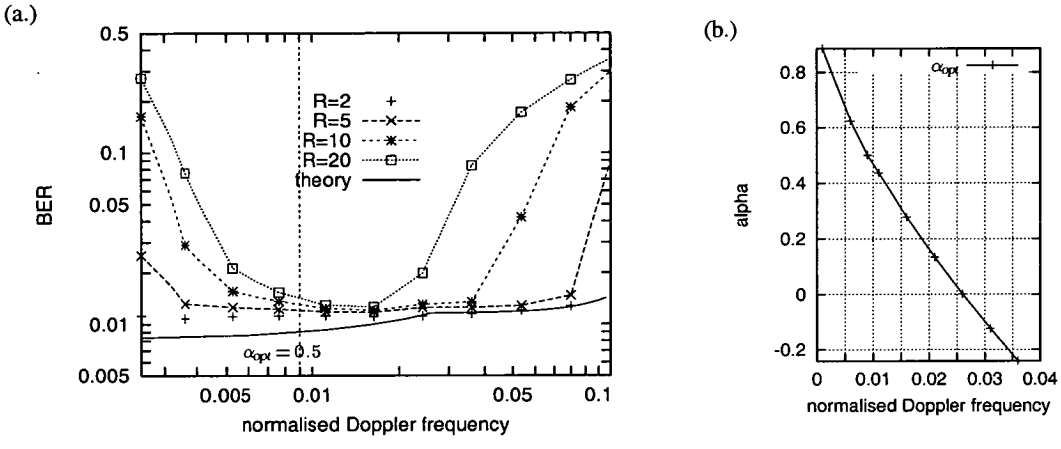


Figure 6.18: *(a.) BER vs ν'_{max} with $\alpha = \alpha_{opt}$, for different R.*
(b.) $\alpha = \alpha_{opt}$ as a function of ν'_{max} obtained through analysis.
$Q=1$; $\bar{\gamma}=15$ dB.

The BER performance of a receiver using an estimation filter with $\alpha = \alpha_{opt}$ chosen according to (2.22) is shown in Figure 6.18, which plots the BER against ν'_{max}, for different pilot to data ratios R. It can be observed that the performance is highly dependent on R. For small R, the receiver performance is close to the lower bound for arbitrary fading rates. When R increases, however, the performance is poor for slow and fast fading, while it is relatively unaffected for medium fading rates $\nu'_{max} \approx 0.01$. Only for small R is the theoretical lower bound of (5.41) tight (label "theory"). For slow fading and $R \geq 5$, similar stability problems are observed as previously for FIR–PSP. The cut-off frequency of a filter with α_{opt} is not responsible for the poor slow fading performance, as it is still much larger than ν'_{max}. Otherwise, the performance of the lower bound should also degrade. The reason for this frequency dependent instability is similar to the discussion of the FIR–PSP. The filter parameter α_{opt}, is a function of ν'_{max} and $\bar{\gamma}$, such that the noise error equals the phase lag error [28], as discussed in section 2.1.5. The approximation of α_{opt} from (2.22) is drawn in Figure 6.18.b. Note, that the lower the fading rate, the larger α_{opt} becomes. So for slow fading, if α_{opt} becomes larger than a certain value, the influence of a transmitted pilot in order to estimate the CIR is not sufficient to track the phase of the fading fluctuations. This is in accordance with the discussion in section 6.3.3.1, if the pilot–to–data ratio (PDR) defined in (6.7) is over a certain threshold, the receiver becomes unstable.

As a result of the previous graph, α_{opt} of (2.22) is not optimum to minimise the probability of error; this was already pointed out in section 5.2.4, where the lower bound of VA-MLSD using a 1st order IIR filter was studied. So, it is useful to investigate how to determine a filter constant which yields more satisfactory results. Thus, the effects of the choice of the filter constant α, are shown in Figure 6.19. The BER is plotted against ν'_{max} for different values of α. Note, for the simulations in this graph, α

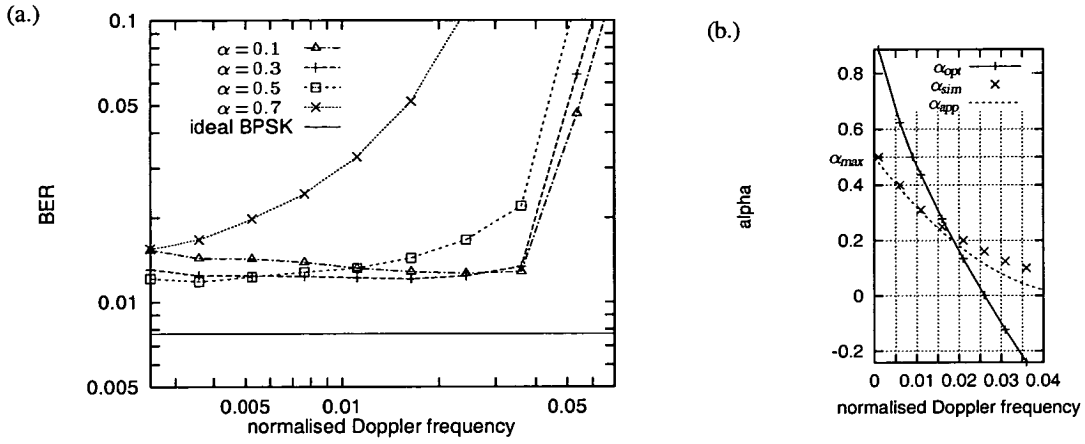


Figure 6.19: *(a.) BER vs ν'_{max}, for different α; $Q=1$, $R=10$, $\bar{\gamma}=15$ dB. (b.) α_{opt} obtained through analysis and simulations.*

does not adapt to the changing fading conditions, but stays constant. Results shown in Figure 6.19 are similar to those found for a mismatched FIR filter in Figure 6.10. This implies that similar effects are responsible for the stability problems. Comparison of Figure 6.18 with the graph for various α in Figure 6.19 indicates that for low Doppler, the poor performance is because α_{opt} is too large. It is seen in Figure 6.19, that for $\alpha > 0.5$ the receiver performance is poor even for low Doppler. Thus, the performance can be improved simply by imposing an upper bound α_{max} on α_{opt}. Simulations have shown that a value of $\alpha_{max} \approx 0.5$ is usually a good approximation to overcome the degrading effects at low Doppler.

For fast fading on the other hand, the performance cannot be improved by means of adopting α_{opt}. This is due to a *run away* phenomena of the receiver; the α–tracker is unable to track the carrier phase, which becomes worse with increasing R. Since for fast fading α_{opt} in Figure 6.18 approaches zero, the performance is similar to $\alpha = 0.1$ in Figure 6.19 (with $R = 10$). That small α perform better for high Doppler is an expected result, bearing in mind that with decreasing α the phase-lag error, which is dominant for high Doppler also decreases.

The filter constant which minimises the BER can also be determined through simulations. In order to find this parameter, α_{sim} from Figure 6.19.a, that α is chosen, associated with the graph, which has the minimum BER for a certain Doppler frequency ν'_{max}. These parameters $\alpha = \alpha_{sim}$ are drawn in Figure 6.19.b. It is seen in Figure 6.19.b, that the filter coefficient which yielded the best results in the simulations α_{sim}, differs significantly from the analytically derived α_{opt}. The parameter α_{sim} as a function of ν'_{max} can be approximated by the quadratic function

$$\alpha_{app} = \begin{cases} a\nu'_0\,(\nu'_{max} - \nu'_0)^2\,; & \nu'_{max} \leq \nu'_0 \\ 0\,; & \text{elsewhere} \end{cases} \tag{6.10}$$

where a and ν'_0 are constants chosen appropriately to approximate the simulation results for α_{sim}. The pole ν'_0 represents the Doppler frequency from which $\alpha_{app}=0$, i.e. for ν'_{max} larger than ν'_0 the receiver does not filter the received signal any longer, it simply takes its previous sample $y_q(k\text{–}1)$. Through comparison

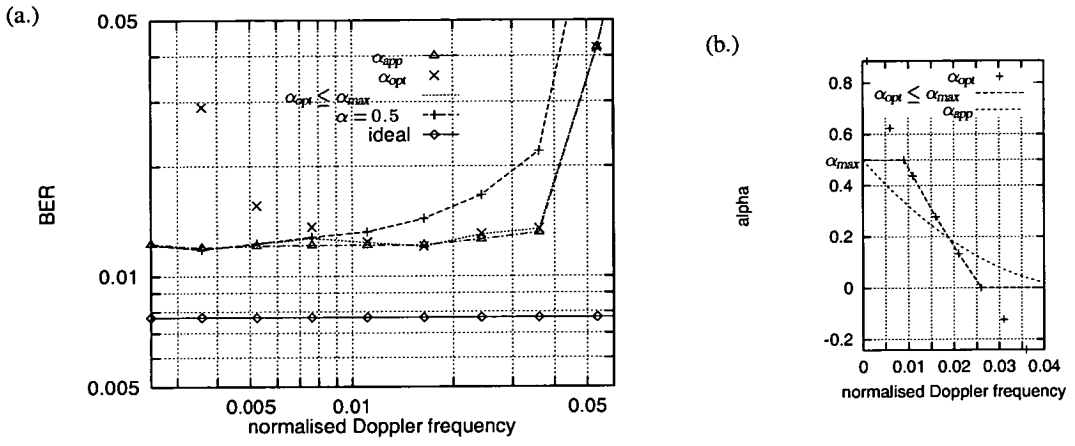


Figure 6.20: *(a.) BER vs ν'_{max} for different filter constants, α_{opt}, α_{app} and $\alpha = 0.5$.*
(b.) Different filter constants as a function of ν'_{max}.
$Q = 1$; $R = 10$; $\bar{\gamma} = 15$ dB.

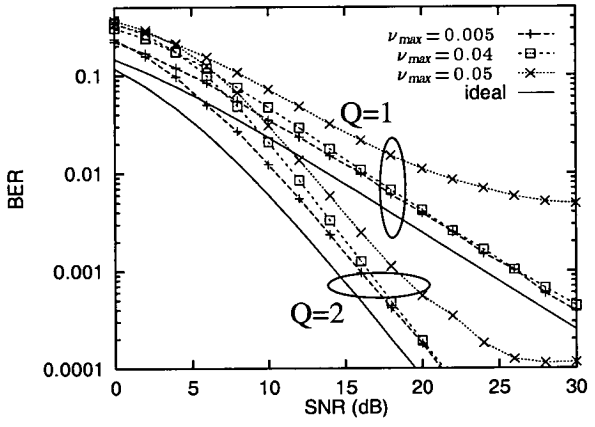


Figure 6.21: *BER vs SNR of a filter with α_{app}, for various Doppler frequencies ν'_{max}. $R = 10$.*

with Figure 6.19.a the constants $\nu'_0 = 0.05$ and $a = 10$ are obtained, yielding $\alpha_{app} = (20\,\nu'_{max} - 1)^2/2$, for $\bar{\gamma} = 15$ dB. Note, this approximation is only valid for a given SNR.

Figure 6.20 (part a.) shows the performance obtained from different realisations of the filter constant α against ν'_{max}. In Figure 6.20 (part b.) the filter coefficients α_{opt} and α_{app} are drawn as a function of ν'_{max}. Also shown in the graph is the case where α_{opt} is upper and lower bounded by $0 \leq \alpha_{opt} \leq 0.5$. When α_{opt} is upper bounded by $\alpha_{max} = 0.5$ the performance is virtually identical to the filter with α_{app}, although α_{opt} can differ from α_{app} quite significantly, as seen in Figure 6.20.b. It can therefore be concluded that unless $\alpha > \alpha_{max}$ the system performance is not critically dependent on the choice of α.

In Figure 6.21 results are shown for IIR–PSP with the filter coefficient α_{app} chosen according to (6.10). It can be observed that the receiver performance is excellent for slow fading. On the other hand, IIR–PSP degrades significantly for Doppler frequencies $\nu_{max} > 0.04$. The observed irreducible bit error rate (IBER) is due to the induced phase lag of the IIR filter, which is essentially a low-pass filter and can be reduced by introducing diversity to the system. The IIR–PSP is therefore not suitable for fast fading. Although for $Q = 2$ the IBER is much lower, the fast fading performance is still significantly

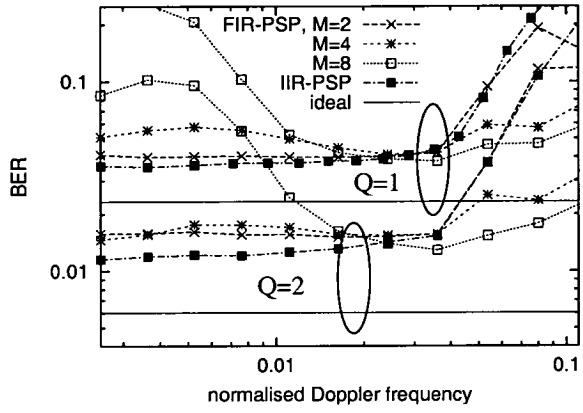


Figure 6.22: *BER vs ν'_{max}, for FIR–PSP and IIR–PSP. $R=10$, $\bar{\gamma}=10$ dB.*

poorer compared to the FIR–PSP from Figure 6.7.b. The fast fading performance can only be improved by reducing R (see Figure 6.18).

Comparison: PSP with FIR and IIR filtering Figure 6.22 shows the BER against ν'_{max}, for various realisations of PSP based on a 2–state trellis. It is seen that IIR-PSP[5] is superior in slow fading conditions, while the FIR-PSP with long predictor ($M=8$) shows better performance for high Doppler.

To conclude this section, the restrictions of pilot aided plus decision directed channel estimation have been studied extensively, with the constraint of a 2–state trellis representation of the receiver. Stability problems of the considered receiver structures were investigated and analysed, by means of Monte Carlo simulations and theoretically by a Markov model. Conditions which indicate instability were identified, which were shown to be critically dependent on the Doppler frequency ν'_{max}, the filter order M and the pilot multiplexing rate R. Based on the knowledge of this parameters, stability problems can be anticipated on a run time basis. Furthermore, robust receivers applicable to both slow and fast fading channels, consisting of a hybrid filter will be addressed in section 6.5.

6.4 PSP based on an expanded trellis structure

The discussion in the previous section was restricted to PSP based on a 2–state trellis. Now other state reduction techniques discussed in section 6.1 are examined. Compared to 2–state PSP, the receiver now operates on an expanded trellis structure with $L = 2^D$ states. Apart from that, all other assumptions concerning the channel model or the simulation setup correspond to section 6.3.2. For the numerical results of the lower bound, equation (5.41) in section 5.2.4 has been evaluated. Generally, curves labelled "ideal" identify the case where the CIR is known *a priori* according to (4.5), and "theory" is the label for the lower bound of (5.41). Binary modulation (BPSK with $A_m=2$) was used for all results presented in

[5] with the modification that $\alpha \leq 0.5$.

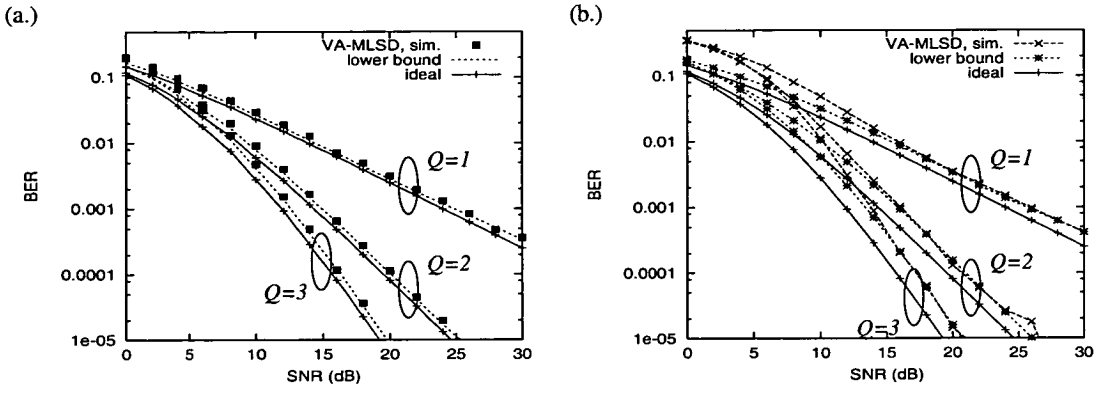


Figure 6.23: *BER vs SNR for VA-MLSD with $Q=1, 2$ and 3 diversity taps; $D=M=4$, $R=10$.*
(a.) slow fading: $\nu'_{max}=0.005$, *(b.) fast fading: $\nu'_{max}=0.05$*

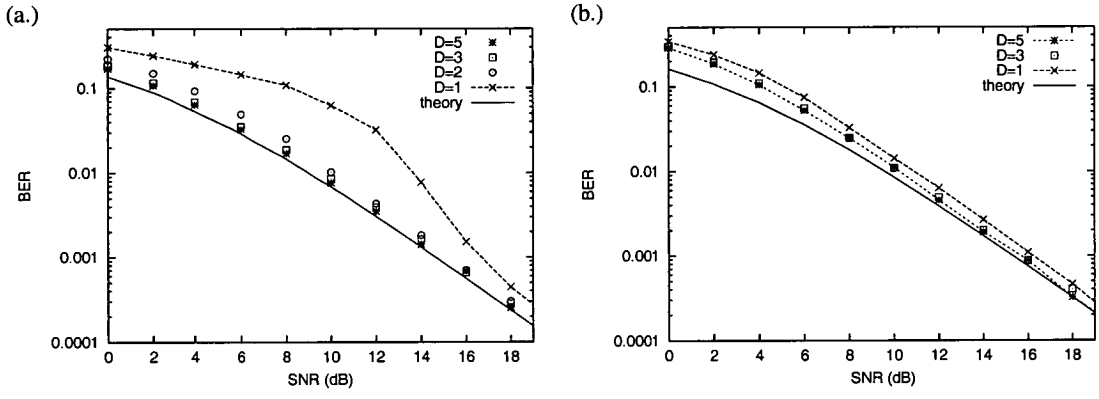


Figure 6.24: *BER vs SNR for different numbers of D; $M=8$, $R=10$, $Q=2$,*
(a.) slow fading: $\nu'_{max}=0.005$, *(b.) fast fading: $\nu'_{max}=0.05$*

this section.

6.4.1 FIR estimation filter

Results for VA-MLSD with a full trellis described in section 5.2.3 with $D = M = 4$, are shown in Figure 6.23. The BER is plotted against the SNR, $\bar{\gamma}$, for a diversity receiver with $Q = 1, 2$ and 3 taps. The BER performance of a system in a slow fading channel, with a normalised Doppler frequency of $\nu'_{max} = \nu_{max}T_s = 0.005$, is shown in part (a.). It is seen that the simulations match the lower bound. Thus, it can be postulated that VA-MLSD without state dependent decision feedback is robust for slow fading. The BER performance of a system in a fast fading channel, with a normalised Doppler frequency of $\nu'_{max} = \nu_{max}T_s = 0.05$, is shown in part (b.). It is seen that the simulation results ($\bar{\gamma} > 15$ dB) match the lower bound of (5.41) for high SNR, independently of the number of diversity taps Q. For low SNR however, long error bursts and error propagation cause the bound to become loose. During an error event, the channel estimation process is corrupted, leading to an inaccurate CIR estimate, thus subsequent symbols are more likely to be detected incorrectly. Hence the single error assumption for the derivation of the lower bound is not valid. It can be observed that adding diversity significantly improves the system performance, while the relative behaviour of the algorithms is independent of Q.

In the following, the number of diversity taps is set to $Q=2$. The effect of varying D on the BER as

a function of the SNR and the normalised Doppler frequency ν'_{max} is shown in Figures 6.24–6.25. The performance for various D as a function of the SNR, is shown in Figure 6.24,[6] for a slow (Figure 6.24.a) and a fast fading channel (Figure 6.24.b). For fast fading the difference of PSP in terms of D never exceeds 1 dB. The PSP performance with $D \leq 2$ for the slow fading case (Figure 6.24.b), is seen to break down for low SNRs, an effect which was extensively analysed in section 6.3.3. Provided that $D \geq 3$, the simulated error probability closely matches the lower bound.
In Figure 6.25, the BER is shown against ν'_{max}, again for different numbers of D. It is seen that receivers with low D have stability problems dependent on ν'_{max}.

In Figures 6.26 and 6.27 simulation results of the BER against the pilot insertion rate R are shown for various trellis memory length D, and filter orders M. For slow fading (Figure 6.26), the short predictor ($M = 4$ in part (a.)) performance is better for short delays D. The long filter ($M = 8$ in part (b.)) the performance is marginally better. The performance of 2–state PSP ($D = 1$) is poor even for low R, which corresponds to the poor slow fading performance observed in section 6.3.2.
For fast fading (Figure 6.27), increasing the filter order to $M = 8$, results in the receiver becoming less sensitive to variations of R. Clearly the BER degrades monotonically with R. For $M = 8$ the BER is considerably lowered by increasing D from 1 to 2. However, if $D \geq 2$, there is little performance gain to be obtained. For $M = 4$ on the other hand, the BER degradation is more severe for increasing R. The reason why shorter predictors are more affected by increasing R is due to their poorer phase tracking capability, leading to run-away effects of the estimated CIR phase. A relatively short predictor is more likely to suffer a phase slip and the receiver stays locked in a burst state, producing errors with a probability of virtually 100%, at least until the reception of the next pilot. Furthermore, this penalty can only be slightly reduced by expanding the trellis, i.e. setting $D > 1$.
An interesting result in this context is, that expanding the trellis is more beneficial for slow fading than for fast fading. A possible explanation for this observation is described in the following. For 2–state PSP the path associated with the final decision may not be the ML path, since the receiver is only sub-optimum. However, the ML path may not be the actual transmitted sequence. For fast fading, due to poorer phase tracking of the predictor, this is more likely. Thus, even if the ML path was correctly selected by the receiver, the performance of the receiver may still be poor, due to run-away effects of the estimated CIR phase. For slow fading it is easier to track the carrier phase, so run-away effects are less severe. Hence, provided the ML path is chosen correctly, which is more likely for $D > 1$, the performance is close to the theoretical lower bound.

6.4.2 IIR estimation filter

The expanded trellis MLSD receiver employing an IIR estimation filter is considered in the following graphs. The filter constant α_{opt} was calculated using (2.22). The BER performance against ν'_{max} for different numbers of D is shown in Figure 6.28. The slow fading performance ($\nu'_{max} < 0.02$) is only close to the lower bound for $D = 5$ and is seen to be similar to the receiver employing a FIR filter

[6] Due to the exponential growth of the trellis with respect to D, values larger than 5 result in a vast computational complexity, while the performance is only marginally improved. Thus graphs with $D > 5$ have been omitted.

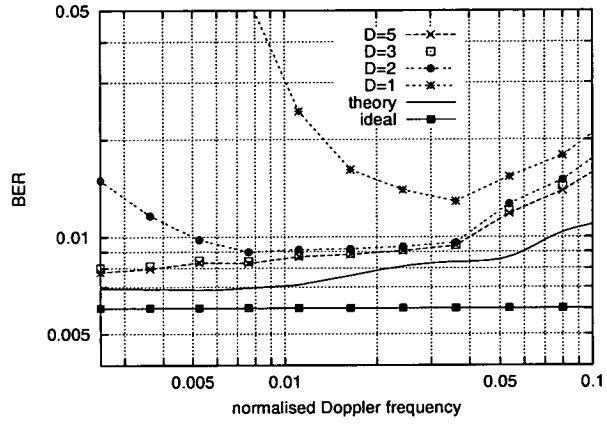


Figure 6.25: *BER vs ν'_{max} for different numbers of D. $M=8$, $R=10$, $Q=2$, $\bar{\gamma}=10$ dB.*

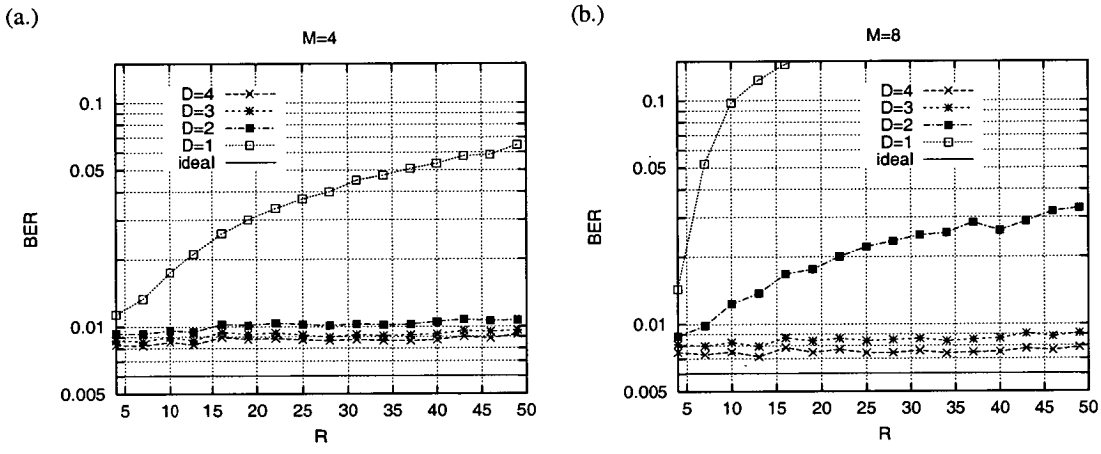


Figure 6.26: *BER vs pilot insertion rate R for different numbers of D.*
Slow fading, $\nu'_{max}=0.005$, $Q=2$, $\bar{\gamma}=10$ dB; (a.) $M=4$ and (b.) $M=8$.

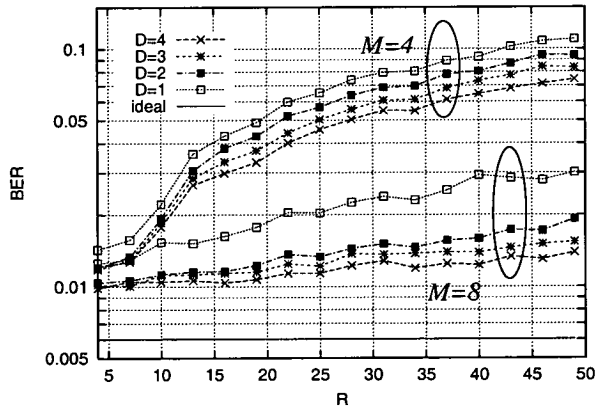


Figure 6.27: *BER vs pilot insertion rate R for different numbers of D.*
Fast fading, $\nu'_{max}=0.05$, $M=\{4,8\}$, $Q=2$, $\bar{\gamma}=10$ dB.

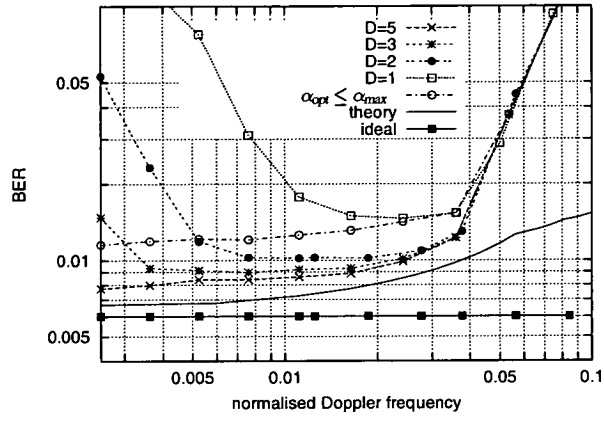


Figure 6.28: *BER vs ν'_{max} using an IIR estimation filter for different numbers of D. $M=8$, $R=10$, $Q=2$, $\bar{\gamma}=10$ dB.*

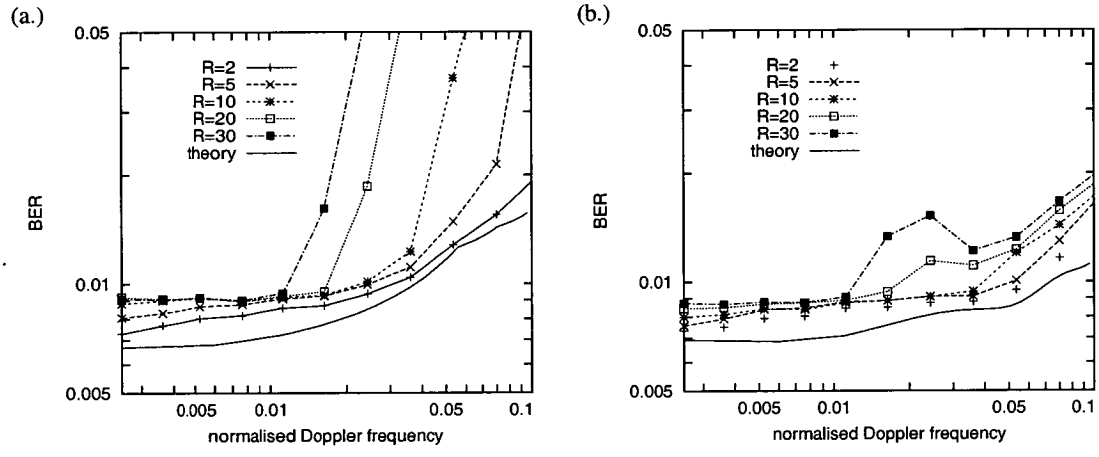


Figure 6.29: *BER vs ν'_{max} for different R, with: (a.) a receiver with IIR filter and filter constant $\alpha_{opt} \leq \alpha_{max}$; (b.) a receiver with FIR filter, $M=8$. $D=3$, $Q=2$, $\bar{\gamma}=10$ dB.*

depicted in Figure 6.25. The fast fading performance ($\nu'_{max} > 0.04$) of IIR–PSP is absolutely unaffected. Thus, increasing the number of decoder states cannot improve the ability of the IIR estimation filter to track the carrier phase, if ν'_{max} is high. The error floor for fast fading experienced by IIR–PSP cannot be lowered, independent of D. The reason is that for fast fading the filter constant becomes negligible, i.e. $\alpha_{opt} \approx 0$. Now the CIR estimate becomes $\widehat{h}_q(i,k) = y'(i,k-1)$ and the receiver is similar to FIR–PSP with a short predictor (in this case $M=1$), described in the previous section.

Note, it was shown that for IIR-PSP, the slow fading performance can be significantly improved by upper bounding the filter constant $\alpha_{opt} \leq 0.5$. This is also shown in Figure 6.28, where the graph labelled "$\alpha_{opt} \leq \alpha_{max}$" is shown for comparison purposes. It is seen that the performance for slow fading can be improved by expanding the trellis. However, this is achieved by considerably increasing the complexity. For the parameters chosen for the plots, with $A_m = 2$, the trellis has 2^D more states with also 2^D more transitions per time step.

Comparison of FIR and IIR filtering In Figure 6.29.a the BER is plotted against ν'_{max} of an IIR filter with α_{opt}, for various values of the pilot multiplexing rate R. It is seen that the performance of the IIR–PSP with $D = 3$ for fast fading is severely affected by the choice of R. For improved performance R must be decreased, note this is not desirable as this increases the system overhead. The slow fading performance of IIR filtering, observed in Figure 6.28, can be improved by upper bounding $\alpha_{opt} \leq \alpha_{max}$, shown in Figure 6.29.a. The upper bound α_{max} is chosen as a function of R. For $R \leq 5$ the filter parameter α_{opt} does not need to be truncated, while for $R \geq 10$, α_{opt}, was upper bounded by$\alpha_{max} = 0.6$. These are similar observations as for the IIR-PSP ($D = 1$), hence expanding the trellis does not yield to a significant improvement in system performance.

In comparison, the results for the same set of parameters for a receiver with FIR filter are shown in Figure 6.29.b. Unlike the IIR filtering receiver, FIR filtering is not severely affected by the choice of R. To conclude, as for $D \geq 3$, FIR filtering does not suffer from stability problems, it appears to be the more appropriate choice, due to its advantages in fast fading conditions.

6.4.3 List type Viterbi decoding

In this section, the performance of the list type Viterbi algorithm (LVA) having $L^\star = 2$ super states is investigated. The LVA was described in section 6.1.2. Unlike the VA based PSP, the LVA has not been applied to MLSD with unknown parameter estimation yet. The performance of the LVA proposed by Auer *et al* [138] is investigated in this section. Simulation work concentrates on the effects of the choice of the following parameters: the list size J, which determines the complexity of the algorithm; and the time delay D, specifying the length of the time window for the ambiguity check (number of samples for which candidate sequences must be different). Note for $J = 2^{D-1}$ the LVA operates on a full list. Setting the number of super states to $L^\star = 2$ results in a receiver complexity similar to PSP with $L = 2J$ states. The LVA using an M^{th} order linear predictor was investigated in Figures 6.30 and 6.31.

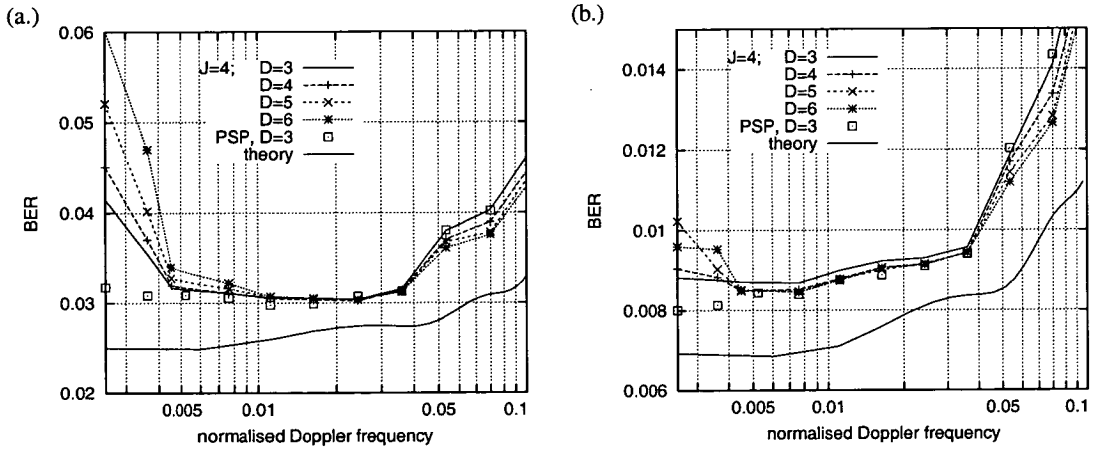


Figure 6.30: *BER vs ν'_{max} for 4-SP, with various D.*
(a.) flat fading $Q = 1$, (b.) diversity $Q = 2$. $J = 4$, $R = 10$, $M = 8$, $\bar{\gamma} = 10$ dB.

Figure 6.30 shows the BER against the normalised Doppler frequency ν'_{max} for 4-SP, with various D. Part (a.) shows results for the flat fading and part (b.) for a double diversity channel ($Q = 2$). For low Doppler ($\nu_{max} < 0.01$) the performance improves with decreasing D. This means that J-SP performs

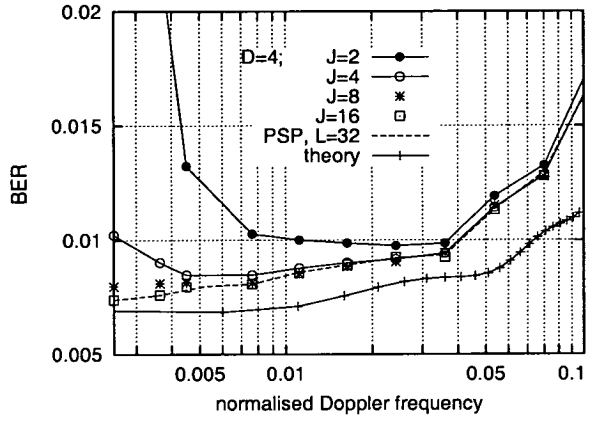


Figure 6.31: *BER vs ν'_{max} for J-SP, with various J and fixed $D=5$. $Q=2$, $R=10$, $\bar{\gamma}=10$ dB.*

better with a full list. The slow fading performance is relatively poor compared to medium fading rates ($0.01 < \nu'_{max} < 0.025$). This can only be improved by increasing the list size J. However, J-SP is outperformed by PSP with $D = 3$ having approximately the same complexity, even if J-SP operates with a full list. Hence, the J-SP is found to be ineffective for slow fading. Similar results are obtained by employing the M-algorithm in the detector [153, 154]. In addition for the sub-optimality due to the reduced set of survivor sequences, the dominant mechanism responsible for this unsatisfactory behaviour is the fact that, on fading channels the received signal $y_q(k)$ may undergo deep fading. Since the noise part dominates the signal part of $y_q(k)$ there are many paths metrics which have virtually the same distance. In addition the number of contenders having virtually the same path metric may quickly grow when the deep fade continues. This is more likely to happen for slow fading, thus J-SP works well for fast fading, but degrades when the fading rate decreases. Furthermore, with diversity, the occurrence of deep fades is less likely, so the algorithm performs better with increasing diversity, as seen in Figure 6.30.b. For fast fading the situation changes as receivers with larger D are slightly superior. Now, with deep fades unlikely to persist longer than a few samples, the list size J is of less importance and the algorithm is able to slightly improve the performance for increasing D.

Figure 6.31 shows the BER against ν'_{max} for the same receiver, with various J and fixed $D = 5$. A list size of $J = 4$ appears to be required to allow good performance on arbitrary fading conditions. This receiver has approximately the same complexity as PSP with $D = 3$ (see e.g. Figure 6.30.b) but with poorer performance for slow fading and slightly better performance for fast fading. Again little is gained if the list size exceeds $J > 4$.

Differences between PSP and J-SP To conclude, reducing the complexity of VA-MLSD through list type decoding was seen to be ineffective, compared to PSP. The list type approach with J-SP has failed to reduce the decoder complexity compared to PSP based on a $L = 32$ state trellis ($D = 5$). Instead, for receivers with comparable complexity, the fast fading performance can be slightly improved. There are some profound differences of the list type algorithm and the PSP based receivers studied earlier on, the

main points are highlighted below.

PSP with $L = 2^D$ states:	J**-SP:**
The A_m candidate paths which are to be discarded were different for at least D samples.	The list of the J best paths is generated comparing *all* $A_m J$ candidates of the previous sample.
$\Rightarrow$ The survivor of sample $k+1$ is found for the state sequences, $\chi(k)$, which differ on the last bit with delay, $k-D+1$, only.	$\Rightarrow$ Paths which would remain with PSP may be discarded falsely.
	$\Rightarrow$ J-SP is not as selective as $L = 2^D$ state PSP.

6.5 Hybrid receiver structures

Thus far, several MLSD receiver structures have been discussed. The PSP based receiver from section 6.3 is very sensitive to the normalised Doppler frequency ν'_{max}. For 2–state PSP, depending on the choice of the estimation filter the system performance was poor either on slow fading for the FIR–PSP, or fast fading for the IIR–PSP. On the other hand, the FIR–PSP with an expanded trellis works well in both slow and fast fading conditions; with the drawback of a significantly increased complexity.

In this section hybrid filtering techniques will be considered for PSP based on a 2–state trellis. Such techniques are based on the observation that FIR–PSP works excellently on fast fading channels. On slow fading conditions constructing an appropriate receiver is thought to be less difficult than for fast fading. Generally, the proposed hybrid receiver consists of PSP with FIR filtering and a second, robust receiver. The two hybrid receivers chosen for further investigation are: first, a combination of PSP with FIR and IIR filtering, which was proposed by Auer *et al* [144], and secondly, a hybrid of FIR–PSP and the pilot aided receiver (PA–RAKE), which exclusively uses pilot symbols for channel estimation, described in section 4.2. Such hybrid techniques are potentially less complex than the VA–MLSD, although suffer some degradation in system performance.

The main problem for the poor performance of FIR–PSP for slow fading were stability problems caused by phase-slips after deep fades. The receiver was subsequently unable to recover during good reception areas and therefore the BER temporarily approached 100%, illustrated in Figure 6.9. In this section only flat fading receivers ($Q = 1$) are investigated. So the subscript q for the received signal variables is dropped. However, the results can be easily extended to diversity reception.

Two different hybrid structures are investigated:

(i.) The two receivers run in parallel continuously. The decisions of both receivers are compared, and if a phase slip between the reference receiver and the FIR–PSP is detected, the receiver simply inverts the output of the FIR-PSP, until the occurrence of the next phase slip.

(ii.) The receiver may switch from FIR-PSP to a robust technique whenever conditions are identified as unstable.

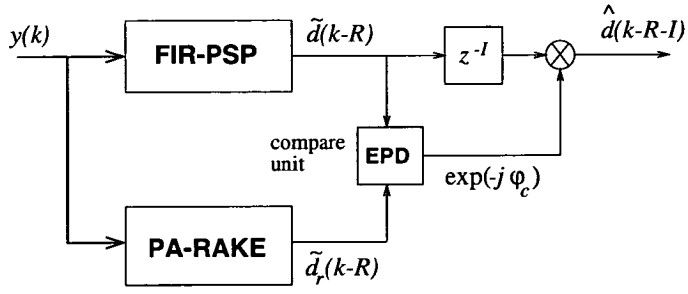


Figure 6.32: *Block diagram of PSP with error propagation detector (EPD).*

6.5.1 Error propagation detector

The hybrid design utilising an error propagation detector (EPD) consists of a robust reference receiver and the FIR-PSP, providing the tentative decisions $\tilde{d}(k)$. The pilot aided receiver of section 4.2, where channel estimation is exclusively based on the pilots, is used as the robust reference, delivering the reference output $\tilde{d}_r(k)$. By comparing the corresponding outputs between $\{\tilde{d}(k-i)\}$ and $\{\tilde{d}_r(k-i)\}$ in the interval $i = \{0, \cdots, I-1\}$, the number of the unmatched phases $N_\Delta \leq I$ is recorded for every sample. Whenever N_Δ is larger than a certain threshold $N_{th} \leq N_\Delta \leq I$, the error propagation detector detects a phase slip and the receiver inverts the tentative decisions $\tilde{d}(k)$. After a delay of I samples the receiver output is given by

$$\hat{d}(k-I) = e^{-j\varphi_c}\,\tilde{d}(k-I)\;; \quad \varphi_c = \arg\left[d(k)\right] \tag{6.11}$$

where φ_c is the output of the EPD which conjugates $\{\tilde{d}(k-i)\}$ if a phase slip is detected. A block diagram of the error propagation detector is depicted in Figure 6.32. A similar error propagation detector was proposed in [155].

The optimal length of the interval I and the threshold N_{th} are determined through experiments. The total decision delay comprises to R+I due to the delay of R of PSP, thus I imposes an additional decision delay. On the other hand, it is possible that the error propagation detector gives a *false alarm*, i.e. the EPD falsely detects an error. In order to minimise the probability of a false alarm, the threshold N_{th} and the observation interval I need to be high. Hence, there is a trade-off in terms of I, between the imposed decision delay and the probability of a false alarm. Good results were achieved with $I = 10$ and $N_{th} = 8$. Also, since for high ν'_{max} the reference receiver becomes unreliable, while FIR-PSP is then robust, the error propagation detector should be switched off for fast fading. The system performance is assessed through computer simulations. Results are shown in comparison with other hybrid techniques in Figure 6.34 studied in the next section.

6.5.2 Hybrid filtering

In order to circumvent the stability problems, the receiver may switch from FIR–PSP to the robust receiver if conditions are identified to be unstable. This may be a gradual change from one filter to the

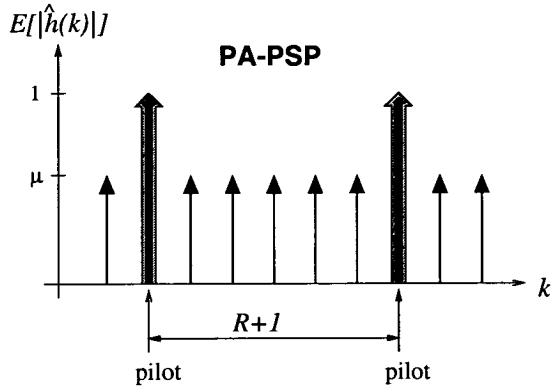


Figure 6.33: *Average contribution of received signal samples for PA-PSP.*

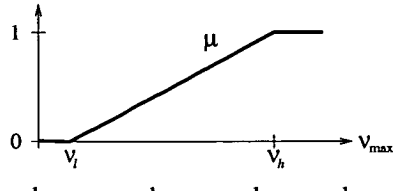
other or a binary switch. The CIR estimate of the hybrid PSP is defined by the piecewise linear equation

$$\widehat{h}_h(k) \triangleq \mu\,\widehat{h}_{\text{fir}}(k) + (1-\mu)\,\widehat{h}_r(k)\;; \qquad \mu \in [0,1] \tag{6.12}$$

where $\widehat{h}_{\text{fir}}(k)$ is the CIR estimate of a linear predictor and $\widehat{h}_r(k)$ is the estimate of a robust reference. The estimated response $\widehat{h}_h(k)$ is subsequently processed with the VA, in per-survivor fashion. The parameter μ is dependent on the channel statistics, ν'_{max} and $\bar{\gamma}$. As it is difficult to find μ analytically, the optimum values are to be assessed through computer simulations. Furthermore the particular choice of μ is dependent on the filter which provides $\widehat{h}_r(k)$. Two types of hybrid receivers were investigated, which are addressed in the following.

6.5.2.1 Emphasise the pilot symbols for channel estimation

FIR–PSP combined with the PA–RAKE, termed PA–PSP is considered first. Here the basic idea is to weight pilots more than data symbols for channel estimation, if FIR–PSP alone appears unstable. That means the pilots are "emphasised" through pilot aided channel estimation to maintain stability of the receiver at low Doppler. This may be achieved by using a receiver that uses the pilots only for channel estimation as the reference filter to provide $\widehat{h}_r(k)$. The estimate $\widehat{h}_r(k) = \widehat{h}_{\text{pa}}(k)$ is given by (4.11). As only pilot symbols are involved in estimating the CIR the receiver is robust. On the other hand, without decision feedback errors $\widehat{h}_{\text{fir}}(k)$ the performance is superior compared to $\widehat{h}_{\text{pa}}(k)$, since the sampling rate of the PA–RAKE is $1/R$ times the sampling rate of FIR–PSP. With decision errors, however, the situation might change, particularly at low SNR and low Doppler. The factor μ is described by the linear piecewise equation



$$\mu \triangleq \begin{cases} 0 & \nu'_{max} < \nu_l \\ \frac{\nu'_{max}-\nu_l}{\nu_h-\nu_l} & \nu_l \le \nu'_{max} \le \nu_h \\ 1 & \nu'_{max} > \nu_h \end{cases} \tag{6.13}$$

where ν_l and ν_h are a lower and upper boundary. Since the receiver gradually changes between the reference and PSP it is termed *soft* PA-PSP. For low Doppler the pilot symbols have more impact on

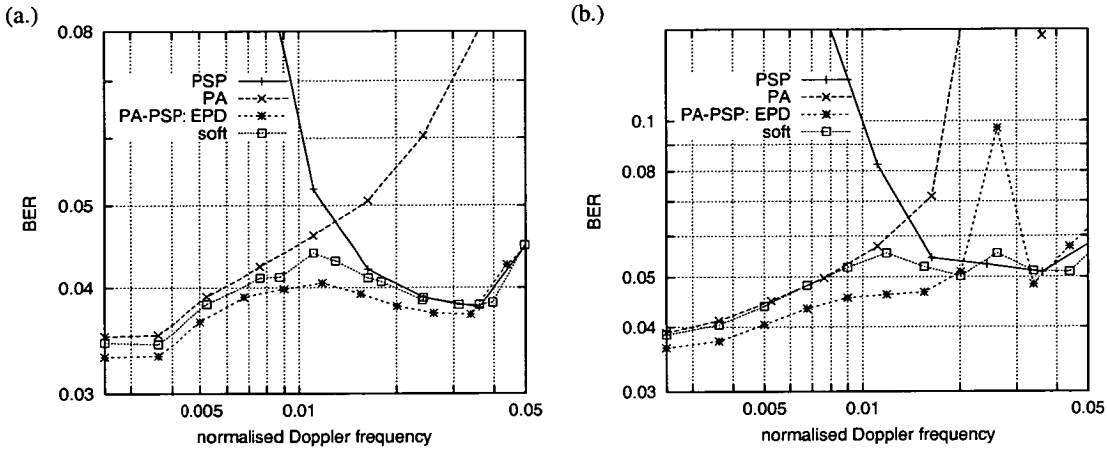


Figure 6.34: *BER vs normalised Doppler frequency, ν'_{max}, for soft PA–PSP in comparison with PSP using EPD.*
(a.) $R=10$, (b.) $R=20$;
$M=8$, $Q=1$, $\bar{\gamma}=10$ dB.

channel estimation through $\widehat{h}_{pa}(k)$. The boundaries ν_l and ν_h may be adapted for changing $\bar{\gamma}$. The concept of the PA–PSP is illustrated in Figure 6.33. The larger μ, the more the receiver acts in a decision directed fashion. For $\mu=0$ channel estimation is performed by only using the pilots, while $\mu=1$ results in ordinary FIR–PSP. For most applications appropriate values for the boundaries are: $\nu_l=0$ and $\nu_h=0.02$.

soft PA-PSP:	ν_l	0
	ν_h	0.02
PSP with EPD:	N_Δ	10
	N_{th}	8
PA-RAKE:	M	8
	M_f	0

Table 6.3: *Hybrid filter parameters.*

Simulation results of the soft PA–PSP compared to PA– PSP with an error propagation detector (EPD) are shown in the following graphs. The filter parameters are given in Table 6.3. Figure 6.34 shows the BER against the normalised Doppler frequency, ν'_{max}, for a pilot multiplexing rate of (a.) $R=10$ and (b.) $R=20$. In part (a.) the PSP with EPD is superior to the soft PA-PSP for slow fading. For $R=20$ in part (b.) the results are similar, however the differences in BER are more pronounced. The performance for high Doppler can be improved by switching off the EPD for $\nu'_{max}>0.017$. Given that, PA–PSP with EPD appears preferable for implementation.

6.5.2.2 Hybrid filtering MLSD

In section 6.3.4 PSP based receiver structures were discussed which work well in slow fading, but not in fast fading. Those were PSP with IIR filtering and linear predictors with short filters ($M=2$). The proposed hybrid filtering MLSD switches from FIR-PSP with high filter order, to a robust receiver whenever conditions are identified to be unstable. This is mainly the normalised Doppler frequency ν'_{max}, as seen in Figure 6.35, where the BER is plotted against ν'_{max}, for some values of the SNR, $\bar{\gamma}$. For slow fading

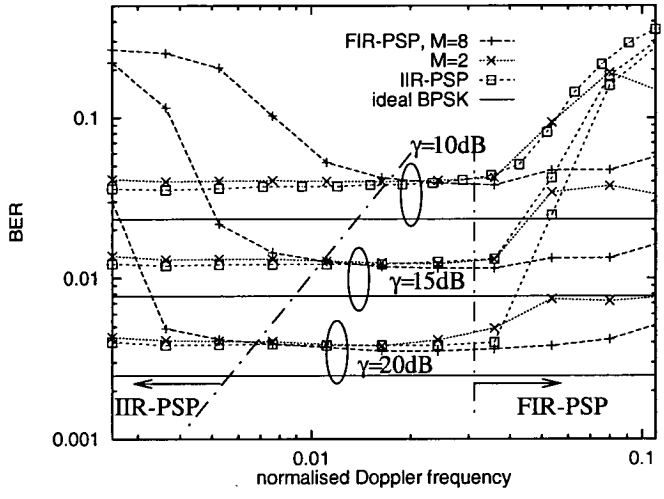


Figure 6.35: *BER vs normalised Doppler frequency, ν'_{max}, for FIR–PSP and IIR–PSP.* $R=10$, $M=8$, $Q=1$.

PSP with IIR filtering is slightly superior to the short predictor with $M = 2$. Thus the combination of IIR and FIR–PSP will be referred to as hybrid PSP (HPSP). Hybrid FIR and IIR channel estimation $\widehat{h}_r(k) = \widehat{h}_{iir}(k)$ is provided by the IIR–PSP of equation (5.30) [138, 144]. Here a switch $\mu = \{0, 1\}$ appears to be more appropriate, so either $\widehat{h}_{iir}(k)$ or $\widehat{h}_{fir}(k)$ is used to generate the CIR estimate, as also shown in Figure 6.35. There is a great overlap between FIR–PSP and IIR–PSP where both estimation filters have virtually the same performance. Hence, for the design of a receiver which can operate in arbitrary fading conditions, a hybrid solution of combined FIR and IIR–PSP appears attractive. The HPSP simply switches on changing fading rates between the FIR and IIR estimation filter, dependent on ν'_{max}. Only a rough estimate of $\widehat{\nu}_{max}$ is required, due to the large overlap of the FIR and IIR–PSP performance.

For low SNR $\bar{\gamma}$ the overlap between FIR and IIR filtering becomes smaller, which can also be observed in Figure 6.35. Furthermore, the received signal power experiences vast changes compared to the average value, due to signal fading. Hence, it may be worthwhile switching between the FIR and IIR filter dependent on the instantaneous SNR $\gamma(k)$ rather than the average $\bar{\gamma}$. A technique which aims to achieve that is described in Appendix B.1.2. Unfortunately, this approach was not successful and hence switching will be performed dependent on the average channel parameters $\bar{\gamma}$ and ν'_{max} in the following.

Figure 6.36 shows the BER against ν'_{max}, for some pilot multiplexing rates R. It is seen that for $R \geq 20$ both FIR and IIR filtering fail to produce reasonable performance for medium fading rates ($0.01 \leq \nu'_{max} \leq 0.03$). Hence HPSP is only applicable for $R \leq 10$ and a value of $R = 10$ will be assumed in the following.

6.5.3 Comparison of hybrid filters

Hybrid techniques were applied to PSP with FIR filtering studied earlier in this chapter. The objective was to create a receiver to work in arbitrary fading conditions, being less complex than the VA-MLSD. The HPSP, being an IIR and FIR–PSP combination, and PA-PSP with an error propagation detector

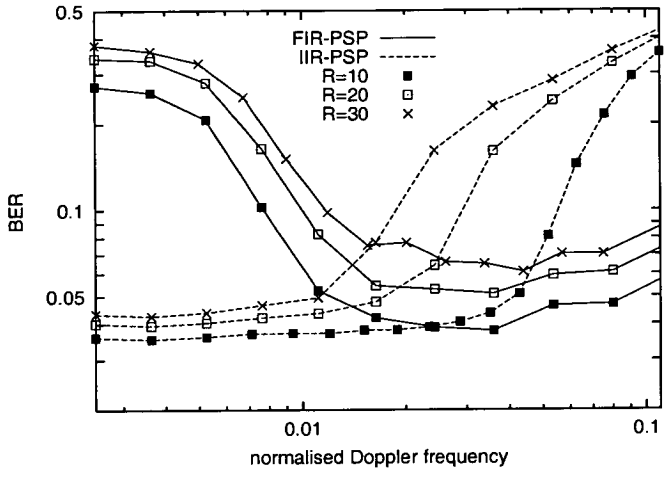


Figure 6.36: *BER vs ν'_{max}, for FIR–PSP and IIR–PSP for various R.*
$M=8$, $\bar{\gamma}=10$ dB, $Q=1$.

(EPD) appear to be appropriate choices for hybrid filter designs. Good results were achieved for both receivers at a pilot multiplexing rate of $R=10$, however, as R increases the degradation in performance was significant. Note, the HPSP is far easier to implement, since most of the receiver components can be re-used for both receiver parts, while the PA–PSP requires a Viterbi decoder for fast fading in the FIR–PSP, and a significantly different receiver, i.e. a filter bank for the pilot aided channel estimation for slow fading.

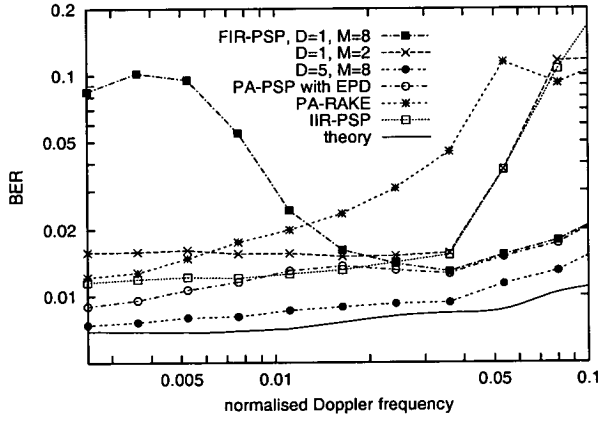


Figure 6.37: *BER vs normalised Doppler frequency, ν'_{max}, for hybrid receiver structures.*
$Q=2$, $R=10$, $\bar{\gamma}=10$ dB.

Simulation results for a two tap diversity channel presented in Figure 6.37 show that the PA-PSP with EPD works very well, provided that the EPD is switched off for fading rates $\nu'_{max} > 0.02$. For comparison purposes the FIR–PSP with $D=5$ is also shown in the graph. In fact, it is seen that for slow fading the error performance of PA-PSP with EPD approaches the expanded trellis PSP. Diversity particularly benefits the EPD technique, since the reference is more reliable and the error propagation detector can detect phase slips of the FIR-PSP more efficiently. IIR-PSP works better in slow fading than

the predictor with $M=2$ and so it is the preferable choice for HPSP.

6.6 Assessing the effects of multiple access interference

Thus far, diversity systems have been studied where the interference have been additive white Gaussian noise (AWGN). However, on a direct sequence (DS) CDMA system, the interference caused by other users transmitting through the same channel, is coloured. Thus, the results for the diversity receivers of Chapters 4–6 cannot be applied to a RAKE receiver, without taking into account the interference caused by other users. The focus is on the the base station to mobile link (downlink), so all users transmit through the same channel and the CIR of tap q, is the same for all users. Furthermore, synchronised transmission is assumed, which grossly simplifies the analysis.

The optimum receiver for a system with MAI is very similar to the optimum receiver of an ISI channel [104, 156]. In this section, however, no attempt is made to reduce MAI other than through the choice of appropriate spreading codes. Instead the performance of the algorithms investigated so far will be assessed for non-Gaussian interference such as MAI.

The limiting effects of multiple access interference (MAI) are examined in the following. In order to do so, section 3.3 where the RAKE receiver front-end was addressed is re-visited. In section 3.3 the receiver front-end was described, consisting of a tapped delay line of decorrelators for the desired user (see Figure 3.7) The received signal for user u of the q^{th} RAKE finger tap, after despreading with the desired users' signature waveform is given by (3.19):

$$y_{qu}(k) = d_u(k)\, h_q(k)\, \rho_{uu}(0) + \underbrace{\sum_{p=1}^{Q} \sum_{\substack{v=1 \\ v\neq u \cap p\neq q}}^{U} d_v(k)\, h_p(k)\, \rho_{uv}(p-q)}_{\text{MAI}} + n'(k) \tag{6.14}$$

where $\rho_{uv}(p-q)$ denotes the cross-correlation between signature waveforms u and v, with relative chip delay $p-q$, which accounts for the interference user v has on the desired user u. Subscripts identifying a certain user $u, v \in \{1, \cdots, U\}$ have been introduced. This notation implies that inter-symbol interference (ISI) caused by the multipath fading channel are not taken into account, according to the discussion in section 3.3. The discussion in this chapter is limited to the binary case, so the spreading codes take on the values 1 and -1, yielding $\rho_{uv}(\cdot) \in \mathbb{R}$. It is seen that in addition to the term for the desired user and the background noise, we have to deal with a third quantity, the MAI term, i.e. the interference from other users. MAI ultimately limits the system performance resulting in an error floor. Therefore, the choice of code families with low correlation coefficients $\rho_{uv}(p-q)$, for $u \neq v$ and $q \neq p$, is of great importance for the performance of a DS–CDMA system.

Subsequently, the received signal in (6.14) is processed according to the receiver structures studied in the previous chapters. The difference is that the channel estimator needs to be modified due to the increased noise term of MAI, which is coloured and data dependent. Hence, the channel estimation filter,

determined by the MMSE criterion $\mathrm{E}\left[\,\left|y_{qu}(k) - d_u^{(o)}(k)\,\widehat{h}_q(\ell,k)\right|^2\,\right]$, become dependent on the data bits of all interfering users. One way out of this dilemma is to jointly estimate and detect the signals for all users, rather than separately detect the signal for each user. Techniques aiming to achieve this are summarised in the next section. This section follows the conventional approach to detect each user's signal separately. The additional noise introduced by the MAI term can be taken into account by the Gaussian approximation, i.e. the binomial distributed sum $\sum_{v,p} d_v(k)\,h_p(k)\,\rho_{uv}(p-q)$ from (6.14) is replaced by a Gaussian pdf with identical variance $\sum_{v,p} \bar{\alpha}_p^2 \rho_{uu}^2(p-q)$, where $\bar{\alpha}_p^2 = \mathrm{E}\left[\,|h_p(k)|^2\,\right]$ denotes the tap weight for the p^{th} channel tap. The Gaussian approximation implies that the auto-correlation of $y_{qu}(k)$ as a function of the time k can be approximated as

$$\mathrm{E}\left[\,y_{qu}(k)\;y_{qu}^*(k+\Delta k)\,\right] \approx \begin{cases} \bar{\alpha}_q^2 + N_0 + \sum_{p=1}^{Q} \sum_{\substack{v=1 \\ v\neq u \cap p\neq q}}^{U} \bar{\alpha}_p^2 \rho_{vv}^2(p-q)\;; & \Delta k = 0 \\ \mathrm{E}\left[\,h_q(k)\;h_q^*(k+\Delta k)\,\right]\;; & \text{elsewhere} \end{cases} \tag{6.15}$$

where $\rho_{uu}^2(0)$ is normalised to one. Note that with the Gaussian approximation the contribution of the MAI term has become independent of the data. Thus, the Wiener filter $\mathbf{w}$ can now be determined according to (2.18) in section 2.1.3.

Simulation Results Monte Carlo simulations were carried to assess the effects of MAI. For results throughout this chapter Gold codes with a processing gain of $N_c = 31$ were used. A signal generated according to (6.14) on the symbol level was fed to various detectors studied in previous chapters. The receiver parameters of the investigated receivers, including a reference where a more detailed description can be found, are depicted in Table 6.4. Generally, results presented in this chapter concentrate on the comparison between different receivers which estimate the channel response in the presence of MAI. The stability of decision directed receivers under this conditions is of particular interest. Various decision directed receivers are compared with the pilot aided (PA) receiver, which only uses the pilots to estimate the channel response. This will allow a distinction between MAI and decision feedback related performance degradation.

Receiver	**FIR–PSP**	**IIR–PSP**	**DD–RAKE**	**PA–RAKE**
Filter	$M=8$	$\alpha_{opt} \leq 0.5$	$M=8$	$\substack{M=8 \\ \Delta\kappa=0}$
Modulation	BPSK	BPSK	DEPSK	BPSK
PSI rate	$R=10$	$R=10$	-	$R=10$
Section	6.3.2	6.3.4	4.3.2	4.2
Codes:	**Gold codes:** Processing gain $N_c=31$			

Table 6.4: *Parameters for the considered receiver structures.*

Figure 6.38 shows results for slow fading of the PA–RAKE from section 4.2. The BER is plotted against the SNR for different numbers of active users, on (a.) a flat fading channel and (b.) a frequency selective fading channel with $Q=2$. MAI is more severely affected by the frequency selective channel, than the flat fading counterpart, which indicates that the diversity benefit is reversed to a penalty. One reason for the observed performance penalty is the high cross-correlation values of not synchronised spreading codes, due to the imposed spread in time of the received signal by the frequency selective

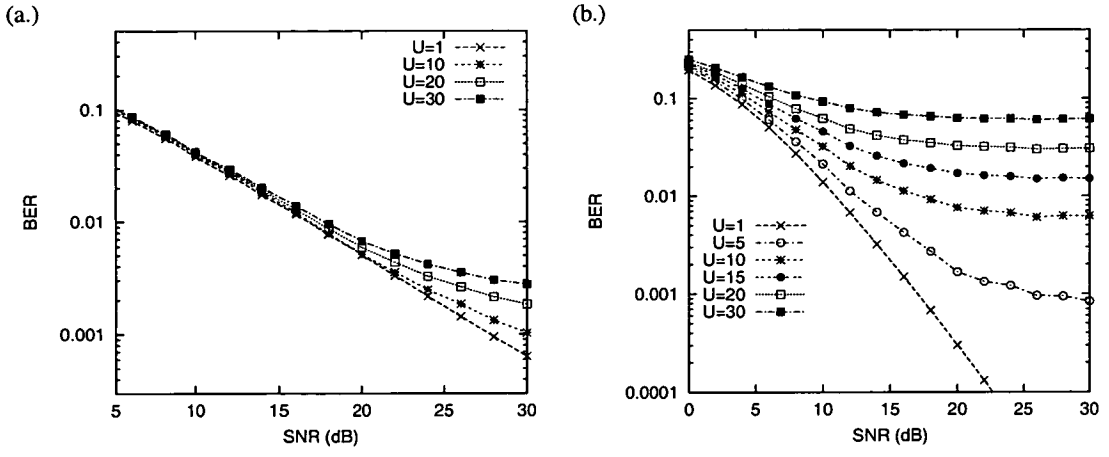


Figure 6.38: *BER vs SNR for the PA–RAKE with different numbers of active users U.*
Slow fading: $\nu'_{max}=0.005$, $M=8$, $\Delta\kappa=0$;
(a.) flat fading: $Q=1$, *(b.) double diversity: $Q=2$.*

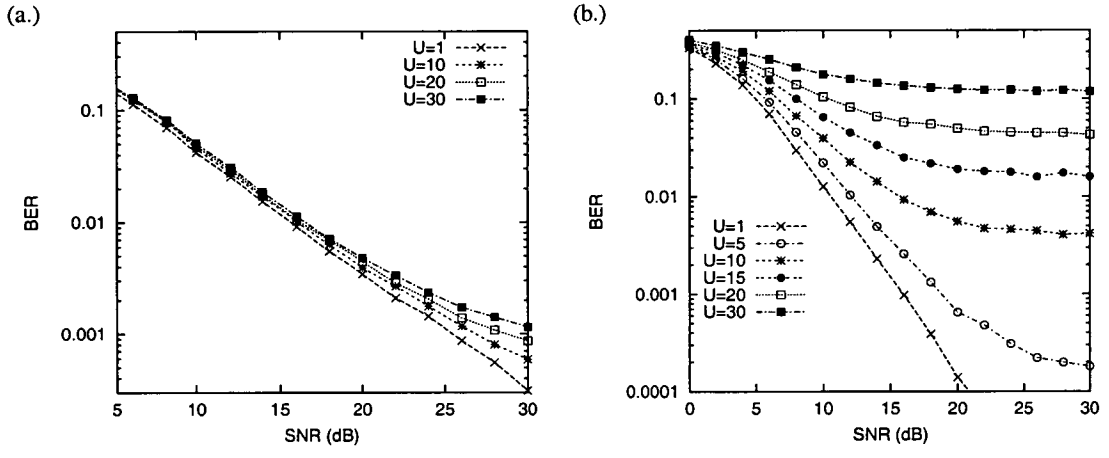


Figure 6.39: *BER vs SNR for 2–state PSP with different numbers of active users U.*
Fast fading: $\nu'_{max}=0.05$, $M=8$,
(a.) flat fading: $Q=1$, *(b.) double diversity: $Q=2$.*

fading channel. Gold codes with a processing gain of $N_c = 31$ take on the cross-correlation values $\rho_{uv}(p-q) = \{-1/31, \pm 7/31, -9/31\}$, being up to 9 times higher in magnitude than synchronised codes, which always are $\rho_{uv}(0) = -1/31$. In addition to the high cross-correlation values, receiving the codes asynchronously results in the *near-far effect*. That is, when the desired user's signal is in a deep fade, the signal strength of at least one interfering user may be several times stronger, with the effect that the desired user's signal is overshadowed by the interference, severely degrading the performance. On the other hand, since synchronous transmission is assumed, there is no near-far effect for flat fading, therefore the performance is far less compromised by MAI.

In Figure 6.39 the performance of PSP based on a 2–state trellis from section 6.39 is shown for fast fading. Similar results as for the PA–RAKE shown in the previous figure are obtained. For flat fading there is relatively little degradation due to MAI observed, whereas the BER degradation for the 2 tap channel is very severe. Consider $Q = 2$, compared to the PA–RAKE in Figure 6.38.b, the performance of PSP is superior for low load ($U \leq 15$) despite the higher fading rate, whereas for heavy load the BER is slightly poorer. Note that error propagation due to decision feedback effects is not responsible for

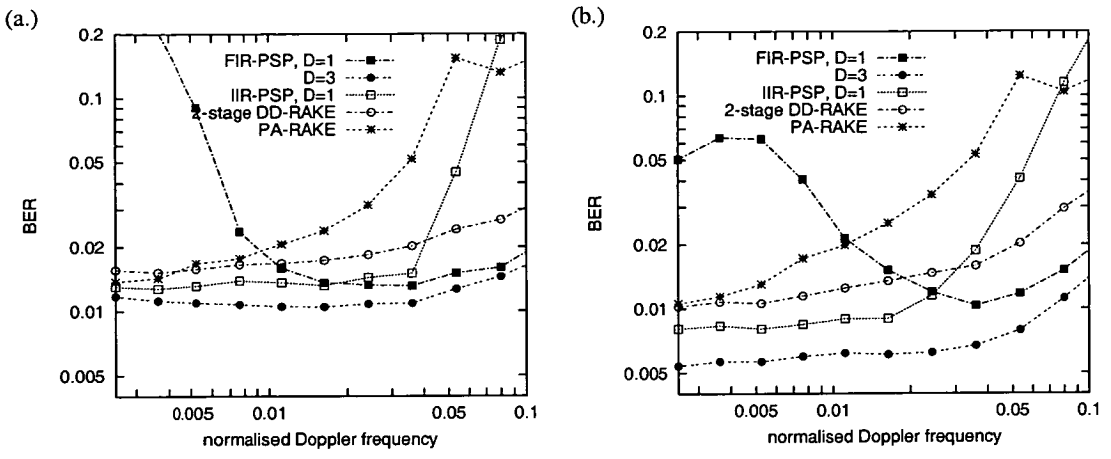


Figure 6.40: *BER vs ν'_{max} for various receiver structures;* $\bar{\gamma} = 15$ *dB,*
(a.) flat fading: $Q = 1$ *and* $U = 30$ *(b.) double diversity:* $Q = 2$ *and* $U = 10$.

the poor performance of PSP in this case, since the PA–RAKE which is a purely non-decision feedback receiver shows a similar behaviour.

Various receiver designs specified in Table 6.4, are compared with each other as a function of ν'_{max} in Figure 6.40. Part (a.) shows the performance of the flat fading case with heavy load $U = 30$ and part (b.) a channel with $Q = 2$ taps was chosen with $U = 10$ users. With these parameters receivers of parts (a.) and (b.) work in the same BER region. Comparing the receiver structures on parts (a.) and (b.) indicates that the difference in BER of the investigated receivers becomes larger for the 2 tap diversity channel of Figure 6.40.b. Generally, the features of the shown receivers are similar. So, although MAI does affect the performance, the phase tracking characteristics of the single user case are retained, which are discussed in the following. The 2–state FIR–PSP ($D = 1$) is found to be unstable for low SNR. The stability problems can be overcome by expanding the trellis, shown in the graph for the 8–state PSP ($D = 3$). The IIR–PSP, with the filter constant chosen according to (2.22) and upper bounded $\alpha_{opt} \leq 0.5$, does work well for slow fading but breaks down for fast fading ($\nu'_{max} > 0.04$). The difference in the BER between the 2–stage decision directed (DD) RAKE and the 8–state PSP is mainly due to usage of DPSK of the former receiver. The PA–RAKE has poor performance for fast fading, since the channel predictor works at a $R = 10$ times higher sampling rate compared to the decision directed receivers. The slow fading performance of the PA–RAKE can be improved by employing a smoothing type channel estimator with $0 < \Delta\kappa \leq M/2$.

As mentioned earlier, work presented in this section is focused on the downlink. Note that on a flat fading channel the downlink degradation in performance is less severe than for the corresponding case on the uplink. On the uplink, the conventional RAKE experiences an error floor, even if the cross-correlation between codes is low, due to the near-far problem [157]. However, as it has been demonstrated, this may change when a frequency selective fading channel is considered, as much of the diversity benefit of the Q independently fading taps is lost through MAI. Amongst other techniques, on way to improve the downlink capacity may be multi-user detection.

Summary and Conclusions In this Chapter the effects of multiple access interference (MAI) were investigated. In section 6.6 simulation results for the downlink of a DS–CDMA system suggest that the effects of MAI are far more severe on a frequency selective channel, compared to a flat fading channel. The diversity benefit, due to independently fading taps $h_q(k)$, was reversed to a penalty due to the near-far effect. Furthermore, the multipath fading channel destroys the synchronous transmission of the spreading codes. This degrades the performance further, since auto- and cross-correlation amplitudes of the codes for asynchronous transmission is several times higher in magnitude, than for synchronised transmission. This observation is independent of the chosen receiver structure. However, for the frequency selective channel ($Q = 2$) the differences between different receiver realisations become more pronounced. MAI does not inflict further stability problems for decision directed receiver architectures.

6.7 Summary and conclusions

A comparison of the main points of the addressed receiver structures in terms of performance and complexity is now presented. The computational cost of receivers based on the VA, in terms of the number of survivor paths in the trellis, $O(\cdot)$, is shown in Table 6.5.

Receiver:	**VA–MLSD**	**PSP**	J**–SP**
complexity order:	$O(2^M)$	$O(2^D)$	$O(2J)$
		$D \leq M$	$J \leq 2^{D-1}$

Table 6.5: *Computational cost of the considered receiver architectures, for binary modulation ($A_m = 2$).*

VA–MLSD States are defined incorporating the current and past M received signal samples. The receiver needs to process $L = 2^M$ states and as many survivor paths, resulting in a computational cost of order $O(2^M)$. FIR filtering appears to be more appropriate, due to its better fast fading performance. Complexity grows exponentially with the filter order M. The performance is close to the theoretical lower bound for a given M.

FIR–PSP Compared to VA–MLSD the number of states in the trellis are reduced due to state dependent decision feedback, by introducing the parameter D. A trellis with $L = 2^D$ states is obtained, resulting in a computational cost of order $O(2^D) \leq O(2^M)$. For the considered system model, little is to be gained if $D > 3$. Alternatively, for a given D, longer predictors may be chosen than for VA-MLSD ($D = M$), then PSP outperforms VA-MLSD.

The performance is only close to the lower bound for all receiver types, for a ratio of data to pilot symbols of $R = 2$, almost independent of D. On the other hand, for larger R, the performance is more dependent on D and M. Suppose $M < 4$: increasing R has degrading effects on fast fading performance, particularly for low SNR. Conversely, if $M \geq 4$ the receiver performance becomes generally less dependent on R, but more dependent on D. This is particularly true for slow fading, where choosing $D < 3$ results in stability problems. The effect that the dependence on R decreases is encouraging, since an increasing R maximises the spectral efficiency, by reducing the number of pilots per frame. However, the requirement for an expanded trellis PSP is not desirable due to the increased complexity.

IIR–PSP This algorithm is similar to FIR–PSP, with the difference that a 1st order IIR filter is employed for channel estimation. The complexity is similar to FIR–PSP with $M = 2$. The performance is relatively independent of D, so a 2–state trellis PSP is usually sufficient. Generally, IIR–PSP has a good slow fading performance. However, increasing R drastically degrades the fast fading performance. The receiver experiences an error floor dependent on R, which cannot be mitigated by expanding the trellis.

J-SP This receiver is a member of the family of the list-type Viterbi decoders. The decoder processes $L^{\star} = 2$ super states, each having a list of J candidates. The complexity is dependent on J, having the computational cost of $O(2J) \leq O(2^D)$. The state dependent lists are updated using the M-algorithm. The J-SP has poorer slow fading performance than PSP with the same complexity. However, the fast fading performance can be slightly improved.

Hybrid receivers can be employed to make FIR–PSP with a 2–state trellis ($D = 1$) work in arbitrary fading conditions. The receiver either switches to a robust reference when conditions are identified as unstable, or uses a second receiver as a phase reference, via an error propagation detector (EPD). Simulation results suggest that the HPSP, being an IIR and FIR–PSP combination, and PA-PSP with an error propagation detector (EPD), both work well.

Chapter 7
Conclusions

This thesis has been concerned with the design of receivers for operation in an unknown, time-varying Rayleigh fading channel. Particularly, receiver structures being able to cope with a fast fading environment were investigated, where conventional techniques fail, due to the rapidly changing channel impulse response (CIR). Channel estimation based on the minimum mean squared error (MMSE) criterion was investigated in conjunction with various receiver architectures. The studied receiver architectures are divided into two main parts:

One-shot receivers: Here detection of the received signal is carried out independently from estimating the CIR. For estimation, the correlation of adjacent samples is exploited. These estimates are subsequently used to decorrelate the received signal, such that detection is performed on a symbol-by-symbol basis.

Joint detection & estimation: Estimation and detection are performed jointly, taking into account the entire sequence. So, channel estimation is done for the purpose of detection. In contrast to one-shot receivers, more than one signal "hypothesis" (i.e. a possible transmitted sequence) are processed in parallel, which results in increased complexity.

A key objective of the research has been the optimisation of the spectral efficiency, in terms of minimising the system overhead through a necessary phase reference. This was achieved by introducing decision feedback, such that data symbols were used for decision directed channel estimation. The resulting error propagation effects were extensively studied theoretically and through simulations of various receiver types. The final part of this thesis was dedicated to assess and mitigate the effects of multiple access interference (MAI), which is an inherent feature of a CDMA system and ultimately limits its capacity. This chapter will draw together the main conclusions of the work and briefly discuss its limitations. Some suggestions for further research in this area are also presented.

7.1 Summary of the work

In Chapter 2, estimation based on the MMSE criterion was reviewed, and an overview to Bayesian detection was given. In Chapter 3 the features of a mobile DS–CDMA were briefly sketched and the channel model, describing a mobile radio link in an urban environment was considered. The model assumed the equivalent baseband channel and the implementation used to generate a time-variant CIR for computer simulations was also addressed.

Chapter 4 considered one-shot receivers. If coherent modulation is employed, side information is required to estimate the CIR. This side information was provided by time multiplexing known pilot sym-

bols into the data stream with rate $1/R$. For the investigated pilot aided receiver no decision feedback was employed, instead pilot symbols were used exclusively for channel estimation. Therefore this receiver is robust if R is appropriately chosen to allow some degree of oversampling with respect to the fading rate. To estimate the channel response according to the MMSE criterion a filter bank of R Wiener filters is required. The system performance is dependent on whether a smoother or a linear predictor is employed as estimation filter, resulting in a trade-off between improved performance and an imposed decision delay for the smoother compared to the linear predictor.

If a non-coherent modulation technique is chosen, channel estimation can be performed without side information in the form of a phase reference. Then the modulation technique inherently provides the phase reference. For differentially encoded signals a one-shot receiver with decision directed prediction of the channel response was investigated. It was found that decision directed prediction is robust and does not lead to a degradation in performance compared to a decision aided reference receiver, i.e. a non decision feedback receiver, which assumes that correct decisions for channel prediction are available. In fact, in many cases the decision directed receiver did outperform the decision aided reference receiver, an effect which has been thoroughly analysed in this thesis.

Post processing by means of iterative channel estimation (ICE) improves the performance further of both the pilot aided and decision directed receiver, given the BER of the 1st stage is reasonably low (BER $< 5 \cdot 10^{-2}$).

The limitations of one-shot receivers are very much dependent on the particular receiver realisation. The pilot aided receiver with coherent modulation is ineffective for fast fading, since a high pilot insertion rate $1/R$ is required to allow a sufficient degree of oversampling, which in turn compromises the spectral efficiency. On the other hand, the decision directed receiver suffers an approximately 3 dB performance penalty due to differential modulation. These limitations can be overcome by combining the pilot aided with the decision directed receiver. However, this requires a more sophisticated receiver design, hence receiver architectures which jointly estimate the CIR and detect the received signal become attractive.

Before proceeding with the investigation of techniques for pilot symbol aided plus decision directed phase tracking, the optimal maximum likelihood receiver for the detection of the entire sequence was reviewed in Chapter 5. The optimum receiver can be separated into a correlator–estimator structure. That is the receiver consists of an estimator that delivers the MMSE estimates of the fading distortion and a detector that utilises these estimates, by decorrelating the received signal. Although the optimum receiver is very powerful and applicable for arbitrary linearly modulated systems operating in a channel described by a Gaussian distribution, it is in most cases too complex for implementation.

Using Bayes' rule a recursive formulation of the optimum receiver has been derived. That receiver jointly estimates the CIR and detects the most likely data sequence. The optimal channel estimator is a one-step linear predictor. Thus, a causal Wiener filter which minimises the mean squared error for one-step prediction is optimum for detection. Based on a recursive formulation of the optimal receiver the receiver complexity can be drastically reduced by application of the Viterbi algorithm. Although, the resulting receiver is generally sub-optimum, it is the best possible solution, when the channel estimator

is truncated by a M^{th} order linear predictor. Alternatively, an IIR-type channel estimation filter can be applied to the receiver. Furthermore, performance bounds were derived for both a M^{th} order linear predictor and a 1^{st} order IIR filter. Unfortunately, on a Rayleigh fading channel these bounds are loose for low SNR. Thus, Monte Carlo simulations were carried out to assess the receiver performance more accurately.

In Chapter 6 the implementation of recursive maximum likelihood sequence detector (MLSD) utilising the Viterbi algorithm (VA) was studied. Even with the VA, recursive MLSD may still be too complex for implementation, since the number of decoder states grows exponentially with the estimation filter order, M. To circumvent this problem state reduction techniques were devised. These techniques employ decision directed techniques, by reducing the number of hypotheses (decoder states) which are processed in parallel, thus reducing the complexity. One possibility to reduce the number of states is by means of state dependent decision feedback, resulting in a receiver based on the principle of per-survivor processing (PSP). Alternatively, the list-type Viterbi algorithm (LVA) can be applied to the problem. The more decision feedback is employed, the less complex is the receiver, while compromising the optimality. The necessary phase reference was provided by time multiplexed pilot symbols, however, for channel estimation and data detection both pilot and data symbols were used.

One of the main achievements in this thesis was that conditions were identified where a receiver utilising pilot symbol aided plus decision directed techniques can become unstable. Increasing the degree of state dependent decision feedback was shown to corrupt the robustness of the receiver under certain conditions. Specifically, consider the low complexity implementation of PSP based on a 2–state trellis. Rather surprisingly, the receiver was found to be more vulnerable for stability problems in slow fading conditions, which did severely degrade the system performance, whereas in fast fading the receiver was found to be more robust. The nature and order M of the channel estimation filter was shown to be responsible for the instability. For large M or low fading rates, the pilots were unable to impose a sufficient phase reference, causing the receiver to lose track of the CIR phase and to enter a burst state having a BER of virtually 100%. The occurrence of these stability problems can be successfully predicted analytically by means of error propagation analysis based on the Gilbert–Elliott channel (GEC). The GEC is a special case of a hidden Markov model having two states, a good state and a burst state. With the GEC the receiver performance can be predicted more accurately, compared to the conventional performance bounds suggested in the previous chapter.

Employing IIR filtering, the same receiver was found to be more robust, given the filter constant α_{opt} was upper bounded to $\alpha_{opt} \leq \alpha_{max} \approx 0.5$. Moreover, the performance was seen to be not significantly dependent on the receiver complexity, i.e. the number of states in the trellis, for slow fading. However, increasing R drastically degrades the fast fading performance, resulting in an error floor dependent on R. This effect was due to poor channel estimation of the IIR filter, resulting in insufficient phase tracking performance, which was entirely independent of the receiver complexity.

Reducing decision feedback by increasing the number of states did improve the stability of PSP using FIR filtering. Furthermore, the choice of R did not have a significant impact in the receiver performance

(except for high Doppler frequencies in combination with low SNR), which allows the system overhead due to pilot symbols to be less than 10%, independent of the Doppler frequency.

The receiver based on the LVA failed to improve the slow fading performance compared to PSP with comparable complexity, but the fast fading performance could be slightly improved. Similar conclusions concerning the stability of the algorithm can be drawn as for PSP with FIR filtering.

Furthermore, hybrid receiver designs were proposed that made the receiver robust in arbitrary fading conditions. On the one hand, a receiver was developed that consists of a robust reference receiver and the FIR–PSP, running in parallel. The basic idea is that a compare unit, termed the error propagation detector, assisted by the robust reference receiver, detects when the FIR–PSP is entering the burst state. Then, the receiver simply inverts the decisions of the FIR–PSP until the receiver leaves the burst state. On the other hand, a hybrid receiver was proposed which switches between the FIR–PSP and a robust receiver, whenever conditions are identified to be unstable. These conditions are mostly dependent on the Doppler frequency. Since PSP with IIR-type channel estimation works very well for slow fading, it offers an appropriate choice for the robust reference receiver.

Finally, in section 6.6 the base station to mobile link (downlink) of a DS–CDMA system was studied under more realistic assumptions, taking into account the effects of multiple access interference (MAI). A received signal corrupted with MAI was applied to receiver designs investigated for a general diversity system (i.e. the single user case) from Chapters 4–6. Simulation results for the downlink suggest that the effects of MAI are far more severe on a frequency selective channel than for flat fading, due to the near-far effect and higher code correlation magnitudes of shifted spreading sequences. It was shown that MAI does not inflict further stability problems for decision directed receiver architectures.

7.2 Suggestions for further work

There are a number of points not covered within this thesis, which merit much more work.

Adaptive channel estimation: Little work has been done in this thesis to estimate the 2^{nd} order channel statistics. This may be done adaptively using the LMS or RLS adaptive algorithms [18]. It is desirable to update the coefficients of the channel estimation filter in a decision directed manner, otherwise periodic blocks of known pilots had to be transmitted to re-train the filter coefficients. However, decision directed training may cause degradation to the receiver performance which need to be assessed, and if necessary techniques are required which mitigate the effects due to decision directed training. Another possibility is the application of higher order Kalman filters [18]. Kalman filters do not require *a priori* information about the channel statistics, furthermore they are applicable to a non-stationary channel. Although Kalman filters have been applied to MLSD in several publications, for instance in [95–97], the study of error propagation effects may merit further research.

Hidden Markov models: The analysis of error propagation effects in section 6.3.3.2 was limited was limited to a 2–state Markov model, known as the Gilbert–Elliott channel (GEC). However, the GEC does

not incorporate any assumptions about the ratio of data to pilot symbols R. Hence, if the performance of the receiver is significantly dependent on R, the GEC cannot be used for the burst error analysis. This limitations may be overcome by employing a higher order hidden Markov model, which also accounts for R. On the other hand, for cases where the dependence of R on the performance was not dominant, results for the GEC were quite encouraging, so that higher order Markov models may be able deliver tighter performance bounds for arbitrary decision directed receivers operating in fading channels.

Extension to coded modulation: Virtually, every mobile radio system employs some kind of channel coding to improve the performance and thus increase the capacity. Particularly for soft-output decoding the task of the inner receiver and the channel decoder cannot be separated any longer and should be considered simultaneously. Moreover, due to the development of iterative decoding algorithms, the MAP symbol–by–symbol detector has gained more importance since it utilises and delivers soft inputs and soft outputs of the *a priori* and *a posteriori* probability of a symbol error, respectively. The MAP symbol–by–symbol detector and MLSD are conceptually similar algorithms, so MLSD receiver structures investigated in Chapters 5 and 6 may be converted for implementation of the MAP symbol–by–symbol detector.

Multi-user detection: The effects of MAI (see section 6.6) may be mitigated by means of multi-user detection. The algorithms for channel estimation and data detection addressed in this thesis may be combined with an multi-user detector. Moreover, the model for the received signal with multiple access interference, is idealised in the way that it does neglect inter-symbol interference (ISI) caused by the delay spread of the fading channel. The development of wideband CDMA [158] means that the delay spread of the multipath may not be negligible with respect to the symbol duration, resulting in significant ISI. Non-linear multi-user detectors, for instance based on radial basis functions (RBF), may also be combined with algorithms for channel estimation and detection in operating in a fast fading, frequency selective fading channels. Finally, the system model may be generalised to incorporate asynchronous transmission, thus extending the model to the uplink.

References

[1] M. Prögler, C. Evci, and M. Umehira, "Air Interface Access Schemes for Broadband Mobile Systems," *IEEE Communications Magazine*, vol. 37, pp. 106–115, Sep. 1999.

[2] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[3] T. Ojanperä and R. Prasad, "An Overview of Third Generation Wireless Personal Communications: A European Perspective," *IEEE Personal Communications*, pp. 59–65, Dec. 1998.

[4] A. Samukic, "UMTS Universal Mobile Telecommunications System: Development of Standards for the Third Generation," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 1099–1104, Nov. 1998.

[5] E. Dahlman, P. Beming, J. Knutsson, F. Ovesjö, M. Persson, and C. Roobol, "WCDMA - The Radio Interface for Future Mobile Multimedia Communications," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 1105–1118, November 1998.

[6] K. Ito, "Spreading and Modulation (TDD)," tech. rep., 3rd Generation Partnership Project (3GPP); Technical Secification Group (TSG) Radio Access Network (RAN); Working Group 1 (WG1), TS S1.23 V2.0.0 (1999-04) 1999.

[7] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, NY, USA, 3rd ed., 1995.

[8] D. P. Taylor, G. M. Vitetta, B. D. Hart, and A. Mämmelä, "Wireless Channel Equalisation," *European Transactions on Telecommunications*, vol. 9, pp. 155–172, Mar./Apr. 1998.

[9] K. S. Gilhousen, I. M. Jacobs, R. Padovani, and L. A. Weaver, Jr., "Increased Capacity using CDMA for Mobile Satellite Communication," *IEEE Journal Selected Areas Communication*, vol. 8, pp. 503–514, May 1990.

[10] M. L. Moher and J. H. Lodge, "TCMP—A Modulation and Coding Strategy for Rician Fading Channels," *IEEE Journal Selected Areas Communication*, vol. SAC-7, pp. 1347–55, Sep. 1989.

[11] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. VT-40, pp. 686–693, Nov. 1991.

[12] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, Jun. 1968.

[13] A. J. Viterbi, "Wireless Digital Communication: A View Based on Three Lessons Learned," *IEEE Communications Magazine*, pp. 33–36, Sep. 1991.

[14] J. D. Parsons, *The Mobile Radio Propagation Channel*. Pentech Press, London, 1992.

[15] G. L. Turin, "Introduction to Spread Spectrum Anti-multipath Techniques and their Application to Urban Digital Radio," *Proceedings of the IEEE*, vol. 68, pp. 328–353, Mar. 1980.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall International, 1993.

[17] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part 1*. Wiley, New York, 1968.

[18] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 2nd ed., 1991.

[19] L. L. Scharf, *Statistical Signal Processing; Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.

[20] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[21] N. Wiener, *The Extrapolation, Interpololation, and Smoothing of Stationary Time Series*. Wiley, New York, 1949.

[22] C. W. Therrien, "On the Relation Between Triangular Matrix Decomposition and Linear Prediction," *Proceedings of the IEEE*, vol. 71, pp. 1459–60, Dec. 1983.

[23] T. Kailath, "The Innovations Approach to Detection and Estimation Theory," *Proceedings of the IEEE*, vol. 58, pp. 680–95, 1970.

[24] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE*, vol. 63, pp. 561–80, Apr. 1987.

[25] P. M. Grant, C. F. N. Cowan, B. Mulgrew, and J. H. Dripps, *Analogue and Digital Signal Processing and Coding*. Bromley, UK: Chartwell-Bratt, 1989.

[26] N. Zhou and N. Holte, "Least Squares Channel Estimation for a Channel with Fast Time Variations," in *Proc. IEEE Int. Conference on the Applications of Speech and Signal Processing (ICASSP'92), San Francisco, CA, USA*, vol. 5, pp. V–165–168, Mar. 1992.

[27] R. E. Kalman, "A New Approach to Linear Filtering and Prediction," *Trans. ASME. J. Basic Engineering*, vol. 82, pp. 34–35, Mar. 1960.

[28] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. Wiley Interscience, 2nd ed., 1998.

[29] P. Y. Kam and S. Y. Tay, "Adaptive Reception of MPSK on Fading Channels," *IEE Electronics Letters*, vol. 30, pp. 1022–23, June 1994.

[30] F. Adachi, "Adaptive Reception of MPSK Signals in Fast Fading and AWGN Channels," *IEE Electronics Letters*, vol. 32, pp. 1944–45, Oct. 1996.

[31] T. Kailath, "Correlation Detection for Signal Perturbed by a Random Channel," *IRE Trans. Inf. Theory*, vol. IT-6, pp. 361–66, June 1960.

[32] T. Kailath, "Optimum Receivers for Randomly Varying Channels," in *Proceedings of the Fourth London Symposium of Information Theory*, pp. 109–22, Butterworth Scientific Press, London, UK, 1961.

[33] T. Kailath, *Adaptive Matched Filters*. University of California Press; Berkley, CA, 1963.

[34] T. Kailath, "A General Likelihood-Ratio Formula for Random Signals in Gaussian Noise," *IEEE Transactions on Information Theory*, vol. IT-15, pp. 350–61, May 1969.

[35] R. Häb and H. Meyr, "A Systematic Approach to Carrier Recovery and Detection of Digitally Phase Modulated Signals on Fading Channels," *IEEE Transactions on Communications*, vol. COM-37, pp. 748–54, July 1989.

[36] J. H. Lodge and M. L. Moher, "Maximum Likelihood Sequence Estimation of CMP Signals Transmitted over Rayleigh Flat Fading Channels," *IEEE Transactions on Communications*, vol. COM-38, pp. 787–94, June 1990.

[37] X. Yu and S. Pasupathy, "Innovations-Based MLSE for Rayleigh Fading Channels," *IEEE Transactions on Communications*, vol. COM-43, pp. 1534–44, Feb.-Apr. 1995.

[38] L. R. Bahl, J. Cocke, F. Jenilek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[39] P. Y. Kam, "Optimal Detection of Digital Data over the Nonselective Rayleigh Fading Channel with Diversity Reception," *IEEE Transactions on Communications*, vol. COM-39, pp. 214–219, Feb. 1991.

[40] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook (Revised Edition)*. McGraw–Hill, 1994.

[41] D. V. Sarwate and M. B. Pursley, "Cross-correlation Properties of Pseudorandom and Related Sequences," *Proceedings of the IEEE*, vol. 68, pp. 593–619, May 1980.

[42] F. J. MacWilliams and N. J. A. Sloane, "Pseudo-random Sequences and Arrays," *Proceedings of the IEEE*, vol. 64, pp. 1715–29, Dec. 1976.

[43] R. Gold, "Optimal Binary Sequences for Spread Spectrum Multiplexing," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 619–621, Oct. 1967.

[44] K. G. Beauchamp, *Walsh Functions and their Applications*. London: Academic Press, 1975.

[45] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation Measurements and Models for Wireless Communication Channels," *IEEE Communications Magazine*, vol. 1, pp. 42–49, Jan. 1995.

[46] P. A. Bello, "Characterisation of Randomly Time-variant Linear Channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 4, pp. 360–393, 1963.

[47] W. C. Jakes Jr, *Microwave Mobile Communications*. Wiley, NY, USA, 1974.

[48] W. C. Y. Lee, *Mobile Communications Engineering*, ch. 6. McGraw–Hill, NY, USA, 1982.

[49] M. Nakagami, "The m-Distribution, a General Formula of Intensity of Rapid Fading," in *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium held at the University of California* (W. Hoffman, ed.), pp. 3–36, Pergamon Press, 1960.

[50] U. Charash, "Reception Through Nakagami Fading Multipath Channels with Random Delays," *IEEE Transactions on Communications*, vol. COM-43, pp. 1937–45, May 1995.

[51] ETSI, "The ETSI UMTS terrestrial radio access (UTRA) ITU-R RTT candidate sumbmission," 1998.

[52] R. H. Clarke, "A Statistical Theory of Mobile-Radio Reception," *Bell System Technical Journal*, vol. 47, pp. 957–1000, Jul-Aug 1968.

[53] M. Failli, ed., *Digital Land Mobile Radio Communications - COST 207: Final Report*. Luxembourg: Commision of the European Communities, 1989.

[54] D. I. Laurenson, D. G. M. Cruickshank, and G. J. R. Povey, "A Computationally Efficient Multipath Channel Simulator for the COST 207 Models," *IEE Colloquium on Computer Modelling of Communication Systems*, May 1994.

[55] D. I. Laurenson and G. J. R. Povey, "Channel Modelling for a Predictive RAKE Receiver System," *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'94), The Hague, The Netherlands*, pp. 715–719, Sep 1994.

[56] R. Price and P. E. Green, "A Communications Technique for Multipath Channels," *Proceedings of the IRE*, vol. 2, pp. 555–70, Mar. 1958.

[57] D. G. Brennan, "Linear Diversity Combining Techniques," *Proc. IRE*, pp. 1075–1102, Jun. 1959.

[58] P. Y. Kam, "Adaptive Diversity Reception over a Slow Nonselective Fading Channel," *IEEE Transactions on Communications*, vol. COM-35, pp. 572–574, May 1987.

[59] K. Pahlavan and J. M. Matthews, "Performance of Adaptive Matched Filter Receivers Over Fading Multipath Channels," *IEEE Transactions on Communications*, vol. COM-38, pp. 2106–13, Dec. 1990.

[60] M. Barrett, "Error Probability for Optimal and Suboptimal Quadratic Receivers in Rapid Rayleigh Fading Channels," *IEEE Journal Selected Areas Communication*, vol. SAC-5, pp. 302–304, Feb. 1987.

[61] G. J. R. Povey, P. M. Grant, and R. D. Pringle, "A Decision Directed Spread Spectrum RAKE Receiver for Fast Fading Mobile Channels," *IEEE Transactions on Vehicular Technology*, vol. VT-45, pp. 491–502, Aug. 1996.

[62] H. Abeta, M. F. Sawahashi, and F. Adachi, "Performance Comparison between Time-Multiplexed Pilot Channel and Parallel Pilot Channel for Coherent Rake Combining in DS-CDMA Mobile Radio," *IEICE Transactions on Communications*, vol. E81-B, pp. 1417–25, July 1998.

[63] F. Sampei and T. Sunaga, "Rayleigh Fading Comansation for QAM in Land Mobile Radio Communications," *IEEE Transactions on Vehicular Technology*, vol. VT-42, pp. 137–147, May 1993.

[64] C. Liu and K. Feher, "Pilot-Symbol Aided Coherent M-ary PSK in Frequency-Selective Fast Rayleigh Fading Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 54–62, Jan. 1994.

[65] H. Andoh, M. F. Sawahashi, and F. Adachi, "Channel Estimation Filter Using Time-Multiplexed Pilot Channel for Coherent RAKE Combining in DS-CDMA Mobile Radio," *IEICE Transactions on Communications*, vol. E81-B, pp. 1517–26, July 1998.

[66] N. Kalouptsidis, *Signal Processing Systems*. Wiley Interscience, 1997.

[67] P. Schramm and R. R. Müller, "Pilot Symbol Assisted BPSK on Rayleigh Fading Channels with Diversity: Performance Analysis and Parameter Optimisation," *IEEE Transactions on Communications*, vol. COM-46, pp. 1561–63, Dec. 1998.

[68] V. P. Kaasila and A. Mämmelä, "Bit Error Probability for an Adaptive Diversity Receiver in a Rayleigh Fading Channel," *Proc. IEEE Int. Symposium on Spread Spectrum Techniques and Applications (ISSSTA'94), Oulu, Finland*, pp. 450–454, July 1994.

[69] W. H. Press, S. A. Teukoldy, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge University Press, 2nd ed., 1992.

[70] T. M. Schmidl, A. G. Dabak, and S. Hosur, "The use of Iterative Channel Estimation (ICE) to Improve Link Margin in Wideband CDMA Systems," in *Proc. IEEE Vehicular Technology Conference (VTC'99 spring), Houston, TX, USA*, vol. 2, pp. 1307–11, June 1999.

[71] D. I. Laurenson and G. J. R. Povey, "The Application of a Generalised Sliding FFT Algorithm to Prediction for a RAKE Receiver System Operationg Over Mobile Channels," *Proc. IEEE Int. Conference on Communications (ICC), Seatle, USA*, pp. 1823–27, June 1995.

[72] G. Auer, G. J. R. Povey, and D. I. Laurenson, "Robust Channel Predicton Technique for Decision Directed RAKE Receivers," *IEE Electronics Letters*, vol. 34, pp. 338–40, Feb. 1998.

[73] G. Auer, G. J. R. Povey, and D. I. Laurenson, "Mobile Channel Estimation for Decision Directed RAKE Receivers Operating in Fast Fading Radio Channel," in *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'98), Sun City, South Africa*, vol. 2, pp. 576–579, Sep. 1998.

[74] A. Mämmelä and V. P. Kaasila, "Prediction, Smoothing and Interpolation in Adaptive Diversity Reception," *Proc. IEEE Int. Symposium on Spread Spectrum Techniques and Applications (ISSSTA'94), Oulu, Finland*, pp. 475–478, July 1994.

[75] A. Mämmelä, *Diversity Receivers in a Fast Fading Multipath Channel*. PhD thesis, Univ. Oulu, Oulu, Finland, 1995.

[76] W. C. Dam and D. P. Taylor, "An Adaptive Maximum Likelihood Receiver for Correlated Rayleigh-Fading Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 2684–92, Sep. 1994.

[77] B. D. Hart, D. K. Borah, and S. Pasupathy, "Interpretation of the MLSD Metric for Time Varying, Frequency Selective Rayleigh Fading Channels," in *Proc. IEEE Vehicular Technology Conference (VTC'99 fall), Amsterdam, The Netherlands*, pp. 1321–25, Sep. 1999.

[78] G. L. Turin, "On Optimal Diversity Reception," *IRE Trans. Inf. Theory*, vol. IT-7, pp. 154–166, July 1961.

[79] G. L. Turin, "On Optimal Diversity Reception, II," *IRE Trans. Commun. Syst.*, vol. CS-10, pp. 22–31, Mar. 1962.

[80] T. Kailath, "On Multilink and Multidimensional Channels," *IRE Trans. Inf. Theory*, vol. IT-8, pp. 260–62, April 1962.

[81] J. G. Proakis, "Probabilities of Error for Adaptive Reception of M-Phase Signals," *IEEE Transactions on Communications*, vol. COM-16, pp. 71–81, Feb. 1968.

[82] G. L. Turin, "The Characteristic Function of Hermitian Quadratic Forms in Complex Normal Variables," *Biometrica*, vol. 47, pp. 199–201, June 1960.

[83] M. Stojanovic, J. G. Proakis, and J. A. Catipovic, "Analysis of the Impact of Channel Estimation Errors on the Performance of a Decison-Feedback Equalizer in Fading Multipath Channels," *IEEE Transactions on Communications*, vol. COM-43, pp. 877–886, Feb.-Apr. 1995.

[84] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–78, March 1973.

[85] G. D. Forney, "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Transactions on Information Theory*, vol. IT-18, pp. 363–378, May 1972.

[86] X. Yu and S. Pasupathy, "Error Performance of Innovations-Based MLSE for Rayleigh Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. VT-45, pp. 631–642, Nov. 1996.

[87] B. D. Hart and D. P.Taylor, "Extended MLSE Receiver for the Frequency-Flat, Fast-Fading Channel," *IEEE Transactions on Vehicular Technology*, vol. VT-46, pp. 381–89, May 1997.

[88] R. E. Morley and D. L. Snyder, "Maximum Likelihood Sequence Estimation of Randomly Dispersive Channels," *IEEE Transactions on Communications*, vol. COM-27, pp. 833–839, June 1979.

[89] K. M. Chugg, "The Conditions of the Applicability of the Viterbi Algorithm with Implications for Fading Channel MLSD," *IEEE Transactions on Communications*, vol. COM-46, pp. 1112–16, Sep. 1998.

[90] D. Makrakis, P. Mathiopoulos, and D. Bouras, "Optimal Decoding of Coded PSK and QAM Signals in Correlated Fast Fading Channels: a Combined Envelope, Multiple Differential and Coherent Detection Approach," *IEEE Transactions on Communications*, vol. COM-42, pp. 63–75, Jan. 1994.

[91] H. Kubo, K. Murakami, and T. Fujino, "Comparison on Maximum-Likelihood Sequence Estimation Schemes Incorporating Carrier Phase Estimation," *IEEE Transactions on Communications*, vol. COM-47, pp. 14–17, Jan. 1999.

[92] R. Raheli, A. Polydoros, and C. Tzou, "Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments," *IEEE Transactions on Communications*, vol. COM-43, pp. 354–364, Feb.-Apr. 1995.

[93] K. M. Chugg and A. Polydoros, "MLSE for an Unknown Channel — Part I: Optimality Considerations," *IEEE Transactions on Communications*, vol. COM-44, pp. 836–846, July 1996.

[94] G. Colavolpe, P. Castoldi, and R. Raheli, "Linear Predictive Receivers for Fading Channels," *IEE Electronics Letters*, vol. 34, pp. 1289–90, June 1998.

[95] D. Lee and J. Oh, "An Adaptive Innovations-Based MLSE Receiver for Frequency-Flat Rayleigh-Fading Channels," in *Proc. IEEE Vehicular Technology Conference (VTC'99 spring), Houston, TX, USA*, pp. 717–721, June 1999.

[96] Q. Dai and E. Shwedyk, "Detection of Bandlimited Signals over Frequency Selective ayleigh Fading Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 941–950, Feb./Mar./Apr. 1994.

[97] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian Algorithms for Blind Equalization Using Parallel Adaptive Filtering," *IEEE Transactions on Communications*, vol. COM-42, pp. 1017–32, Feb./Mar./Apr. 1994.

[98] G. M. Vitetta and D. P. Taylor, "Maximum Likelihood Decoding of Uncoded and Coded PSK Signal Sequences Transmitted over Rayleigh Flat-Fading Channels," *IEEE Transactions on Communications*, vol. COM-43, pp. 2750–58, Nov. 1995.

[99] D. K. Borah and B. D. Hart, "A Robust Receiver Structure for Time-Varying, Frequency-Flat, Rayleigh Fading Channel," *IEEE Transactions on Communications*, vol. COM-47, pp. 360–4, Mar. 1999.

[100] K. M. Chugg, "MLSE for an Unknown Channel — Part II: Tracking Performance," *IEEE Transactions on Communications*, vol. COM-44, pp. 949–958, Aug. 1996.

[101] P. Y. Kam and P. Sinha, "A Viterbi Type Algorithm for Efficient Estimation of M-PSK Sequences of the Gaussian Channel with Unknown Carrier Phase," *IEEE Transactions on Communications*, vol. COM-43, pp. 2429–33, Sep. 1995.

[102] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 260–269, 1967.

[103] R. E. Bellman and S. E. Dreyfus, *Applied Dynamic Programming*. Princeton, NJ: Princeton University Press, 1962.

[104] S. Verdú, "Minimum Probability of Error for Asynchronous Gaussian Multiple Access Channels," *IEEE Transactions on Information Theory*, vol. IT-32, pp. 85–96, Jan 1986.

[105] B. D. Hart and D. P.Taylor, "Maximum-Likelihood Synchronization, Equalization, and Sequence Estimation for Unknown Time-Varying Frequency-Selective Rician Channels," *IEEE Transactions on Communications*, vol. COM-46, pp. 211–221, Feb. 1998.

[106] G. Ungerböck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," *IEEE Transactions on Communications*, vol. COM-22, pp. 363–378, May 1974.

[107] P. Castoldi and R. Raheli, "On Recursive Optimal Detection of Linear Modulations in the Presence of Random Fading," *European Transactions on Telecommunications*, vol. 9, pp. 209–220, Mar./Apr. 1998.

[108] R. D'Avella, L. Moreno, and M. Sant'Agostino, "An Adaptive MLSE Receiver for TDMA Digital Mobile Radio," *IEEE Journal Selected Areas Communication*, vol. SAC-7, pp. 122–129, Jan. 1989.

[109] S. A. Fechtel and H. Meyr, "Optimal Parametric Feedforward Estimation of Frequency-Selective Fading Radio Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 1639–50, Feb.-Apr. 1994.

[110] F. R. Magee and J. G. Proakis, "Adaptive Maximum Likelihood Sequence Estimation for Digital Signalling in the Presence of Intersymbol Interference," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 120–124, Jan 1973.

[111] S. Qureshi, "Adaptive Equalisation," *Proceedings of the IEEE*, vol. 73, pp. 1349–1387, 1985.

[112] H. Kubo, K. Murakami, and T. Fujino, "Adaptive Maximum-Likelihood Sequence Estimator for Fast Time-Varying Intersymbol Interference Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 1872–80, Feb.-Apr. 1994.

[113] N. Seshadri, "Joint Data and Channel Estimation Using Blind Trellis Search Techniques," *IEEE Transactions on Communications*, vol. COM-42, pp. 1000–11, Feb.-Apr. 1994.

[114] A. P. Clark and S. Hariharan, "Efficient Estimators for an HF Radio Link," *IEEE Transactions on Communications*, vol. COM-38, pp. 1173–80, Aug. 1990.

[115] D. K. Borah and B. D. Hart, "MLSD-type Detection Using an Estimator for Time-Varying Frequency-Selective Channels," in *Proc. IEEE Vehicular Technology Conference (VTC'99 fall), Amsterdam, The Netherlands*, pp. 2403–7, Sep. 1999.

[116] J. B. Anderson and S. Mohan, "Sequential Decoding Algorithms: A Survey and Cost Analysis," *IEEE Transactions on Communications*, vol. COM-32, pp. 169–176, Feb.-Apr. 1984.

[117] T. Hashimoto, "A List-Type Reduced-Constraint Generalisation of the Viterbi Algorithm," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 866–76, Nov. 1987.

[118] M. V. Eyuboğlu and S. U. H. Qureshi, "Reduced-State Sequence Estimation with Set Partioning and Decision Feedback," *IEEE Transactions on Communications*, vol. COM-36, pp. 13–20, Jan. 1988.

[119] M. V. Eyuboğlu and S. U. H. Qureshi, "Reduced-State Sequence Estimation for Coded Modulation on Intersymbol Interference Channels," *IEEE Journal Selected Areas Communication*, vol. SAC-7, pp. 989–995, Aug. 1989.

[120] A. N. D'Andrea, A. Diglio, and U. Mengali, "Symbol-Aided Channel Estimation with Nonselective Rayleigh Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. VT-44, pp. 41–48, Feb. 1995.

[121] J. F. Hayes, T. M. Cover, and J. B. Riera, "Optimal Sequence Detection and Optimal Symbol-by-Symbol Detection: Similar Algorithms," *IEEE Transactions on Communications*, vol. COM-20, pp. 152–157, Jan. 1982.

[122] M. J. Gertsman and J. H. Lodge, "Symbol-by-Symbol MAP Demodulation of CPM and PSK Signals on Rayleigh Flat-Fading Channels," *IEEE Transactions on Communications*, vol. COM-45, pp. 788–99, July 1997.

[123] K. Abend and B. D. Fritchman, "Statistical Detection for Communication Channels with Intersymbol Interference," *Proceedings of the IEEE*, vol. 58, pp. 779–785, 1970.

[124] Y. Li, B. Vucetic, and Y. Sato, "Optimum Soft-output Detection for Channels with Intersymbol Interference," *IEEE Transactions on Information Theory*, vol. IT-41, pp. 704–713, May 1995.

[125] A. Anastasopoulos and A. Polydoros, "Adaptive Soft Decision Algorithms for Mobile Fading Channels," *European Transactions on Telecommunications*, vol. 9, pp. 183–190, Mar./Apr. 1998.

[126] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Soft-Input Soft-Output Modules for the Construction and Distributed Iterative Decoding of Code Networks," *European Transactions on Telecommunications*, vol. 9, pp. 155–172, Mar./Apr. 1998.

[127] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error Correcting Coding and ecoding: Turbo-Codes," in *Proc. IEEE Int. Conference on Communications (ICC), Geneva, Switzerland*, pp. 281–285, 1993.

[128] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and ecoding: Turbo-Codes," *IEEE Transactions on Communications*, vol. COM-44, pp. 1261–71, Oct. 1996.

[129] W. Koch and A. Baier, "Optimum and Sub-optimum Detection of Coded Data Disturbed by Time-varying Intersymbol Interference," in *Proc. IEEE Global Telecommunications Conference (Globecom '90). San Diego, CA, USA*, pp. 1679–84, Dez. 1990.

[130] J. A. Erfanian, S. Pasupathy, and G. Gulak, "Reduced Complexity Symbol Detectors with Parallel Structures for ISI Channels," *IEEE Transactions on Communications*, vol. COM-42, pp. 1661–71, Feb./Mar./Apr. 1994.

[131] R. Mehlan, J. Wittkopp, and H. Meyr, "Soft Output MAP Equalization and Trellis Coded Codulation for Severe Frequency-Selective Fading Channels," *Proc. IEEE Int. Conference on Communications (ICC), Chicago, IL, USA*, pp. 331.4.1–5, 1992.

[132] N. Seshadri and C. E. W. Sundberg, "List Viterbi Decoding Algorithms with Applications," *IEEE Transactions on Communications*, vol. COM-42, pp. 313–323, Feb.-Apr. 1994.

[133] J. Hagenauer and P. Höher, "Viterbi Algorithm with Soft-Decision Outputs and its Applications," in *Proc. IEEE Global Telecommunications Conference (Globecom '89). Dallas, TX, USA*, vol. 3, pp. 1680–86, Nov. 1989.

[134] P. Robertson, E. Villebrun, and P. Höher, "A Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log Domain," in *Proceedings of the 1995 IEEE International Conference on Communications (ICC '95), Seattle, WA, USA*, vol. 2, pp. 1009–13, Jun. 1995.

[135] M. P. C. Fossorier, F. Burkert, and J. Hagenauer, "On the Equivalence Between SOVA and Max–Log–MAP Decodings," *IEEE Communications Letters*, vol. 2, pp. 137–9, May 1998.

[136] P. Höher, "Advances in Soft-Output Decoding," in *Proc. IEEE Global Telecommunications Conference (Globecom '93). Houston, TX, USA*, vol. 2, pp. 793–797, Nov. 1993.

[137] P. Y. Kam and A. T. L. Lai, "Adaptive Viterbi Estimation of MPSK Sequences over a Rayleigh Fading Channel," *IEE Electronics Letters*, vol. 31, pp. 2142–43, Dec. 1995.

[138] G. Auer, G. J. R. Povey, and D. I. Laurenson, "On Unknown Parameter Estimation for MLSE on Flat Fading Channels," *Submitted to IEE Proceedings Communications*, 1999.

[139] J. Salz, "Optimum Mean-square Decision Feedback Equalization," *Bell System Technical Journal*, vol. 52, pp. 1341–73, Oct. 1973.

[140] G. M. Vitetta and D. P. Taylor, "Viterbi Decoding of Differentially Encoded PSK Signal Transmitted over Rayleigh Frequency-Flat Fading Channels," *IEEE Transactions on Communications*, vol. COM-43, pp. 1256–1259, Feb.-Apr. 1995.

[141] G. T. Irvine and P. J. McLane, "Symbol Aided Plus Decision-Directed Reception for PSK/TCM Modulation on Shadowed Mobile Satellite Fading Channels," *IEEE Journal Selected Areas Communication*, vol. SAC-10, pp. 1289–98, Oct. 1992.

[142] Y. Liu and S. D. Blostein, "Identification of Frequency Non-selective Fading Channels Using Decision Feedback and Adaptive Linear Prediction," *IEEE Transactions on Communications*, vol. COM-43, pp. 1484–92, Feb.-Apr. 1995.

[143] G. Auer, G. J. R. Povey, and D. I. Laurenson, "Per-Survivor Processing Applied to Decision Directed Channel Estimation for a Coherent Diversity Receiver," in *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'98), Sun City, South Africa*, vol. 2, pp. 580–584, Sep. 1998.

[144] G. Auer, G. J. R. Povey, and D. I. Laurenson, "Hybrid Filtering for Per-Survivor Processing in Flat Fading Channels," in *Proc. IEEE Vehicular Technology Conference (VTC'99 fall), Amsterdam, The Netherlands*, Sep. 1999.

[145] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, pp. 4–16, Jan. 1986.

[146] E. N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell System Technical Journal*, vol. 39, pp. 1253–65, Sep. 1960.

[147] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell System Technical Journal*, vol. 42, pp. 1977–97, Sep. 1963.

[148] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, NY, USA, 3rd ed., 1991.

[149] S. Sivaprakasam and K. S. Shanmugan, "Finite-State Markov Channel — A Useful Model for Radio Communication Channels," *IEEE Transactions on Vehicular Technology*, vol. VT-44, pp. 163–171, Feb. 1995.

[150] B. D. Fritchman, "A Binary Channel Characterization Using Partitioned Markov Chains," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 221–227, Apr. 1967.

[151] S. Sivaprakasam and K. S. Shanmugan, "An Equivalent Markov Model for Burst Errors in Digital Channels," *IEEE Transactions on Communications*, vol. COM-43, pp. 885–899, Feb.-Apr. 1995.

[152] C. Pimentel and J. F. Blake, "Modelling Burst-Noise Channels Using Partitioned Fritchman's Markov Models," *IEEE Transactions on Vehicular Technology*, vol. VT-47, pp. 885–899, Aug. 1998.

[153] A. Baier and G. Heinrich, "Performcnce of M-Algorithm MLSE Equalizers in Frequency Selective Fading Mobile Radio Channels," in *Proc. IEEE Int. Conference on Communications (ICC), Boston, MA, USA*, pp. 281–285, 1989.

[154] A. P. Clark and S. M. Asghar, "Detection of Digital Signals Transmitted over a Known Time-Varying Channel," *Proceedings of the IEE*, vol. 128, pp. 167–174, Feb. 1981.

[155] J. Liu, S. C. Kwatra, and J. Kim, "An Analysis of Decision Feedback Detection of Differentially Encoded MPSK Signals," *IEEE Transactions on Vehicular Technology*, vol. VT-44, pp. 261–267, May 1995.

[156] S. Verdú, *Multiuser Detection*. Cambridge: University Press, 1998.

[157] M. Kavehrad, "Performance of Nondiversity Receivers for Spread Spectrum in Indoor Wireless Communications," *AT&T Technical Jounal*, vol. 64, pp. 1181–85, July–Aug. 1985.

[158] T. S. G. T. R. WG1, "Physical Channel and Mapping of Transport Channels onto Physical Channels," tech. rep., 3rd Generation Partnership Project (3GPP), April, 1999.

[159] S. L. Marple, *Digital Spectrum Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1987.

Appendix A
Derivations for the Optimum Receiver

A.1 Invariance over pre-multiplication

Starting with the pre-multiplied observation $\mathbf{y}'_q(\ell) = \mathbf{D}^{(\ell)H}\mathbf{y}_q$ from (5.9) in section 5.1.4, it shall be shown that the decision variable of the optimum detector–estimator is invariant over pre-multiplication. Therefore, the equivalence of the quadratic terms $\mathbf{y}_q^H \boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1}\mathbf{y}_q$ and $\mathbf{y}_q'^H(\ell)\,\boldsymbol{\Phi}_{\mathbf{y'y'},q}^{-1}\mathbf{y}'_q(\ell)$ is to be shown.

The covariance matrix of $\mathbf{y}_q$, $\boldsymbol{\Phi}_{\mathbf{yy},q}(\ell)$ according to (2.10), has the form [76]

$$\boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)} = \left[\mathbf{D}^{(\ell)}\,\boldsymbol{\Phi}_{\mathbf{hh},q}\,\mathbf{D}^{(\ell)H} + N_0\mathbf{I}\right]$$

Provided that $\mathrm{D}^{(\ell)}\,\mathbf{D}^{(\ell)H} = \mathbf{I}$, which is always valid with the assumptions in section 5.1.4, the inverse of $\boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)}$ can be rewritten as

$$\begin{aligned}
\boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1} &= \left[\mathbf{D}^{(\ell)}\,\boldsymbol{\Phi}_{\mathbf{hh},q}\,\mathbf{D}^{(\ell)H} + N_0\,\mathbf{D}^{(\ell)}\,\mathbf{D}^{(\ell)H}\right]^{-1} \\
&= \left[\mathbf{D}^{(\ell)}\left(\boldsymbol{\Phi}_{\mathbf{hh},q} + N_0\,\mathbf{I}\right)\mathbf{D}^{(\ell)H}\right]^{-1} \\
&= \left[\mathbf{D}^{(\ell)H}\right]^{-1}\left[\boldsymbol{\Phi}_{\mathbf{hh},q} + N_0\,\mathbf{I}\right]^{-1}\mathbf{D}^{(\ell)-1} \\
&= \mathbf{D}^{(\ell)}\left[\boldsymbol{\Phi}_{\mathbf{hh},q} + N_0\mathbf{I}\right]^{-1}\mathbf{D}^{(\ell)H} \qquad \text{(A.1)}
\end{aligned}$$

where the dependence of the matrix $\boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1}$ on ℓ has been conveniently factored out.

The quadratic form of (5.8) for tap q is given by

$$\Lambda_q(\ell) = \mathbf{y}_q^H\,\boldsymbol{\Phi}_{\mathbf{yy},q}^{(\ell)\,-1}\mathbf{y}_q$$

With the definition of the pre-multiplied observation $\mathbf{y}'_q(\ell) = \mathbf{D}^{(\ell)H}\mathbf{y}_q$ and inserting (A.1) into the above expression for $\Lambda_q(\ell)$, the following is obtained

$$\begin{aligned}
\Lambda_q(\ell) &= \mathbf{y}_q^H\,\mathbf{D}^{(\ell)}\left[\boldsymbol{\Phi}_{\mathbf{hh},q} + N_0\mathbf{I}\right]^{-1}\mathbf{D}^{(\ell)H}\,\mathbf{y} \\
&= \left(\mathbf{D}^{(\ell)H}\,\mathbf{y}_q\right)^H\left[\boldsymbol{\Phi}_{\mathbf{hh},q} + N_0\mathbf{I}\right]^{-1}\mathbf{D}^{(\ell)H}\,\mathbf{y} \\
&= \mathbf{y}_{\mathbf{q}}'^H(\ell)\,\boldsymbol{\Phi}_{\mathbf{y'y'},q}^{-1}\,\mathbf{y}'_q(\ell) \qquad \text{(A.2)}
\end{aligned}$$

Finally, with $\Lambda(\ell) = \sum_{q=1}^{Q}\Lambda_q(\ell)$, the result of (5.10) is verified.

A.2 Generalised eigenvalue decomposition

We are concerned with evaluating the eigenvalues of a product of two Hermitian matrices, being in general a non-symmetric matrix $\mathbf{\Phi Q}$. Its eigenvalues λ must satisfy

$$\mathbf{\Phi Q u} = \lambda \mathbf{u}$$

where $\mathbf{u}$ is the associated eigenvector. Since both matrices $\mathbf{Q}$ and $\mathbf{\Phi}$ are Hermitian, the eigenvalue decomposition of $\mathbf{\Phi Q}$ can be circumvented by the generalised eigenvalue problem,[1] giving the equivalent eigenvalue problem treated in [69]

$$\mathbf{Q u} = \lambda \mathbf{\Phi}^{-1} \mathbf{u} \tag{A.3}$$

A symmetrical eigenvalue problem can be recovered by using the Cholesky decomposition [159]

$$\mathbf{\Phi} \triangleq \mathbf{A A}^H \quad \Longleftrightarrow \quad \mathbf{\Phi}^{-1} = \left[\mathbf{A A}^H\right]^{-1} \triangleq \mathbf{B B}^H$$

where the matrices $\mathbf{A}$ and $\mathbf{B}$ have a lower triangular form, referred to as the square-root matrices of $\mathbf{\Phi}$ and $\mathbf{\Phi}^{-1}$, respectively. With this definition, the problem in (A.3) can be cast in the form

$$\begin{aligned} \mathbf{C}\left(\mathbf{B}^H \mathbf{u}\right) &= \lambda\left(\mathbf{B}^H \mathbf{u}\right) \; ; \quad \mathbf{v} \triangleq \mathbf{B}^H \mathbf{u} \\ \mathbf{C} &= \mathbf{B}^{-1} \mathbf{Q} \left[\mathbf{B}^{-1}\right]^H = \mathbf{A}^H \mathbf{Q}\, \mathbf{A} \end{aligned} \tag{A.4}$$

The eigenvalues, λ, of the Hermitian matrix $\mathbf{C}$ are identical with the ones of the original problem in (A.3), with the associated eigenvectors $\mathbf{v}$, yielding an equivalent eigenvalue problem

$$\mathbf{C v} = \lambda \mathbf{v} \,, \qquad \text{with} \quad \mathbf{u} = \mathbf{A v}$$

The desired eigenvectors are obtained by $\mathbf{u} = \mathbf{A v}$.

It can be observed that neither the square-root matrix $\mathbf{B}$ needs to be calculated nor the covariance matrix $\mathbf{\Phi}$ to be inverted.

A.3 Derivations of the quadratic form for performance analysis

In this appendix the quadratic forms $\mathbf{u}^H \mathbf{Q u}$ and the filter matrix $\mathbf{Q}$ are specified for some receiver realisations.

[1] The matrix $\mathbf{\Phi}$must be non singular. A sufficient condition for that is its positive definiteness, which is given for virtual all noise corrupted covariance matrices [159].

A.3.1 Pilot aided receiver

The derivation of the quadratic form for the pilot aided (PA) receiver which entirely relies on pilot symbols to estimate the channel response is considered. To evaluate the error probability the decision variable in (4.13) needs to be cast into a quadratic form. The derivation with the q^{th} fading tap is considered first, later this is generalised to a diversity receiver. The decision variable of the q^{th} diversity tap is defined by

$$\Lambda_q(k,r) = \mathbf{u}_q^H(k,r)\,\mathbf{Q}_q(r)\,\mathbf{u}_q(k,r)$$

where the row vector $\mathbf{u}_q(k,r)$ of dimension $\mathbb{C}^{M+1}$, contains the bit to be detected and the pilots used for channel estimation. The observation $\mathbf{u}_q(k,r)$ and the $(M+1)\times(M+1)$ matrix $Q_q(r)$ are defined by

$$\mathbf{u}_q(k,r) = \begin{bmatrix} y_q(k-M_f R) \\ \mathbf{y}_{Pq}(\kappa) \end{bmatrix} \qquad \mathbf{Q}_q(r) = \begin{bmatrix} 0 & \mathbf{w}_r^{H(\Delta\kappa)} \\ \mathbf{w}_r^{(\Delta\kappa)} & \mathbf{0} \end{bmatrix}$$

The statistics of $\mathbf{u}_q(k,r)$ are specified by its time independent auto correlation matrix

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{uu},q}(r) &= \mathrm{E}\big[\mathbf{u}_q(k,r)\,\mathbf{u}_q^H(k,r)\big] \\ &= \begin{bmatrix} \mathrm{E}\big[y_q(k)\,y_q^*(k)\big] & \mathrm{E}\big[y_q(k)\,\mathbf{y}_{Pq}^H(\kappa)\big] \\ \mathrm{E}\big[y_q^*(k)\,\mathbf{y}_{Pq}(\kappa)\big] & \mathrm{E}\big[\mathbf{y}_{Pq}(\kappa)\,\mathbf{y}_{Pq}^H(\kappa)\big] \end{bmatrix} = \begin{bmatrix} 1+\frac{1}{\bar{\gamma}_q} & d(k)\,\phi_{Pr}^{H(\Delta\kappa)} \\ d^*(k)\,\phi_{Pr}^{(\Delta\kappa)} & \mathbf{\Phi}_P \end{bmatrix} \end{aligned}$$

where $\mathbf{\Phi}_P$ denotes the covariance matrix of the pilots defined in (4.9) and $\phi_{Pr}^{(\Delta\kappa)}$ is the cross correlation vector between the desired bit and the pilots $\mathbf{y}_{Pq}(\kappa)$ of (4.10). It is seen, that $\mathbf{\Phi}_{\mathbf{uu},q}(r)$ is dependent on the bit being transmitted, but as the test is symmetric, the error probability does not depend on $d(k)$.

Diversity reception Using the independent fading assumption for the combining of the Q diversity taps the following decision variable is obtained

$$\begin{aligned} \Lambda(k,r) &= \sum_{q=1}^{Q} \Lambda_q(k,r) = \sum_{q=1}^{Q} \mathbf{u}_q^H(k,r)\,\mathbf{Q}_q(r)\,\mathbf{u}_q(k,r) \\ &= \mathbf{u}^H(k,r)\,\mathbf{Q}(r)\,\mathbf{u}(k,r) \end{aligned} \tag{A.5}$$

and choosing $\widehat{d}(k) = 1$ if $\Lambda_q(k,r)$ is larger than zero and $\widehat{d}(k) = -1$ otherwise. So, the probability of error for one bit in the sequence corresponds to the probability that $\widehat{d}(k) = -1$ was chosen, when $\widehat{d}(k) = 1$ was transmitted. Denoting the dimension of $\mathbf{u}(k,r)$ by $N = Q(M+1)$ the quantities $\mathbf{u}(k,r)$ and $\mathbf{Q}(r)$ are defined by

$$\mathbf{u}(k,r) = \begin{bmatrix} \mathbf{u}_1(k,r) \\ \mathbf{u}_2(k,r) \\ \vdots \\ \mathbf{u}_Q(k,r) \end{bmatrix} \qquad \mathbf{Q}(r) = \begin{bmatrix} \mathbf{Q}_1(r) & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_2(r) & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{Q}_Q(r) \end{bmatrix}$$

The covariance matrix of $\mathbf{u}(k,r)$ is

$$\mathbf{\Phi_{uu}}(r) = \begin{bmatrix} \mathbf{\Phi_{uu,1}}(r) & 0 & \cdots & 0 \\ 0 & \mathbf{\Phi_{uu,2}}(r) & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{\Phi_{uu,}}_Q(r) \end{bmatrix} \in \mathbb{C}^{N\times N}$$

A.3.2 Recursive MLSD

Our aim is to transform the decision variable $\Delta\Lambda(\mathcal{E})$ of an error event $\mathcal{E}$, to a quadratic form $\mathbf{u}^H\mathbf{Q}(\mathcal{E})\mathbf{u}$ as in (5.39). We begin our derivation with deriving an expression for a FIR channel estimation filter. Later an approximation for IIR filtering will also be given. Furthermore, let us consider the q^{th} fading tap, the generalisation to a diversity receiver will follow later on.

FIR filtering: The decision variable of the q^{th} diversity tap, $\Lambda_q(\ell)$, taken from (5.24) is given by

$$\Lambda_q(\ell) = \sum_{k=1}^{K} \left| y_q(k) - d^{(\iota)}(k)\,\widehat{h}_q(\ell,k) \right|^2 = \sum_{k=1}^{K} \left| \epsilon_q(\ell,k) \right|^2 \tag{A.6}$$

where $\epsilon_q(\ell,k)$ denotes the prediction error or the Euclidean distance. The mean squared value of $\epsilon_q(\ell,k)$ is equivalent to the MSE used to find the estimation filter coefficients $\{w_m\}$. With the definition $w_0 \triangleq -1$ the prediction error can be re-written as

$$\left|\epsilon_q(\ell,k)\right|^2 = \left| \sum_{m=0}^{M} w_m^*\, y'_q(\ell,k) \right|^2 \Bigg|_{w_0 \triangleq -1} \tag{A.7}$$

With the filter matrix $\mathbf{w} = [-1, w_1, \cdots, w_M]^T$, the prediction error can be more conveniently expressed in vector notation: $\epsilon_q(\ell,k) = \mathbf{w}^H\mathbf{y}'_q(\ell,k)$. After straightforward transformations and recalling the definition of $\mathbf{y}'_q(\ell,k) = \mathbf{D}^{(\iota)H}(k)\,\mathbf{y}_q(k)$ the following is obtained

$$\left|\epsilon_q(\ell,k)\right|^2 = \mathbf{y}_q^H(k)\,\mathbf{D}^{(\iota)}(k)\;\mathbf{w}\mathbf{w}^H\,\mathbf{D}^{(\iota)H}(k)\,\mathbf{y}_q(k) \tag{A.8}$$

where all used vector quantities are assumed to have the appropriate dimension of $M+1$, being truncated versions of the entire signals. The data matrix $\mathbf{D}^{(\iota)}(k)$ of dimension $M+1 \times M+1$ is defined as $\text{diag}[d^{(\iota)}(k-M), \cdots, d^{(\iota)}(k)]$.

The LRT [17] of (5.39) calls for the difference of the two tested decision variables, $\Delta\Lambda_q(\mathcal{E})$. Using the result of (A.8), $\Delta\Lambda_q(\mathcal{E})$ can be expressed as

$$\begin{aligned} \Delta\Lambda_q(\mathcal{E}) &= \Lambda_q(\ell) - \Lambda_q(0) \\ &= \sum_{k=1}^{K} \mathbf{y}_q^H(k)\,\Omega_q(\mathcal{E},k)\,\mathbf{y}_q(k) \end{aligned} \tag{A.9}$$

where

$$\Omega_q(\mathcal{E},k) = \mathbf{D}^{(\ell)}(k)\,\mathbf{w}\mathbf{w}^H\,\mathbf{D}^{(\ell)H}(k) - \mathbf{D}^{(0)}(k)\,\mathbf{w}\mathbf{w}^H\,\mathbf{D}^{(0)H}(k) \quad \in \mathbb{C}^{(M+1)\times(M+1)}$$

The pairwise probability of detecting hypthesis ℓ in favour of hypothesis 0, is the probability that $\Delta\Lambda_q(\mathcal{E}) < 0$. The notation in (A.9) is already a quadratic form, however, it is an unbounded sum.

Note, $\Delta\Lambda_q(\mathcal{E})$ is only non-zero, where the hypotheses $\widehat{d}(k)$ is different from the transmitted sequence $d(k)$. Recall the definition of the state sequence $\chi_e(k) = \{e^{(\ell)}(k),\cdots,e^{(\ell)}(k-M)\}$ as shift register process of the most likely sequence. By observing $\chi_e(k)$ it can be seen that the summation over k in (A.9) has only a finite number of non-zero elements. Let k_1 denote the time where the first decision error is observed, then the samples where $\Delta\Lambda_q(\mathcal{E})$ is essentially non-zero, are in the range $k \in \{k_1,\cdots,k_1+L_b+M\}$, as illustrated in Figure 5.9. Without loss of generality let $k_1 = M$, then $\Delta\Lambda_q(\mathcal{E})$ becomes

$$\Delta\Lambda_q(\mathcal{E}) \;=\; \sum_{k=M}^{L_e+M} \mathbf{y}_q^H(k)\,\Omega_q(\mathcal{E},k)\,\mathbf{y}_q(k) \tag{A.10}$$

with $L_e = L_b + M$.

A notation of $\Delta\Lambda_q(\mathcal{E}) = \mathbf{u}_q^H\,\mathbf{Q}_q(\mathcal{E})\,\mathbf{u}_q$ of the required form of (5.39) can be obtained by re-arranging (A.10). The matrix containing the received signal samples, $\mathbf{u}_q = [y_q(k_1-M),\cdots,y_q(k_1+M+1)]^T$, is defined in (5.40). The matrix $\mathbf{Q}_q(\mathcal{E})$ emerges from the set of matrices $\{\Omega_q(\mathcal{E},k);\, k = M,\cdots,L_e+M\}$ in order to satisfy (A.9), that is

$$\begin{aligned}\mathbf{Q}_q(\mathcal{E}) &= \sum_{k=M}^{L_e+M}\left[\mathbf{D}_c^{(\ell)}\mathbf{v}(k)\mathbf{v}^H(k)\mathbf{D}_c^{(\ell)H} - \mathbf{D}_c^{(0)}\mathbf{v}(k)\mathbf{v}^H(k)\mathbf{D}_c^{(0)H}\right]\\ &= \mathbf{D}_c^{(\ell)}\left[\sum_{k=M}^{L_e+M}\mathbf{v}(k)\mathbf{v}^H(k)\right]\mathbf{D}_c^{(\ell)H} - \mathbf{D}_c^{(0)}\left[\sum_{k=M}^{L_e+M}\mathbf{v}(k)\mathbf{v}^H(k)\right]\mathbf{D}_c^{(0)H}\end{aligned} \tag{A.11}$$

The matrix $\mathbf{Q}_q(\mathcal{E})$ is seen to be a composition of the data matrix $\mathbf{D}_c$ and the vector $\mathbf{v}(k)$. Assuming the transmitted sequence is the all one sequence and $\mathbf{d}^{(\ell)} = \mathbf{d}^{(0)}$, the diagonal data matrix is now of the time independent form

$$\mathbf{D}_c^{(\ell)} \;=\; \mathrm{diag}\Big[\underbrace{1,\cdots,1}_{M},\, d^{(\ell)}(M),\cdots,d^{(\ell)}(M+L_b),\underbrace{1,\cdots,1}_{M}\Big]$$

The filter vector has become time dependent and is defined by

$$\mathbf{v}(k) \;=\; [\underbrace{0,\cdots,0}_{k-M},\,\underbrace{w_M,\cdots,w_1,-1}_{\mathbf{w}^T},\,\underbrace{0,\cdots,0}_{L_e-k+M}]^T$$

It can be shown that $\Delta\Lambda_q(\mathcal{E})$ is not changed by the transformation from (A.10) to (A.11). The dependency of time k has been changed from the received signal to the filter matrix. The filter vector $\mathbf{w}$ moves one step to the right every sample, within $\mathbf{v}(k)$, while $\mathbf{u}_q$ and $\mathbf{D}_c^{(\ell)}$ on the other hand, are fixed relative to time k. All vector and matrix quantities are of dimension L_{tot} and $L_{tot} \times L_{tot}$ respectively.

IIR filtering: If an α–tracker is used to generate the CIR estimate, $\widehat{h}_q(\ell, k+1) = (1-\alpha)\, y'_q(\ell, k) + \alpha\, \widehat{h}_q(\ell, k)$ one way to evaluate the MSE $|\epsilon_q(\ell, k)|^2$ is to use an expression given [28]. However, this requires us to re-formulate the decision variable (A.6) for the FIR case. So it is desirable to expand the α–tracker estimate of (5.30) to a series

$$\widehat{h}_q(\ell, k+1) \;=\; \frac{1-\alpha}{\alpha} \sum_{m=0}^{k} \alpha^m y'_q(\ell, k-m)\;; \qquad 0 \le \alpha < 1\,. \tag{A.12}$$

It is seen that the influence of $y'_q(\ell, k-m)$ decreases exponentially with delay m. Due to the exponential decline of the terms contributing to (A.12), the summation over k samples in (A.12) may be truncated after a M symbol delay, with M sufficiently large to obtain

$$\widehat{h}_q(\ell, k) \;=\; \frac{1-\alpha}{\alpha} \sum_{m=1}^{M} \alpha^m y'_q(\ell, k-m)\;; \qquad 0 \le \alpha < 1\,. \tag{A.13}$$

Thus, the IIR filter in (A.12) may be viewed as a M^{th} order FIR filter, with coefficients

$$w_m \;=\; \frac{1-\alpha}{\alpha}\alpha^m \;=\; \alpha^{m-1} - \alpha^m \tag{A.14}$$

With this expression the recursive MLSE with IIR filtering can be cast into decision variable (A.6). The corresponding error probability for this case is obtained by proceeding in the same way as demonstrated for the FIR filter case.

Diversity Reception: After an expression of the quadratic form for the flat fading channel has been found, these results can be easily generalised to the case of diversity reception. To accommodate the diversity receiver in the decision variable of (5.39) we employ Kailath's diversity receiver for multilink and multidimensional channels [80]. The quadratic form of the LRT of (5.39) for a Q tap diversity receiver is

$$\Delta\Lambda(\mathcal{E}) \;=\; \sum_{q=1}^{Q} \Delta\Lambda_q(\mathcal{E}) = \mathbf{u}^H\, \mathbf{Q}(\mathcal{E})\, \mathbf{u} \tag{A.15}$$

Following [80] the quantities $\mathbf{u}$ and $\mathbf{Q}(\mathcal{E})$ are given by

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_Q \end{bmatrix} \qquad \mathbf{Q}(\mathcal{E}) = \begin{bmatrix} \mathbf{Q}_1(\mathcal{E}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2(\mathcal{E}) & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_Q(\mathcal{E}) \end{bmatrix}$$

where the dimension of $\mathbf{u}$ and $\mathbf{Q}(\mathcal{E})$ are denoted by $N = QL_{tot}$ and $N \times N$ respectively. The matrix $\mathbf{Q}(\mathcal{E})$ has a block diagonal structures whose elements are themselves matrices. The elements $\mathbf{Q}_q(\mathcal{E})$ and $\mathbf{u}_q$ are given by (A.11) and (5.40) respectively.

Appendix B
Additional Plots

B.1 Parameter optimisation for hybrid filters

B.1.1 Emphasis the pilot symbols for channel estimation

In order to find the optimum fiter parameters ν_l and ν_h for the receiver studied in 6.5.2.1, the soft PA-PSP was employed in channels with different parameters. In general, the same parameters as in 6.5 apply.

The performance of the PA–PSP is depicted in Figure B.1, for different filter parameters ν_l and ν_h. The constant ν_l and ν_h are varied in part (a.) and (b.) respectively. The hybrid receiver is seen to be robust for arbitrary fading conditions. However, the hybrid receiver cannot outperform the entirely pilot aided (PA) receiver for slow fading. There is a slight performance advantage when the FIR–PSP and PA receiver performance are similar ($0.01 < \nu'_{max} < 0.02$). The constants ν_l and ν_h should be chosen such that the μ is changing gradually between slow and fast fading receiver. Otherwise, if μ switches on/off ($\nu_l = \nu_h$), the point where the performance of the fast and slow fading receiver are equal needs to be know precisely. This may result in a performance degradation around the switching point, as it is shown in the graph for $\nu_l = \nu_h = 0.02$.

B.1.2 Hybrid filtering MLSD

We are interested to find the switching parameters in terms of the SNR, $\gamma(k)$, of hybrid per-survivor processing (HPSP) from 6.5.2.2. The SNR may strongly deviate from its mean $\bar{\gamma}$. Furthermore, estimates of the current SNR, $\widehat{\gamma}(k)$ can be generated easily as the channel statistics are monitored extensively

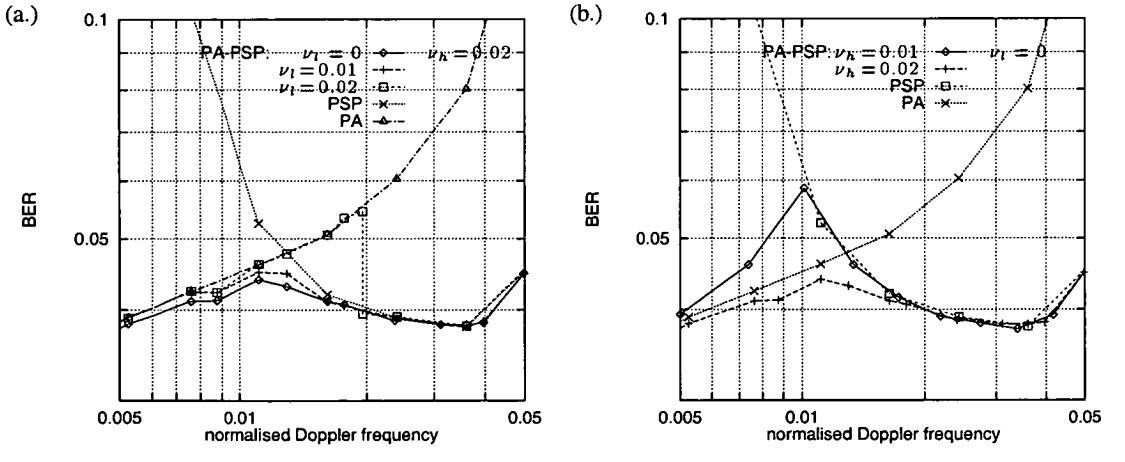


Figure B.1: *BER vs normalised Doppler frequency, ν'_{max}, for PA–PSP.*
(a.) $\nu_h = 0.02$, various ν_l; *(b.) $\nu_l = 0$, various ν_h;*
$R = 10$, $M = 8$, $Q = 1$, $\bar{\gamma} = 10$ dB.

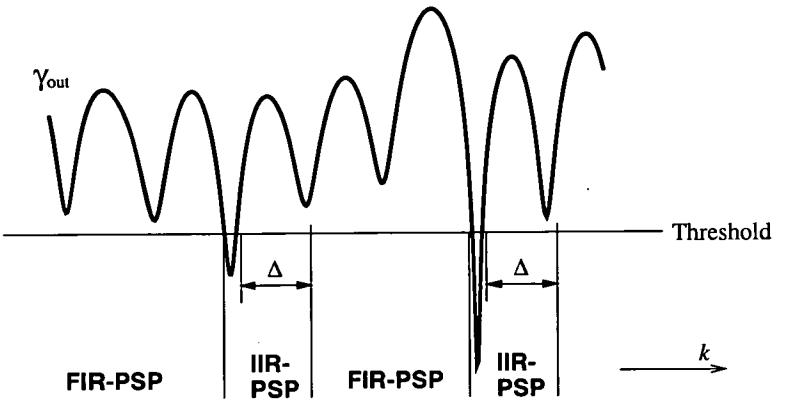


Figure B.2: *Switching between the FIR and IIR–PSP dependent on the threshold SNR, γ_{th}.*

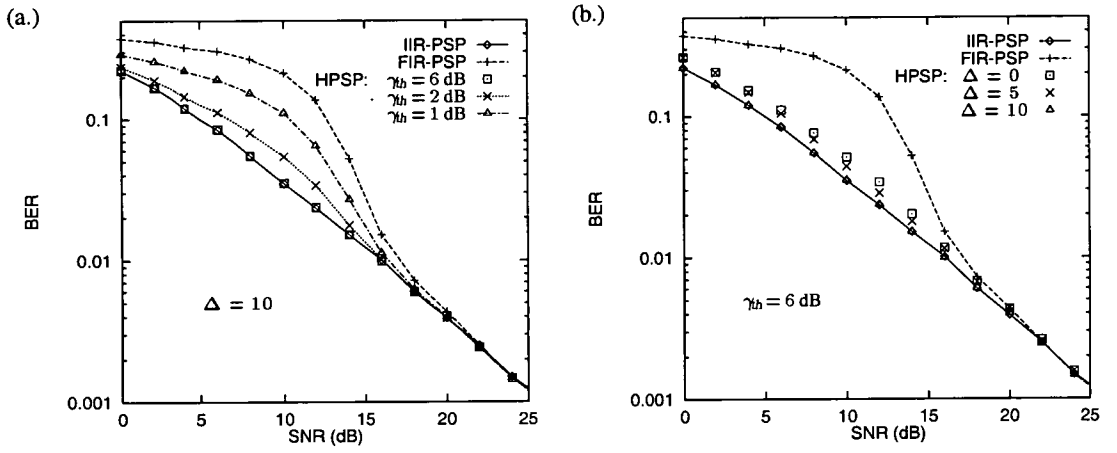


Figure B.3: *BER vs SNR of the HPSP for different ratios of data to known symbols.*
$\nu'_{max} = 0.005$, $R = 10$, $M = 8$, $Q = 1$;
(a.) $\Delta = 10$, *various* γ_{th}; *(b.)* $\gamma_{th} = 6$ *dB, various* Δ.

already for channel estimation. Thus for Doppler frequencies where the IIR–PSP is superior for low SNR and the FIR–PSP for high SNR, typically found in mid fading rates ($0.01 > \nu'_{max} > 0.04$), the receiver needs to switch between FIR and IIR filter. The SNR dependence of μ in (6.13) is taken into account by the threshold γ_{th}. When the received signal is in a deep fade the receiver switches from FIR to IIR channel estimation. As the phase slips of the FIR–PSP may persist through the subsequent good reception area, an time delay Δ is imposed, in order to prevent that. This is illustrated in Figure B.2. In terms of the Doppler frequency ν'_{max}, the threshold γ_{th} is lowered for increasing Doppler, since the FIR–PSP becomes more robust.

Simulation results for some parameters pairs γ_{th} and Δ are shown in Figure B.3. It can be observed that for the threshold $\gamma_{th} = 6$ dB and the delay $\Delta = 10$ the best results are obtained. These are minimum values for γ_{th} and Δ, as further increasing them doesn't change the performance. In fact, the HPSP fails to outperform the IIR–PSP even for high SNR ($\bar{\gamma} > 20$ dB). Thus it can be concluded that it makes little sense to choose μ on a runtime basis, as the channel parameters ν'_{max} and $\bar{\gamma}$ are assumed to change slowly compared to a symbol interval.

B.2 Transition probabilities for the Gilbert–Elliott channel

Results of the transition probabilities of the Gilbert–Elliott channel (GEC), $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$, are shown in the Figures B.4 and B.5. In Figure B.4, the (a.) slow and (b.) fast fading performance is drawn against the SNR. Considering slow fading in Figure B.4.a, if the transition probability from 1 to 0 (from the burst to the good state), $P(\mathcal{T}_{10})$ is not significantly larger than the transition probability from 0 to 1, $P(\mathcal{T}_{01})$, the receiver becomes unstable. This can be observed for low SNR ($\bar{\gamma} \leq 10\,\text{dB}$) and long filters ($M = 8$), by comparing Figure B.4.a with Figure 6.15.a from section 6.3.3.2. On the other hand, if $P(\mathcal{T}_{10}) \approx 1$ the receiver is robust. This is seen for short filters ($M = 2$), by comparing the same graph with Figure 6.17. For fast fading in Figure B.4.b, $P(\mathcal{T}_{10})$ is close to 1 for arbitrary SNRs, thus the receiver is always robust, for short and long estimation filter orders. Note that for $M = 2$, $P(\mathcal{T}_{01})$ is above the ideal error probability where the CIR is known to the receiver. Even for longer filters, the transition probability of entering the burst state is in the range of the error probability. Similar conclusion can be drawn by observing Figure B.5, which shows $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$ against ν'_{max}. The receiver is robust if the difference between $P(\mathcal{T}_{10})$ and $P(\mathcal{T}_{01})$ is large, by comparing Figure B.5 with the error probability of the GEC in that case, shown in Figure 6.16.

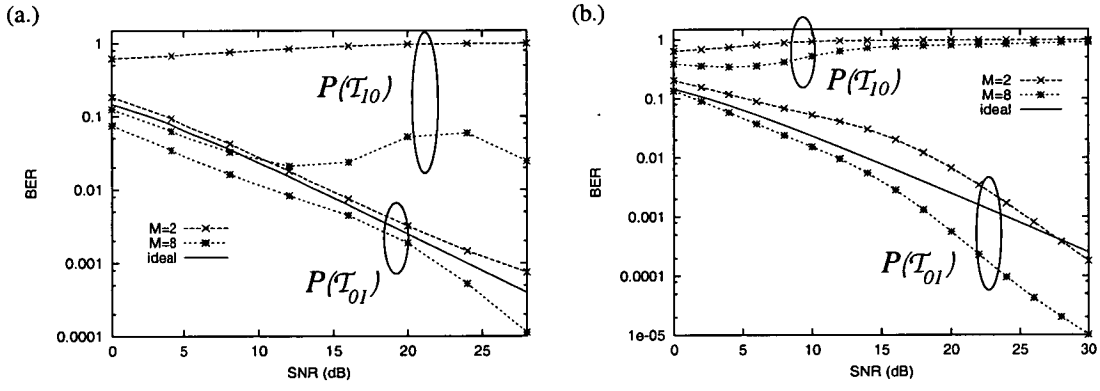


Figure B.4: *Transition probabilities for the GEC, $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$ vs SNR; $M = \{2, 8\}$.*
(a.) slow fading: $\nu'_{max} = 0.005$; *(b.) fast fading: $\nu'_{max} = 0.05$.*

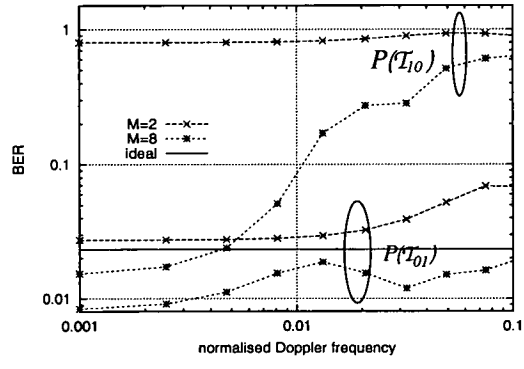


Figure B.5: *Transition probabilities for the GEC, $P(\mathcal{T}_{01})$ and $P(\mathcal{T}_{10})$ vs ν'_{max}; $M = \{2, 8\}$, $\bar{\gamma} = 10$ dB.*

Appendix C
Original Publications

The author of this thesis has the following publications:-

(i.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Adaptive Mobile Channel Prediction for Decision Directed RAKE Receivers," in *IEE Colloquium on Adaptive Signal Processing for Mobile Communication Systems, Digest 1997/383*, pp. 13/1–5, Oct. 1997.

(ii.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Robust Channel Predicton Technique for Decision Directed RAKE Receivers," *Electronics Letters*, vol. 34, pp. 338–40, Feb. 1998.

(iii.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Mobile Channel Estimation for Decision Directed RAKE Receivers Operating in Fast Fading Radio Channel," in *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'98), Sun City, South Africa*, vol. 2, pp. 576–579, Sep. 1998.

(iv.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Per-Survivor Processing Applied to Decision Directed Channel Estimation for a Coherent Diversity Receiver," in *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'98), Sun City, South Africa*, vol. 2, pp. 580–584, Sep. 1998.

(v.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Expanded Trellis MLSE Receiver for Flat Fading Channels," in *Proceedings of the IEEE International Symposium on Personal Indoor and Mobile Communications (PIMRC'99), Osaka, Japan*, Sep. 1999.

(vi.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "Hybrid Filtering for Per-Survivor Processing in Flat Fading Channels," in *Proceedings of IEEE Vehicular Technology Conference (VTC'99 fall), Amsterdam, The Netherlands*, Sep. 1999.

(vii.) G. Auer, G. J. R. Povey, and D. I. Laurenson, "On Unknown Parameter Estimation for MLSE on Flat Fading Channels," *Submitted to IEE Proceedings Communications*, 1999.

Publications (iii.), (v.) and (vi.) are included in the following pages.

ISSSTA 1998, Sun City, South Africa, September 2 – September 4

Mobile Channel Estimation for Decision Directed RAKE Receivers Operating in Fast Fading Radio Channels

Gunther Auer[1], Gordon J. R. Povey and David I. Laurenson

Department of Electrical Engineering,
The University of Edinburgh,
King's Buildings,
Mayfield Road,
Edinburgh EH9 3JL,
Scotland, UK.

Tel: +44 (0)131 650 5659
Fax: +44 (0)131 650 6554

Abstract — **The authors describe a robust channel prediction technique for a direct sequence spread spectrum (DS–SS) system in a fast fading environment. For improved performance the RAKE filter taps are coherently combined, hence accurate channel estimation is required. A FIR type linear prediction filter for each RAKE filter tap is used to estimate the channel response. In order to do this, the data decisions are fed back to the prediction filter. The stability of the proposed system is achieved through differential encoding of the data bits. It is demonstrated through simulations that the performance of the proposed decision directed receiver is better than that of an idealised receiver where channel estimation is not corrupted by decision feedback errors (e.g. by means of employing a pilot signal). The channel estimate can be significantly improved by employing a second stage channel estimation filter.**

I Introduction

Although differential encoding of the data bits does not essentially require any knowledge of the channel impulse response (CIR), the performance of the receiver may be significantly improved if an accurate estimate of the CIR is available. In a fast fading environment, a conventional differential receiver suffers from an irreducible bit error rate, due to the induced phase lag of two adjacent samples. One means of overcoming this problem is to use a predictive system where the CIR for a particular symbol is estimated in advance using the previously measured signals.

An optimal diversity receiver in a Rayleigh fading channel was proposed by Kam [1]. The resulting receiver consisted of an estimator that delivered minimum mean square error (MMSE) estimates of the CIR and a detector that utiliseed these estimates. However, the receiver complexity was found to be too high. A suboptimal realisation of a decision directed two stage receiver was also suggested.

The operation and performance of the receiver is compared with a system using a pilot signal for channel estimation.

[1]Email: Gunther.Auer@ee.ed.ac.uk

II System Model

The transmitter uses direct sequence spread spectrum (DS–SS) modulation. The information data bits $\Delta d(k)$ are DPSK modulated [2] to maintain stability. In an typical urban environment where no line of sight between the mobile and the base station exists, the received signal typically consists of a large number of incoming scatterers, termed a Rayleigh fading channel. For simulation work, the time variant statistics of the channel are described by the classical Doppler model [3, 4]. The bandwidth of a DS–SS modulated signal can be much larger than the coherence bandwidth of the channel, in this case the mobile radio channel is frequency selective. This provides a form of diversity and can subsequently be used to combat the effects of fading [5].

A Receiver structure

Due to the correlation properties of the spreading sequence, a RAKE type receiver can be applied to recover the received signal taps [5]. It is assumed that the correlation properties of the spreading sequence are ideal, in the way that the post correlation signal can be resolved perfectly into Q independent fading taps. After sampling and acquisition, the received signal at the qth tap of the RAKE receiver is in the form [2]

$$y_q(k) = d(k)\sqrt{E_b}\, h_q(k) + n(k) \tag{1}$$

where E_b represents the energy per transmitted data bit $d(k)$ and $h_q(k)$ denotes the CIR of the qth tap. Finally, $n(k)$ represents a sample of a complex additive white Gaussian noise (AWGN) process with zero mean and variance N_0.

In order to coherently combine the RAKE filter taps, the magnitude and phase of the channel response needs to be estimated and applied to the received signal, yielding for the RAKE receiver output

$$z(k) = \sum_{q=1}^{Q} y_q(k)\, \widehat{h}_q^*(k) \tag{2}$$

where $\widehat{h}_q^*(k)$ is the complex conjugate estimate of the CIR of the qth tap, which should ideally match $h_q^*(k)$. The estimate of the kth information bit $\Delta\widehat{d}(k)$ is obtained after passing $z(k)$ to the decision

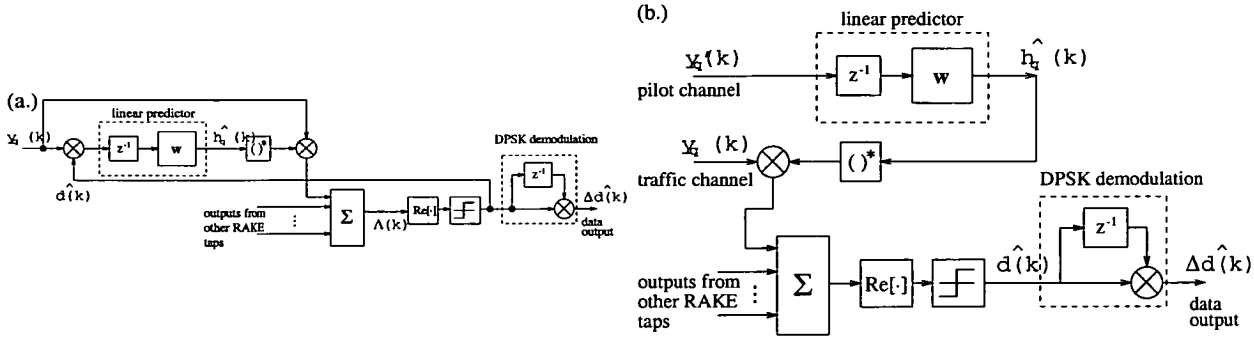


Fig. 1: Block diagram of the qth tap of a (a.) decision directed (b.) decision aided RAKE receiver.

circuit and performing DPSK demodulation

$$\widehat{d}(k) = \mathrm{sgn}[\,\mathrm{Re}\{z(k)\}\,]\;; \quad \Delta\widehat{d}(k) = \widehat{d}(k)\;\widehat{d}(k-1) \qquad (3)$$

where $\widehat{d}(k)$ denotes the estimate of $d(k)$. Two types of system models were investigated:

- A decision directed (DD) RAKE receiver, where the estimated data bit $\widehat{d}(k)$ is fed back to estimate the channel response.
- the second system uses a pilot signal, termed decision aided (DA). Such a DA communication system transmits a different spreading code on the same carrier frequency. Thus, the pilot signal experiences exactly the same fading as the information bearing signal $y_q(k)$ and is therefore used to form an estimate of $h_q(k)$. The SNRs of the pilot and traffic signals are assumed to be equal.

A block diagram of both system models are shown in Fig. 1.

B Linear prediction

The estimate of the CIR is obtained by applying the estimates of the M previous samples to an linear prediction filter with filter coefficients $\{w_m\}$, described by the equation

$$\widehat{h}_q(k) = \sum_{m=1}^{M} w_m\; \tilde{y}_q(k-m) \qquad (4)$$

where $\tilde{y}_q(k)$ denotes the received signal on the input of the RAKE receiver without data modulation, that is $\tilde{y}_q(k) = \sqrt{E_b}\, h_q(k) + n(k)$. Therefore, (4) applies to the DA–RAKE. For decision directed channel estimation, however, $y_q(k)$ is used. $\tilde{y}_q(k)$ may be obtained by multiplying the data estimate $\widehat{d}(k)$ to the received information bearing signal $y_q(k)$, yielding for the qth tap

$$\widehat{h}_q(k) = \sum_{m=1}^{M} w_m\; \widehat{d}(k-m)\, y_q(k-m) \qquad (5)$$

Consequently, in addition to AWGN, decision errors corrupt the channel estimation process. The estimate needs to be calculated for each tap separately. Introducing matrix notation for input samples $\tilde{\mathbf{y}}_q(k) = [\tilde{y}_q(k), \ldots, \tilde{y}_q(k-M+1)]^T$ and the filter weights $\mathbf{w} = [w_1, \ldots, w_M]^T$, both of dimension $\mathbb{C}^M$, (4) can be conveniently expressed as

$$\widehat{h}_q(k) = \mathbf{w}^T \tilde{\mathbf{y}}_q(k-1) \qquad (6)$$

where the superscript ^{T} denotes the matrix transpose operation. To optimise the filter design, we choose to minimise the mean square value of the estimation errors $E[|\tilde{y}_q(k) - \mathbf{w}^T \tilde{\mathbf{y}}_q(k-1)|^2]$, as a function of the weight vector $\mathbf{w}$, termed the MMSE criterion [6].

Decision errors caused by noise may increase the variance of the CIR estimate and thus induce further subsequent errors, giving rise to error propagation. The effects of a decision error will persist for the next M consecutive decision instants. The number of bits L between two consecutive error-free regions that are at least M bits in length are defined as an error burst, shown in Fig. 2. An error burst of approximately the same length as the number of predictor coefficients can cause the channel estimator to converge to false value. This means that the estimate of the impulse response is the inverse of the actual channel response. By examining (2) and (3) it is observed that for this case the received data sequence will be complementary to the transmitted one, i.e. $\{\widehat{d}(k)\} = -\{d(k)\}$. For DPSK such a cycle slip has no degrading effect, other than a single decision error, since the data decision depends on the difference to the previous sample only.

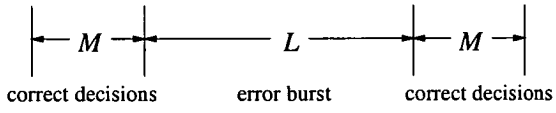


Fig. 2: Definition of an error burst.

C Second stage estimation filter

An improved approximation of the CIR of the DD–RAKE can be made by processing the received signal in two stages, due to Kam [1]. It is based on the fact, that the mean squared error (MSE) of the channel estimate becomes smaller if future samples in addition to the past samples are used. In the proposed two stage receiver, data decisions concerning the future symbols of the second stage are provided by the output of the first stage. In the first stage we employ the linear predictor, discussed in the previous section, to make tentative decisions on the symbols, using for each symbol only the past received signals. The estimated CIR for the second stage, using $M/2$ tentative decisions $\{\widehat{d}_1(k)\}$ of the first and the final decisions $\{\widehat{d}_2(k)\}$ of the second stage, respectively, is given by

$$\widehat{h}_{q,2}(k) = \sum_{\substack{m=-M/2 \\ m\neq 0}}^{M/2} u_m\; \widehat{d}_i(k-m)\, y_q(k-m)\;; \quad i = \begin{cases} 1, & m < 0 \\ 2, & m > 0 \end{cases} \qquad (7)$$

where $\{u_m\}$ are the filter weights, chosen according to the MMSE criterion [6]. The subscript $i = \{1, 2\}$ denotes the decision output of the first and second stage, respectively. Clearly, this adds a decision delay of $M/2$ symbols to the receiver. In matrix notation (7) can be expressed as

$$\hat{h}_{q,2}(k - M/2) = \mathbf{u}^T \tilde{\mathbf{y}}_q(k) \tag{8}$$

where $\mathbf{u} = [u_{-M/2}, \ldots, u_{-1}, 0, u_1, \ldots, u_{M/2}]^T$ denotes the vector containing the filter weights of the second stage estimation filter of dimension $\mathbb{C}^{M+1}$. A block diagram of the filter is shown in Fig. 3.

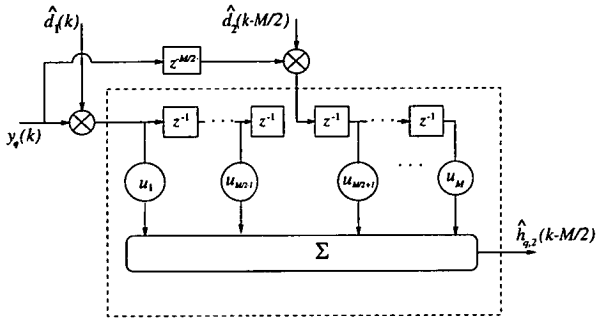


Fig. 3: Decision directed second stage channel estimation filter.

For the DA system no two stage processing is necessary. The received signal of the pilot channel can directly be applied to (8).

III Simulation Results

Simulation work is based on the specification in Tab. 1 for a complex baseband urban channel, defined by the COST-207 [4] report. To generate $h_q(k)$, the complex filter response of a 4th order noise shaping IIR filter is matched to the Doppler power spectra, as described in [7]. The performance of the system was evaluated for one and two taps, both statistics being described by the classical Doppler model. The signal–to–noise ratio (SNR) is defined by the sum of SNRs per tap, according to [2]. To allow fair comparison between a RAKE receiver with different taps, the average received signal power of all Q diversity taps was normalised to one. The SNR of the second tap was set to be 3 dB less relatively to the first one. The maximum Doppler shift was chosen to be 300 Hz, corresponding to a mobile moving with a maximum velocity of 180 km/h. Only the single user case is considered and perfect properties of the PN sequence are assumed.

Carrier frequency	f_c	1.8 GHz
Doppler frequency	ν_{max}	300 Hz
Symbol rate	f_s	8 kbit/s
$\rightarrow \nu_{max}/f_s$		0.0375
Modulation		DPSK DS–SS

Tab. 1: Simulation parameters.

To set up the M filter weights $\{w_m\}$ initially, the predictor for every tap was trained up with a block of 100 known data bits. Subsequently, the channel estimation is performed entirely decision directed, given the channel is stationary. On a wide-sense stationary channel, such as commonly assumed for mobile radio channels, the estimation process can be made adaptive [8].

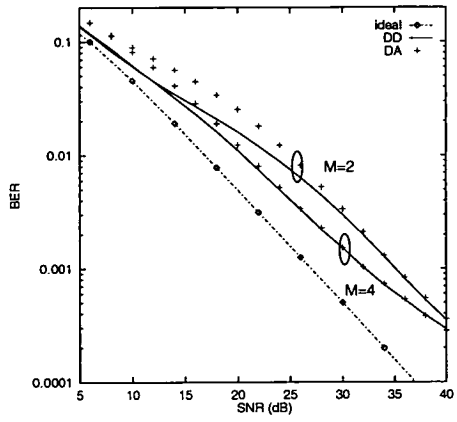


Fig. 4: BER performance of a decision directed (DD) and decision aided (DA) single tap receiver.

The system performance of the receiver using a linear predictor filter is considered in the Figs. 4–6. The bit error rate (BER) versus the SNR of a single tap RAKE receiver for $M = 2$ and 4 predictor coefficients, is shown in Fig. 4. The curve labelled "ideal" shows the results when the CIR is known perfectly to the receiver. It is seen that the DD predictor with the same number of coefficients is never worse than the DA predictor. Furthermore, the DD system performs significantly better for low SNR values. In particular, the fewer coefficients M that the predictor has, the more significant is the difference between the DD and DA system. The reason is that for DPSK modulation, an error in $\hat{d}(k)$ is likely to cause two consecutive errors in $\Delta\hat{d}(k)$, as seen in (3). However, in the decision directed system, the subsequent error caused by $\hat{d}(k-1)$ may be cancelled out by a cycle slip. This is because a cycle slip results in a phase shift for $z(k)$ of π in (2), which causes another error in $\hat{d}(k)$. In other words, the error induced by the DPSK modulation is outweighed by the cycle slip caused by error propagation in the DD channel predictor. For higher SNRs, a cycle slip becomes less likely due to the decreased occurrence of error bursts. Hence the DD and DA curves merge together. It is also seen that, unless the SNR is either very high or low, the $M = 4$ predictor performs considerably better than the one with 2 coefficients. While it is not shown in the graph, hardly any further improvement can be achieved for M larger than 4.

Fig. 5 illustrates the statistical analysis of the cycle slip effect. It shows the probability of an error burst of length L against the SNR for a $M = 2$ predictor. The probability that an observed error burst has the length L, is defined by

$$p(L) = \text{Prob}\{L_i = L\} \tag{9}$$

where $\{L_i\}$ is a set of randomly distributed error bursts obtained by Monte Carlo simulations. The calculation of $p(L)$ was repeated for a number of SNRs. By examining Fig. 5 it is seen that a single error burst is far more likely for the DD than for the DA system for a SNR somewhat smaller than 25 dB. Particularly for a SNR of 16 dB, the probability that an error burst is a single error reaches 80%, while the corresponding probability of the DA system never exceeds 10%. For high SNR values virtually all errors are 2 consecutive errors for both systems. The probability for an error burst larger than 2 is for both systems the same, being small for low and negligible for high SNRs.

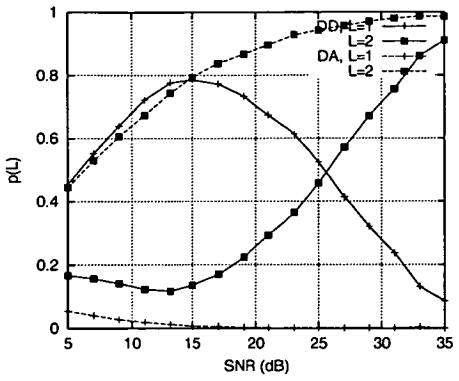


Fig. 5: Probability of an error burst length $L=1$, 2 respectively vs SNR for a $M=2$ predictor.

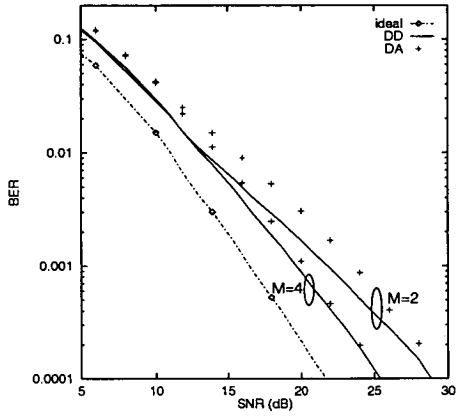


Fig. 6: BER performance of a decision directed (DD) and decision aided (DA) RAKE receiver with 2 taps.

The BER performance of the system can be improved by adding diversity to the system, as illustrated in Fig. 6. It shows the BER against the SNR of a DD and DA–RAKE receiver with 2 taps. Similar results are observed as in the former single tap case, in respect to the behaviour of the DD and DA–RAKE. The BER results however, are much better, allowing acceptable performance at far lower SNRs.

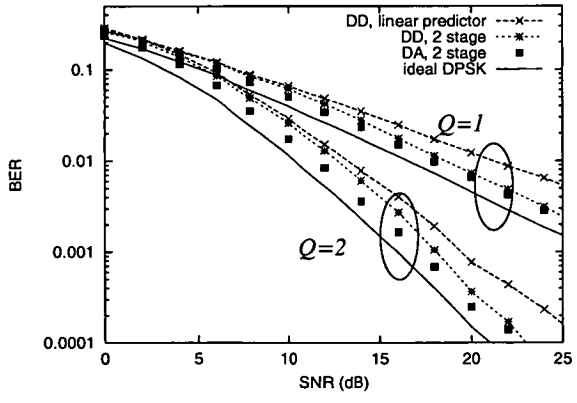


Fig. 7: BER performance of a DD and DA two stage RAKE receiver with one and two diversity taps. $M=4$.

The BER versus mean SNR curves for the two stage receiver are shown in Fig. 7. If the SNR is somewhat larger than 10 dB, processing $y_q(k)$ in two stages becomes worthwhile. For instance, at an error probability of 10^{-3} the difference towards the conventional linear predictor with the same number of coefficients exceeds 2 dB. Unlike the linear predictor, the DD–RAKE does not perform better than the DA–RAKE. As diversity is induced to the system, the difference between the practical receiver and the one which has *a priori* knowledge of the CIR increases. This is because the signal power per tap decreases and since channel estimation is performed for every tap separately, the SNR per tap decreases.

IV Conclusions

This paper has considered the performance of a DD compared to a DA–RAKE receiver, in a fast fading environment. The DD receiver was shown to have no stability problems and the DD system outperformed the system employing a pilot tone. This, however, has been done at the expense of DPSK modulation, which means approximately 3 dB poorer performance compared to BPSK [2]. Given that a DA–RAKE receiver has no stability problems when BPSK modulation is applied, DPSK is not the optimal modulation technique for a DA system. A significant improvement in system performance can be achieved by the two stage estimation filter, on the expense of an induced time delay and higher complexity.

References

[1] P.Y. Kam. "Optimal Detection of Digital Data over the Nonselective Rayleigh Fading Channel with Diversity Reception". *IEEE Trans. Commun.*, COM-39(2):pp214–219, Feb. 1991.

[2] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, 3rd edition, 1995.

[3] R.H. Clarke. "A Statistical Theory of Mobile-Radio Reception". *Bell System Technical Journal*, vol. 47(6):pp957–1000, Jul-Aug 1968.

[4] M. Failli, editor. *Digital Land Mobile Radio Communications - COST 207: Final Report*. Commision of the European Communities, Luxembourg, 1989.

[5] G. L. Turin. "Introduction to Spread Spectrum Anti-multipath Techniques and their Application to Urban Digital Radio". *Proceedings of the IEEE*, vol. 68(3):328–353, Mar. 1980.

[6] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ, 2nd edition, 1991.

[7] D. I. Laurenson and G. J. R. Povey. "Channel Modelling for a Predictive RAKE Receiver System". *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), The Hague, The Netherlands*, pages pp715–719, Sep 1994.

[8] G. Auer, G.J.R. Povey, and D.I. Laurenson. "Adaptive Mobile Channel Prediction for Decision Directed RAKE Receivers. In *IEE Colloquium on Adaptive Signal Processing for Mobile Communication Systems, Digest 1997/383*, pages pp13/1–5, Oct. 1997.

Expanded Trellis MLSE Receiver for Flat Fading Channels

Gunther Auer, Gordon J. R. Povey and David I. Laurenson

Department of Electronics and Electrical Engineering,
The University of Edinburgh,
King's Buildings,
Mayfield Road,
Edinburgh EH9 3JL,
Scotland, UK.

Email: ga@ee.ed.ac.uk
Tel: +44 (0)131 650 5659
Fax: +44 (0)131 650 6554

Abstract—**Receiver designs for maximum likelihood sequence estimation (MLSE) of signals transmitted through a flat fading channel are described. This results in a receiver which jointly estimates the channel impulse response and detects the received data sequence. The receiver uses data as well as pilot symbols, embedded in the Viterbi algorithm (VA). The VA operates on an expanded trellis structure, such that the states depend not only on the current symbol to be detected, but also on previous samples, due to the estimation filter memory. If the estimate of the channel output has a finite memory, then the receiver optimises the maximum likelihood (ML) decision rule. The principle of per-survivor processing (PSP) is shown to be a special case of the proposed receiver structure. The results are not limited to slow or fast fading channels. The PSP detector, having a comparable low complexity, does not suffer a significant performance degradation in fast fading compared to the expanded trellis MLSE; however, PSP performs very poorly in slow fading conditions.**

I. Introduction

In this paper a realisation of MLSE for Gaussian signals in Gaussian noise studied by Kailath [1], is presented. Several approaches of performing MLSE in a time-varying mobile communication link have been studied in literature, based on utilising the Viterbi algorithm (VA) [2]. Morley and Snyder [3] combined the generalised likelihood-ratio formula with the VA. The resulting receiver structure follows Kailath's separation theorem [1]. That is, the receiver consists of an estimator which delivers the minimum mean squared error (MMSE) estimates of the fading distortion and a detector that utilises these estimates. More recently, the idea of *per-survivor processing* (PSP) [4] was used for joint channel estimation and signal detection in a flat fading channel [5].

In order to perform coherent detection, a form of phase reference for the receiver must be provided. The idea of reference symbol phase tracking was studied by [6]. These receivers only used the pilot symbols multiplexed in the data stream for channel estimation. The performance can be improved, especially for fast fading, if pilot symbols as well as data symbols are used for channel estimation, developed by Irvine and McLane [7]. This essentially requires a more sophisticated receiver design and joint estimation and detection of the entire transmitted sequence becomes attractive. In this paper, the MLSE receiver embeds the pilot symbol aided signal in the VA itself.

The transmission and channel model of the mobile radio link is described in section II. Section III considers optimum maximum likelihood detection with unknown parameter estimation. The realisation of an MLSE receiver based on an expanded trellis is discussed in section IV. Simulation results are presented in section V.

II. System Model

We consider binary antipodal transmission, namely binary phase shift keying (BPSK). In the pilot symbol insertion (PSI) technique [6], one of R data bits is known at the receiver and is used for carrier recovery. Hence one known symbol is followed by $R-1$ data symbols, corresponding to the ratio $1:R$.

The BPSK modulated signal, multiplexed with the pilot symbols, is then transmitted over an urban radio channel. As the mobile is in motion, the channel response changes with time due to multipath and the Doppler effect. The channel impulse response (CIR) of time instant k is modelled as a complex time discrete random variable, h_k, being a single realisations of a *wide-sense stationary stochastic process* with zero mean. For simulation work the classical Doppler power spectra was chosen, due to Clarke [8], having the auto correlation function (ACF)

$$E[h_k\, h_{k+\Delta k}^*] = E[|h_k|^2]\; J_0(2\pi\, \nu_{max}\, \Delta k) \qquad (1)$$

where $J_0(\cdot)$ denotes the zero order Bessel function of the first kind, ν_{max} is the normalised maximum Doppler frequency. The superscript * denotes the complex conjugate operation.

After sampling and matched filtering, the received signal is in the form

$$y_k = d_k\sqrt{E_b}\,h_k + n_k \ . \tag{2}$$

The energy per transmitted data bit, d_k, is denoted by E_b, and n_k represents a sample of a complex additive white Gaussian noise (AWGN) process with zero mean and variance N_0.

III. Optimum ML Sequence Detector

In this section optimum MLSE with unknown parameter estimation is considered. The detection and estimation of a whole sequence of the received signal is discussed in this section. The motivation behind this is, adjacent samples of $\{h_k\}$, and hence $\{y_k\}$, are highly correlated. Thus, to minimise the probability of error, the whole transmitted sequence has to be taken into account. Sequences of length K are denoted by column vectors, e.g. for the received signal sequence $\mathbf{y} = [y_1, \cdots, y_K]^T$. We wish to find the sequence $\widehat{\mathbf{d}}$ which minimises the probability of error, out of all possible transmitted sequences $\{\mathbf{d}^{(\ell)};\ \ell \in \mathcal{A}_K\}$. The set $\mathcal{A}_K$ represent all 2^K possible data sequences. All sequences are assumed to occur with the same *a priori* probability. According to the maximum likelihood (ML) decision rule, the *likelihood function* of the observation $\mathbf{y}$, conditioned on the ℓ^{th} transmitted data hypothesis, $\mathbf{d}^{(\ell)}$, is to be maximised:

$$p(\mathbf{y}|\,\widehat{\mathbf{d}}) = \max_{\ell\in\mathcal{A}_K} p(\mathbf{y}|\,\mathbf{d}^{(\ell)}) \tag{3}$$

such that $\widehat{\mathbf{d}}$ is the most likely transmitted sequence.

We consider a *recursive* formulation of MLSE employing present and past samples only. The decision variable of data hypothesis ℓ and time instant k, being essentially the *log-likelihood function* of (3), is given by [9]

$$\Lambda_k^{(\ell)} \ = \ \left|y_k'^{(\ell)} - \widehat{h}_k^{(\ell)}\right|^2 + \Lambda_{k-1}^{(\ell)} \tag{4}$$

where $\widehat{h}_k^{(\ell)}$ is the estimated CIR and $y_k'^{(\ell)} = d_k^{(\ell)*} y_k$ denotes the pre-multiplied received signal. Minimising $\Lambda_k^{(\ell)}$ over all possible hypothesis, ℓ, at the end of the sequence $k = K$ gives the ML estimate

$$\widehat{\Lambda}_K \ = \ \min_{\ell\in\mathcal{A}_K} \Lambda_K^{(\ell)} \ . \tag{5}$$

It can be observed that the ML sequence detector minimises the Euclidean distance between the received signal, $\mathbf{y}$, and the channel estimate, $\widehat{\mathbf{h}}^{(\ell)}$, in respect to all possible transmitted sequences $\{\mathbf{d}^{(\ell)}\}$. This operation minimises the probability of error for detection of the whole data sequence, equivalent to the ML decision rule. The number of sequences, which have to be tested in order to find $\widehat{\Lambda}_K$, grow exponentially with the sequence length K. Thus, this receiver cannot be implemented in this form, due to its prohibitive complexity.

A. Channel estimation

The CIR estimate conditioned on the ℓ^{th} hypotheses, $\widehat{h}_k^{(\ell)}$, is obtained by minimising the mean squared of the prediction error $E\left[|h_k - \widehat{h}_k^{(\ell)}|^2\right]$. The optimum solution is a Wiener filter [10], which provides the MMSE estimate of the CIR. In a stationary channel the channel estimation filter can be truncated by a time independent, M^{th} order, linear predictor, with the coefficients $\{w_m\}$. The channel estimate $\widehat{h}_k^{(\ell)}$ then becomes

$$\widehat{h}_k^{(\ell)} = \sum_{m=1}^{M} w_m^*\, y_{k-m}'^{(\ell)} \ . \tag{6}$$

The formulation of the decision variable in (4), cannot be employed in this form for coherent detection, such as BPSK, since the transmitted signal is not orthogonal. The resulting phase ambiguity at the receiver necessitates the need of a phase reference, which can be provided by pilot aided channel estimation. The multiplexed pilot symbols with rate R can be incorporated in (6) and (4) by redefining the pre-multiplied received signal

$$y_k'^{(\ell)} = \begin{cases} \sqrt{E_b}h_k + n_k \ ; & k \bmod R = 0 \\ d_k^{(\ell)*} y_k \ ; & \text{elsewhere.} \end{cases} \tag{7}$$

IV. Realisation of the MLSE Receiver

It has been described how an exhaustive search over all possible data sequences yields the desired joint ML estimate. In the following the problem of ML sequence estimation using the Viterbi algorithm is analysed based on the decision variable in (4).

The VA represents an optimum solution of ML sequence detection of a finite state, discrete time Markov process observed in memoryless noise. The VA is only optimal as a decoding algorithm if the process is Markovian.[1] That is the probability, $P(s_{k+1}|s_1, \cdots, s_k)$, of being in state s_{k+1} depend only on the state s_k [2]:

$$P\left(s_{k+1}\,|\,s_1, \cdots, s_k\right) = P\left(s_{k+1}\,|\,s_k\right) \ .$$

For PSP (e.g. discussed in [5]), the state s_k is given by the current data symbol of hypothesis ℓ, which is $s_k = \{0, 1\}$. The estimated CIR however, is dependent of the last $M+1$ samples, as observed in (6). Hence, the transition from state s_k to s_{k+1} is a $(M+1)^{\text{th}}$ order Markov process.

This is illustrated in Fig. 1. The sequences $\ell = \ell_\nu$ and $\ell = \ell_\mu$ denote two particular realisations of possible transmitted data sequences. The Euclidean metric from the transition (s_k, s_{k+1}) of hypothesis ℓ_ν and ℓ_μ may be different. Thus if the VA was applied on s_k the corresponding survivor may not be the most likely path in the end of the sequence. Note, there are 2^M sequences per state which have a different path history.

[1] A 1st order Markov process is described as Markovian.

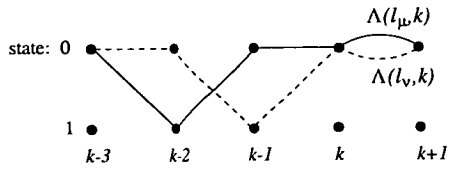


Fig. 1. Distance metric illustrated in a trellis for the recursive MLSE.

A. *Receiver design based on an expanded trellis structure*

Morley and Snyder [3] derived a more general description of the problem, which will be used in the following. Generally speaking, if the trellis is expanded to 2^{M+1} states per sample, then the underlying process is Markovian, and the VA is therefore optimal. This approach leads to a very similar solution as for signals transmitted through channels with inter symbol interference (ISI) [11]. The proposed algorithm is only optimum for channels with a finite channel impulse response. Unfortunately this is not the case on a Rayleigh fading channel. However, based on the approximation of the channel estimation filter to a M^{th} order FIR filter, the algorithm is the best possible decision rule. The only non-optimum approximation is to truncate the estimation of the CIR with a M^{th} order linear predictor.

In the following, an expanded trellis structure is defined, where not only the current data symbol is taken into account, but also the previous symbols up to delay D_{del}. Let the state of the k^{th} sample of a finite state Markov process be denoted by

$$s_k \triangleq \{ d_k^{(\ell)}, \underbrace{d_{k-1}^{(\ell)}, \cdots, d_{k-D}^{(\ell)}}_{\text{used for channel estimation}} \}\,; \quad \ell \in s_k \qquad (8)$$

where $d_k^{(\ell)}$ is the current data symbol, the delayed versions $\{d_{k-m}^{(\ell)},\ m = 1, \cdots, D\}$ are used for channel estimation. There are 2^{D+1} states $\{s_k\}$ per time instant, according to the memory of $D+1$ time samples of one state. The transition (s_k, s_{k+1}) is obtained in accordance to (4), which is the Euclidean distance between the estimated and received signal output of one sample, given by

$$\Delta(s_k, s_{k+1}) \;=\; \left| y'^{(\ell)}_{k+1} - \widehat{h}^{(\ell)}_{k+1} \right|^2 ; \quad \ell \in s_{k+1}\,. \qquad (9)$$

In the equilibrium there are two entering and two leaving branches per state. Initialisation of the trellis takes $D+1$ samples. There is obviously a one-to-one correspondence between state sequences $\{s_k\}$ and transition sequences $\Delta(s_k, s_{k+1})$, given by a *finite state machine* or a *shift register process* [2], since it can be modelled by a shift register of length D. This results in an expanded trellis structure, illustrated in Fig. 2. By observing the trellis, the analogy to equalisation of channels with ISI [11] becomes obvious.

It is seen that there is nothing to be gained, if D is chosen larger than the estimation filter order M. Hence, $D \leq M$, with equality for the best possible decision rule.

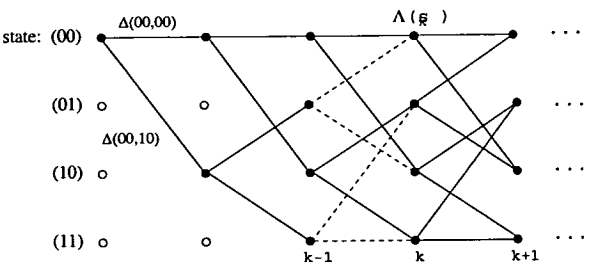


Fig. 2. Expanded distance metric illustrated in a trellis for the recursive MLSE, $D=1$.

Note that if $D = 0$, the algorithm becomes identical to PSP [5].

Now that we have stated the problem, it is straightforward to implement the VA. Let $\widehat{\Lambda}(s_k)$ denote the survivor path of state s_k. That is the metric with the minimum distance entering this state, which is obtained by minimising(4) at sample k

$$\widehat{\Lambda}(s_k) = \min_{\ell \in s_k} \{\Lambda_k^{(\ell)}\}\,. \qquad (10)$$

To extend the survivors to sample $k+1$, we compute the metric from state s_k to s_{k+1}:

$$\Lambda_{k+1}^{(\ell)} \;=\; \widehat{\Lambda}(s_k) + \Delta(s_k, s_{k+1}) \qquad (11)$$

and again applying (10) to $\Lambda_{k+1}^{(\ell)}$. The decision variable up to sample k is already minimised, leaving only 2 transitions $\Delta(s_k, s_{k+1})$ to be minimised. From the branches entering at state s_k, the ones with the larger metric are discarded, giving the survivor at time $k+1$:

$$\widehat{\Lambda}(s_{k+1}) \;=\; \widehat{\Lambda}(s_k) + \min_{(s_k, s_{k+1})} \Delta(s_k, s_{k+1})\,. \qquad (12)$$

This is illustrated in Fig. 2, where the VA calls for choosing between two transitions, printed as dashed and solid lines.

Although the ML decision rule states that the final decision is taken at the end of the sequence, little degradation is expected if the definite decision of the most likely path is done after a finite delay. For simulation work presented here, this time delay is set to R.

The expanded trellis MLSE is closely related to list type Viterbi decoders studied in literature [12], [13]. The list Viterbi algorithm (LVA) constitutes a generalised class of the VA. Its basic concept is to find the L best paths instead of a single survivor.

Now the necessary phase reference in form of the multiplexed pilot symbols needs to be incorporated. The multiplexing rate is R; one known symbol is followed by $R-1$ data symbols. If a pilot is being transmitted, $d_k^{(\ell)} = 1$, and therefore states s_k which imply that $d_k^{(\ell)} = -1$ are *not* allowed. The following D samples the pilot is shifted through the state machine, leaving half of the possible states in the trellis for $D+1$ samples. In general, a state

is not allowed because of a pilot symbol at time instant $k-m$, if the following relation is true:

$$d^{(i)}_{k-m} = -1\,; \qquad (k-m) \mod R = 0\,.$$

B. Computational cost of the expanded trellis

Given the trellis is in its equilibrium, we have to find one survivor out of all the branches entering a particular state (for BPSK it is 2). Note, for every state 2 new transitions need to be calculated. It is seen that MLSE has become independent on the length of the sequence, after initialisation of $D+1$ samples.

The complexity for $D = M$ may still be prohibitively large, due to the exponential growth of the trellis relative to D. The trellis has 2^{D+1} number of states, with 2 branches entering and leaving each state per sample, resulting in 2^{D+2} transitions in total for the whole trellis.

As implied before, the complexity can be further reduced by simply reducing D and therefore the number of states. This can be achieved by simply setting $D < M$. The principle of per-survivor processing (PSP) [4] turns out to be a special case of the proposed algorithm, by setting $D = 0$. The trellis is reduced to two states per sample.

According to the similarity of the trellis with ISI corrupted channels, other state reduction techniques (e.g. the LVA[2] [12]), addressed in literature for channels with ISI, can be applied in a straightforward way. These techniques are based on the fact that only a small number of the 2^{D+2} survivors have a sufficiently small metric to become the overall best survivor path, and therefore the most likely sequence. The vast majority of the survivors can be discarded before the final decision is made, without compromising the system performance.

V. Results

Simulation work is based on a complex baseband flat fading channel, commonly described by a Rayleigh fading channel. To generate h_k, the complex filter response of a 4th order noise shaping IIR filter is matched to the Doppler power spectra of (1), as described in [14].

The bit error rate (BER) was obtained by simulating the MLSE receiver designs over a large number (10^6) of Monte Carlo runs. The error probability, for the idealised case if the CIR is known *a priori* [15, chapter 14], is used as a lower bound for comparison purposes. That is for a frequency-flat fading channel:

$$P_e = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}}{1+\bar{\gamma}}}\right] \qquad (13)$$

where $\bar{\gamma} = E_b/N_0$ denotes the average signal-to-noise ratio (SNR). Generally, curves labelled "ideal" identify the

[2]In [12] this algorithm is referred to as generalised VA (GVA), however the basic concept is the same as for the LVA in other publications.

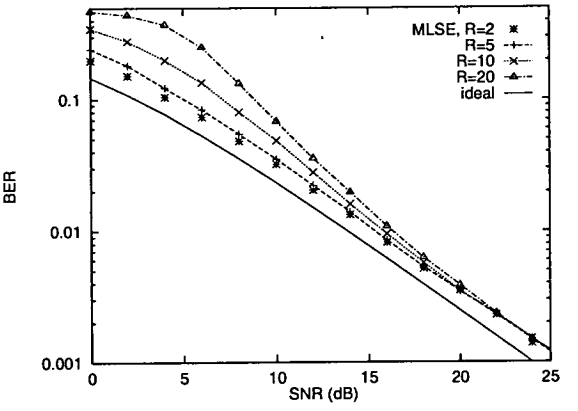


Fig. 3. BER vs SNR for different numbers of the pilot insertion rate R. $M = D = 4$, $\nu_{max} = 0.05$.

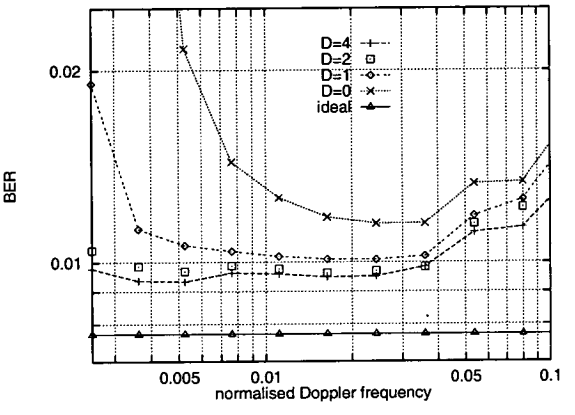


Fig. 4. BER vs ν_{max} for different numbers of D. $M = 8$, $R = 10$, $\bar{\gamma} = 15$ dB.

case where the CIR is known to the receiver. The Doppler frequency ν_{max} is normalised to the sampling frequency.

The BER performance against the SNR, of a system with a trellis memory length of $D = 4$ in a fast fading channel ($\nu_{max} = 0.05$), is shown in Fig. 3. It is seen that varying the multiplexing rate R mainly affects the performance for low SNR, such that for small R the difference to the curve with known CIR becomes smaller. This is because, a smaller R decouples the joint estimation-detection structure of the receiver and therefore reduces error propagation. For a larger R however, the channel estimation process is severely corrupted during deep fades, leading to a poor CIR estimate, so it becomes more likely that subsequent bits will be detected as errors as well.

The system performance dependent on D, is shown in Fig. 4,[3] where the BER is plotted against the normalised Doppler frequency, ν_{max}. The mean SNR was set to $\bar{\gamma} = 15$ dB. It can be observed, that the PSP receiver

[3]Due to the exponential growth of the trellis in respect to D, values larger than 4 result in a vast computational complexity, while the performance is only marginally improved. Thus graphs with $D > 4$ have been omitted.

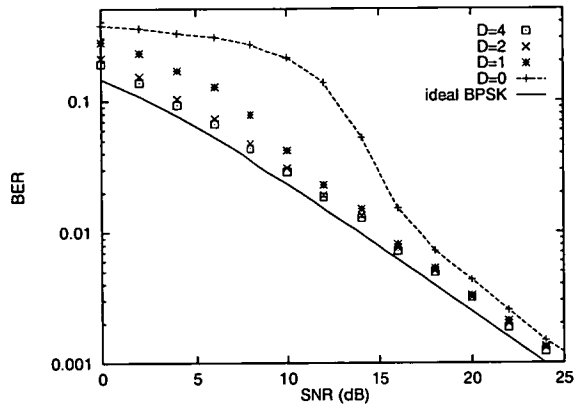


Fig. 5. BER vs SNR for different numbers of D on a slow fading channel. $M=8$, $R=10$, $\nu_{max}=0.005$.

($D = 0$) is not robust for low Doppler. The performance gradually breaks down for ν_{max} somewhat smaller than 0.02. Expanding the trellis ($Dgeq1$) does improve the system performance for low Doppler. However, if $D>2$, virtually no performance gain can be achieved. This is true, even for higher order estimation filters. Hence, almost no performance degradation is observed if $D=2$.

For fast fading ($\nu_{max} > 0.02$), however, PSP performs almost as well as MLSE with $D \geq 1$. The reason is, the contribution of filter taps in (6), whose coefficients w_m are small, is rather insignificant. Especially for fast fading channels, where the correlation between adjacent samples becomes smaller, there are only a few coefficients w_m which have a significantly large magnitude. So expanding the trellis has little benefit as signal samples with larger time delay have less impact on the channel estimation, due to their smaller weight w_m.

Note, the effect of varying D is also dependent on the SNR, as it is illustrated in the following graph. The PSP performance for the slow fading case, as shown in Fig. 5, is seen to have stability problems for low SNRs, while for SNR values above a certain trade-off, the PSP graph gets much closer to MLSE with larger D. It can therefore be concluded that the PSP detector is suitable only for for fast fading channels. However, the PSP performance for low SNR and slow fading improves, if M becomes smaller [16]. Provided that $D \geq 2$, the error probability closely matches the lower bound of the known CIR. It is seen that there is a trade-off where the PSP performance breaks down dependent on ν_{max} and the SNR, $\bar{\gamma}$.

VI. Conclusions

A MLSE receiver, based on an expanded trellis structure, operating in a Rayleigh fading channel with unknown impulse response, has been implemented. The performance of the MLSE receiver is highly dependent on the multiplexing rate R, if the SNR is low. The system performance cannot be improved significantly by choosing $D>2$. This is a pleasing result for systems where the estimator order M is large, due to the exponential growth of the trellis.

The principle of per-survivor processing (PSP) was shown to be a special case of the proposed receiver. Its performance is excellent for fast fading channels, while it is not robust for slow fading. On the other hand the receiver with $D \geq 2$ works for slow as well as fast fading; while having a considerable higher complexity than PSP. However, the complexity may be further reduced through reducing the number of survivors of the trellis, e.g. by means of the LVA [12]. The expanded trellis MLSE receiver may be generalised for channels with ISI and unknown CIR, as the trellis structure is essentially the same.

References

[1] T. Kailath, "Correlation Detection for Signal Perturbed by a Random Channel," *IRE Trans. Inf. Theory*, vol. IT-6, pp. 361–66, June 1960.

[2] G. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–78, March 1973.

[3] R. Morley and D. Snyder, "Maximum Likelihood Sequence Estimation of Randomly Dispersive Channels," *IEEE Trans. Commun.*, vol. COM-27, pp. 833–839, June 1979.

[4] R. Raheli, A. Polydoros, and C. Tzou, "Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments," *IEEE Trans. Commun.*, vol. COM-43, pp. 354–364, Feb.-Apr. 1995.

[5] G. Vitetta and D. Taylor, "Maximum Likelihood Decoding of Uncoded and Coded PSK Signal Sequences Transmitted over Rayleigh Flat-Fading Channels," *IEEE Trans. Commun.*, vol. COM-43, pp. 2750–58, Nov. 1995.

[6] J. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Trans. Veh. Tech.*, vol. VT-40, pp. 686–693, Nov. 1991.

[7] G. Irvine and P. McLane, "Symbol Aided Plus Decision-Directed Reception for PSK/TCM Modulation on Shadowed Mobile Satellite Fading Channels," *IEEE Journal Sel. Areas Commun.*, vol. SAC-10, pp. 1289–98, Oct. 1992.

[8] R. Clarke, "A Statistical Theory of Mobile-Radio Reception," *Bell System Technical Journal*, vol. 47, pp. 957–1000, Jul-Aug 1968.

[9] M. Moher and J. Lodge, "Maximum Likelihood Sequence Estimation of CMP Signals Transmitted over Rayleigh Flat Fading Channels," *IEEE Trans. Commun.*, vol. COM-38, pp. 787–94, June 1990.

[10] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 2nd ed., 1991.

[11] G. Forney, "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 363–378, May 1972.

[12] T. Hashimoto, "A List-Type Reduced-Constraint Generalisation of the Viterbi Algorithm," *IEEE Trans. Inf. Theory*, vol. IT-33, pp. 866–76, Nov. 1987.

[13] N. Seshadri and C. Sundberg, "List Viterbi Decoding Algorithms with Applications," *IEEE Trans. Commun.*, vol. COM-42, pp. 313–323, Feb.-Apr. 1994.

[14] D. I. Laurenson and G. J. R. Povey, "Channel Modelling for a Predictive RAKE Receiver System," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), The Hague, The Netherlands*, pp. 715–719, Sep 1994.

[15] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 3rd ed., 1995.

[16] G. Auer, G. Povey, and D. Laurenson, "Per-Survivor Processing Applied to Decision Directed Channel Estimation for a Coherent Diversity Receiver," in *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA), Sun City, South Africa*, vol. 2, pp. 580–584, Sep. 1998.

Hybrid Filtering for Per-Survivor Processing in Flat Fading Channels

Gunther Auer, Gordon J. R. Povey and David I. Laurenson

Department of Electronics and Electrical Engineering,
The University of Edinburgh,
King's Buildings,
Mayfield Road,
Edinburgh EH9 3JL,
Scotland, UK.

Email: ga@ee.ed.ac.uk
Tel: +44 (0)131 650 5659
Fax: +44 (0)131 650 6554

Abstract—**Receiver designs for maximum likelihood sequence estimation (MLSE) of BPSK modulated signals transmitted through a flat fading channel are described. This results in a receiver which jointly estimates the channel impulse response and detects the received data sequence. The receiver uses data as well as pilot symbols, embedded in the Viterbi algorithm (VA), referred to as per-survivor processing (PSP). It is demonstrated through simulations that the performance of PSP is very sensitive to the choice of the channel estimation filter. PSP employing a FIR filter to estimate the channel impulse response (CIR) is well examined for fast fading conditions. For modest fading rates however, stability problems due to pilot aided channel estimation occur. When channel estimation is performed using an IIR filter, the receiver shows superior performance for slow fading, while it is not suitable for fast fading. In this paper a hybrid realisation of PSP is proposed, using both a FIR and IIR filter, yielding a receiver suitable for slow and fast fading channels.**

I. Introduction

In this paper a realisation of a maximum likelihood (ML) sequence detector for Gaussian signals in Gaussian noise studied by Kailath [1], is presented. That is, the receiver consists of an estimator which delivers the minimum mean squared error (MMSE) estimates of the fading distortion and a detector that utilises these estimates. Several approaches of performing MLSE in a time-varying mobile communication link have been studied in literature, based on utilising the Viterbi algorithm (VA) [2]. This technique is sometimes referred to as the principle of *per-survivor processing* (PSP) [3].

PSP can be applied in a straightforward manner if orthogonal waveforms, or non-coherent modulation such as differential encoding is employed. Differential encoding of the data bits, however, suffers a loss of ≈ 2 dB in system performance. For coherent detection, on the other hand, a form of phase reference for the receiver must be provided. The idea of reference symbol phase tracking was introduced by [4], [5]. These receivers only used the pilot symbols multiplexed in the data stream for channel estimation. The performance can be improved, especially for fast fading, if pilot symbols as well as data symbols are used for channel estimation. This essentially requires a more sophisticated receiver design and joint estimation and detection of the entire transmitted sequence becomes attractive.

PSP was used for joint channel estimation and signal detection for MPSK modulated signals [6], [7], [8]. Kam *et al.* [6] used a first order IIR filter to estimate the channel response, while [7], [8] used a FIR estimation filter. Simulation results suggest that the FIR estimation filter performs superior for fast; while the IIR filter yields better results for the slow fading channel. For slow fading, PSP using a FIR estimation filter experiences stability problems, due to pilot aided channel estimation. This is a rather unexpected result, as normally fast fading is considered to be the more difficult condition. In this paper the reasons for these stability problems are thoroughly analysed.

A hybrid PSP detector is presented which employs both an IIR and FIR channel estimation filter. This enables the PSP based receiver to perform close to the theoretical minimum for coherent detection, on both slow and fast fading channels.

II. System Model

We consider binary antipodal transmission, namely binary phase shift keying (BPSK). In the pilot symbol insertion (PSI) technique [4], [5], one of R data bits is known at the receiver and is used for carrier recovery. Hence one known symbol is followed by $R-1$ data symbols, corresponding to the ratio $1:R$.

The BPSK modulated signal, multiplexed with the pilot symbols, is then transmitted over an urban radio channel. As the mobile is in motion, the channel response changes with time due to multipath and the Doppler effect. The channel impulse response (CIR) of time instant k is modelled as a complex time discrete random variable, h_k, being a single realisations of a *wide-sense stationary*

stochastic process with zero mean. For simulation work the classical Doppler power spectra was chosen, due to Clarke [9], having the auto correlation function (ACF)

$$E\left[h_k\, h^*_{k+\Delta k}\right] = E\left[|h_k|^2\right]\, J_0(2\pi\, \nu_{max}\, \Delta k) \qquad (1)$$

where $J_0(\cdot)$ denotes the zero order Bessel function of the first kind, ν_{max} is the normalised maximum Doppler frequency. The superscript * denotes the complex conjugate operation.

After sampling and matched filtering, the received signal is in the form

$$y_k = d_k\, h_k\, \sqrt{E_b} + n_k \qquad (2)$$

where E_b denotes the energy per transmitted bit, n_k represents a sample of a complex additive white Gaussian noise (AWGN) process with zero mean and variance N_0.

III. Receiver Structure

A. Optimum ML Sequence Detector

The detection and estimation of a whole sequence of the received signal is discussed in this section. We wish to find the sequence $\{\hat{d}_k\}$ which minimises the probability of error, out of all possible transmitted sequences $\{d_k^{(\ell)};\, \ell \in \mathcal{A}\}$. The set $\mathcal{A}$ constitutes all possible transmitted sequences of number 2^K, with K being the sequence length. The receiver is based on the estimator–correlator structure studied by Kailath [1]. We consider a *recursive* formulation of a ML sequence detector employing present and past samples only studied by [10]. The decision variable of data hypothesis ℓ and time instant k, being essentially the *log-likelihood function* of the ML decision rule, is given by

$$\Lambda_k^{(\ell)} \;=\; \left|y_k'^{(\ell)} - \widehat{h}_k^{(\ell)}\right|^2 + \Lambda_{k-1}^{(\ell)} \qquad (3)$$

where $y_k'^{(\ell)} = d_k^{(\ell)*} y_k$ denotes the pre-multiplied received signal and $\widehat{h}_k^{(\ell)}$ is the estimated CIR. Minimising $\Lambda_k^{(\ell)}$ over all possible hypothesis in $\mathcal{A}$, at the end of the sequence $k = K$ gives the ML estimate

$$\widehat{\Lambda}_K \;=\; \min_{\ell \in \mathcal{A}} \Lambda_K^{(\ell)}\,. \qquad (4)$$

It can be observed that the ML sequence detector minimises the Euclidean distance between the received signal, $\{y_k\}$, and the channel estimate, $\{\widehat{h}_k^{(\ell)}\}$, in respect to all possible transmitted sequences $\{d_k^{(\ell)}\}$. This operation minimises the probability of error for detection of the whole data sequence, equivalent to the ML decision rule.

B. Channel estimation

The estimate of the CIR conditioned on the ℓ^{th} hypotheses, $\widehat{h}_k^{(\ell)}$, is obtained by minimising the mean squared of the prediction error $E\left[|h_k - \widehat{h}_k^{(\ell)}|^2\right]$. The optimum solution is a Wiener filter [11], which provides the MMSE estimate of the CIR.

FIR estimation filter: The Wiener filter with k time dependent coefficients can be truncated by a linear prediction filter with M coefficients, $\mathbf{w} = [w_1, \cdots, w_M]^T$, giving the channel estimate [10]

$$\widehat{h}_k^{(\ell)} = \sum_{m=1}^{M} w_m^*\, y_{k-m}'^{(\ell)}\,. \qquad (5)$$

In a stationary channel, $\mathbf{w}$ is time independent and can be pre-computed.

IIR filtering: Another possibility of performing an one-step prediction is to use an *infinite impulse response* (IIR) filter instead of a FIR filter. The optimal IIR-type predictor is given by the *Kalman filter* [11]. In order to keep the complexity to a minimum let us constrain the Kalman filter to be a first order stationary IIR filter with a real valued, scalar parameter α.[1]

The recursive channel estimator can be expressed as [6]

$$\widehat{h}_k^{(\ell)} = (1-\alpha)\, y_{k-1}'^{(\ell)} + \alpha\, \widehat{h}_{k-1}^{(\ell)}\,; \qquad 0 \le \alpha \le \alpha_{max} \quad (6)$$

where α_{max} is a positive constant, smaller than one. This filter is identical with the well known least mean square (LMS) adaptive algorithm [11]. On the other hand, the filter has the form of a low-pass filter. Thus, it reduces the effects noise on the expense of some imposed pass-band distortions.

The gain factor α is to be chosen to optimise the filter design. Generally speaking, a large α reduces the impact of AWGN in the received signal, y_k, but increases the *lag error* induced by the phase lag of the filter, and vice versa. In [12] the gain α is approximated by means of a prediction error analysis and found to be

$$\alpha_{opt} \approx 1 - 3.6\, \sqrt[3]{\nu_{max}^2 \gamma} \qquad (7)$$

Note, the approximation is only valid for modest fading rates ($\nu_{max} \le 0.04$) and for reasonable high signal–to–noise ratios (SNR) $E_b > N_0$ [12]. For larger fading rates, α_{opt} in (7) becomes negative. In this case, setting $\alpha_{opt} = 0$ is the best possible solution. In [6], [13] solutions how α can be calculated adaptively are given, without knowledge of the channel statistics.

C. Pilot aided phase tracking

The formulation of the decision variable in (3), cannot be employed in this form for antipodal modulation, such as BPSK, since the transmitted signal is not orthogonal. The resulting phase ambiguity at the receiver necessitates a phase reference, which is provided by multiplexing pilot symbols in the data stream. The multiplexed pilot symbols with rate R can be incorporated in (3), (5) and (6) by redefining the pre-multiplied received signal

$$y_k'^{(\ell)} = \begin{cases} \sqrt{E_b} h_k + n_k\,; & k \mod R = 0 \\ d_k^{(\ell)*}\, y_k\,; & \text{elsewhere.} \end{cases} \qquad (8)$$

[1] In some publications this filter is referred to as *alpha tracker*.

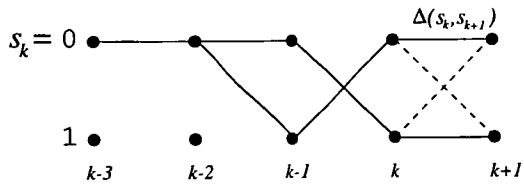


Fig. 1. Survivor Paths illustrated in a trellis for the PSP algorithm.

D. Per-survivor processing

It has been described how an exhaustive search over all possible data sequences yields the desired ML estimate. However, the number of sequences grow exponentially with the sequence length. The prohibitive high complexity of the MLSE detection, can be reduced by employing the Viterbi Algorithm (VA) [2].

Let $\widehat{\Lambda}_k^{(\iota)}$ denote the survivor path of state s_k, that is the metric with the minimum distance entering at this state, which is obtained by truncating (4) at sample k

$$\widehat{\Lambda}_k^{(\iota)} = \min_{\ell \in \tilde{\mathcal{A}}} \{\Lambda_k^{(\ell)}\} \; ; \quad s_k = \begin{cases} 0, & k \bmod R = 0 \\ \{0,1\}, & \text{elsewhere} \end{cases} \tag{9}$$

where the states $s_k = 0$ and $s_k = 1$ are equivalent to a transmitted 1 and -1, respectively. The subset $\tilde{\mathcal{A}}$ denotes the number of hypotheses entering state s_k, up to sample k. We have to distinguish between two cases, whether a data or a known symbol is being detected. A transmitted data bit results in as many survivors as states, i.e. for BPSK it is 2. If a pilot symbol was transmitted there is only one survivor, yielding a unique sequence. After a delay of R symbols only one survivor exists, giving the final data decision $\widehat{d}_{k-R}$.

To illustrate the idea of PSP it is appropriate to regard $d_k^{(\iota)}$ as states in a trellis, as shown in Fig. 1. To compute the survivors to sample $k+1$, we extend the metric from state s_k to state s_{k+1} and combine (3) with (9). From the two branches entering at state s_{k+1}, the one with the larger metric is discarded, leaving one survivor per state

$$\widehat{\Lambda}_{k+1}^{(\iota)} = \widehat{\Lambda}_k^{(\iota)} + \min_{\ell} |y'^{(\ell)}_{k+1} - \widehat{h}_{k+1}^{(\iota)}|^2. \tag{10}$$

IV. Results

Simulation work is based on a complex baseband flat fading channel. Such an environment is commonly described by a Rayleigh fading channel. To generate h_k, the complex filter response of a 4^{th} order noise shaping IIR filter is matched to the Doppler power spectra of (1), as described in [14]. The pilot multiplexing rate is set to $R = 10$, throughout this simulations. For FIR filtering, the number of coefficients is set to $M = 8$.

The bit error rate (BER) was obtained by Monte Carlo simulations. The error probability, for the idealised case

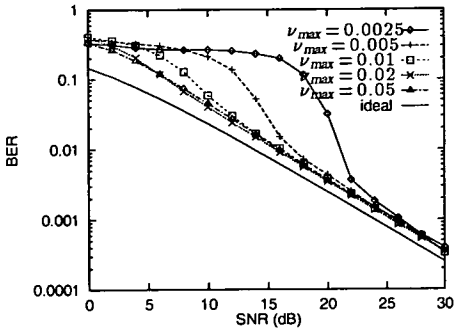


Fig. 2. BER vs SNR for various Doppler frequencies ν_{max}.

if the CIR is known *a priori*, is used for comparison purposes, and is labelled "ideal" in the diagrams. That is for a frequency-flat fading channel [15, chapter 14]

$$P_e = \frac{1}{2}\left[1 - \sqrt{\frac{\gamma}{1+\gamma}}\right]. \tag{11}$$

A. PSP using FIR channel estimation

The performance of PSP with FIR channel estimation (FIR–PSP) is considered first. In Figure 2, the BER is plotted against the SNR, $\gamma = E_b/N_0$, for various Doppler frequencies ν. It can be observed, that the slow ($\nu_{max} < 0.02$) and fast fading case has to be considered separately. For fast fading the system performance is seen to be better than for slow fading conditions. This is a rather unexpected result, as normally fast fading conditions are more complicated to deal with. It is seen that the trade-off, from which the FIR–PSP performance becomes poor is also dependent on the SNR; the performance is able to recover when ν_{max} and γ are above a certain threshold. This effect is due to stability problems of the receiver and is discussed in more detail in the following graphs. On the other hand, for Doppler frequencies larger than $\nu_{max} \approx 0.02$ the receiver is robust.

The stability problems of FIR–PSP for low SNR and low Doppler are addressed in Figure 3. It shows the phase (Figure 3.a) and magnitude (Figure 3.b) of the estimated CIR, $\widehat{h}_k^{(\iota)}$, of FIR–PSP in the time domain, compared to the CIR with and without AWGN. It is seen that the receiver can be looked in a false state, that is the channel estimator phase is flipped, shifted π relative to the CIR phase, i.e. $\arg(\widehat{h}_k^{(\iota)}) = \arg(h_k\, d_k^{(\iota)}) + \varphi_e$, where $\arg(d_k^{(\iota)}) = \pm\pi$, and φ_e denotes the prediction error phase. This can be seen in the graph between samples $k \approx [80, 170]$. That results in an error rate of virtually 100%. The receiver is entering the false state after a deep fade and may stay looked until the following deep fade, as it is shown in the graph. During a deep fade rapid phase changes and an up to 20 dB smaller signal to noise ratio may cause the channel estimator to loose track of the received signal phase. Then the prediction error due to noise becomes large compared to the CIR, h_k, resulting in a poor channel estimates. Sub-

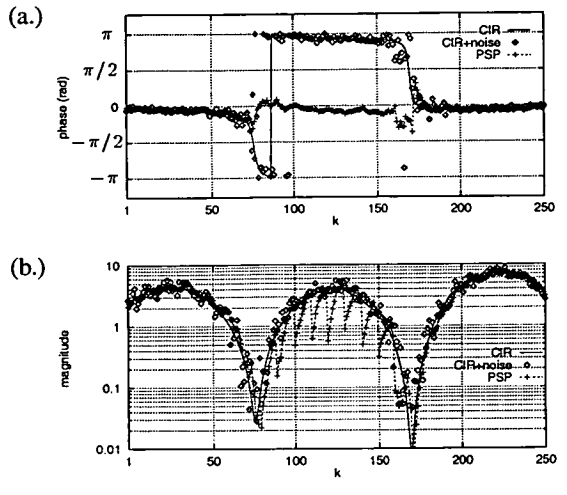


Fig. 3. Phase (a.) and magnitude (b.) of FIR–PSP vs time k, compared to the CIR with and without AWGN. $\nu_{max} = 0.005; \gamma = 10\,\text{dB}$.

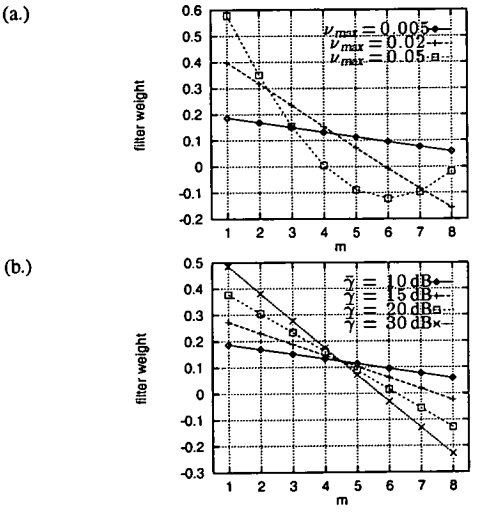


Fig. 4. The filter weights of the FIR estimation filter for various normalised Doppler ν_{max} (a.) and SNR, γ (b.). In (a.) the SNR is set to $\gamma = 10\,\text{dB}$; while in (b.) the normalised Doppler frequency is $\nu_{max} = 0.005$. $M = 8$.

sequently, in a good reception area, the phase estimate of the CIR is staying locked in the false state; despite the fact, that the estimated magnitude (Figure 3.b) diverges from the actual CIR, particularly if a pilot symbol is detected. The pilot symbols are not capable of providing a sufficient phase reference to the receiver. The decision feedback of the data symbols in estimating $\widehat{h}_k^{(i)}$ corrupts the pilot symbol phase tracking. This can cause the receiver to become unstable.

These problems occur since the VA is only a sub-optimum decoding algorithm for the described receiver. Note, the VA is only optimal as a decoding algorithm for a 1st order Markov process. The estimated CIR in (5) however, is dependent on M samples, thus leading to a M^{th} order Markov process. Thus, crucial information may be discarded by applying the VA. For slow fading, adjacent samples are more correlated, which is manifested in the filter weights $\{w_m\}$. Therefore past samples have more impact in calculating (5), and the sub-optimality of the receiver becomes more significant. As a result, the ratio between received signal powers of pilot and data symbols decreases. The phase tracking becomes less reliable and decision feedback effects may cause the receiver to lose track of the CIR phase.

The reason why this occurs more likely in slow fading and low SNR conditions is further analysed in Figure 4. Close studies show, that the filter weights $\{w_m\}$ are responsible for the observed stability problems. The filter weights are a function of the fading statistics and the SNR, with the parameters ν_{max} and γ, determined by the MMSE criterion [11]. Part (a.) shows $\{w_m\}$ matched to different Doppler frequencies ν_{max}. The lower ν_{max} becomes, the less is the gradient on the graph Figure 4.a. Note, the gradient, $\Delta w_m = w_m - w_{m+1}$, becomes smaller for slow fading and low SNR, since adjacent samples are more correlated. Suppose, a pilot is being detected; the smaller the gradient Δw_m gets, the less impact does the pilot symbol have on the channel estimation process, relative to the data symbols. Thus, the phase reference provided by the pilot may be insufficient to trace the random phase of the CIR, $\arg(h_k)$. So, the estimated CIR phase, $\arg(\widehat{h}_k^{(i)})$, may flip and stay locked at a false state, like depicted in Figure 3, leading to a high degradation in system performance. Higher Doppler frequencies or SNR values, on the other hand, as shown in Figure 4.b, increase the gradient Δw_m and hence the receiver becomes more robust.

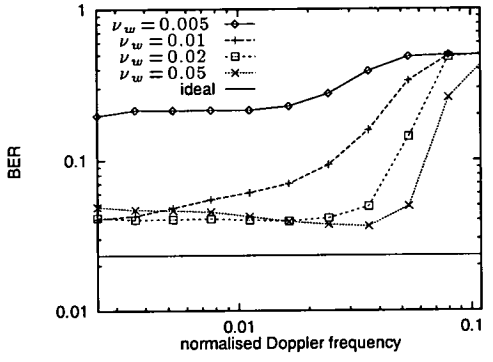


Fig. 5. BER vs normalised Doppler frequency ν_{max}, for filters $\mathbf{w}$ matched to ν_w and $\gamma_w = \gamma = 10\,\text{dB}$.

To prove this claim, let a channel estimation filter matched to the parameters γ_w and ν_w, operate in a scenario with the actual channel parameters γ and ν_{max}, illustrated in Figure 5. It is seen that a filter $\mathbf{w}$ matched to low Doppler, $\nu_w = 0.005$, has very poor performance for all Doppler frequencies, while for filters with $\nu_w \geq 0.02$, the filter shows no stability problems as long as $\nu_{max} \leq \nu_w$.

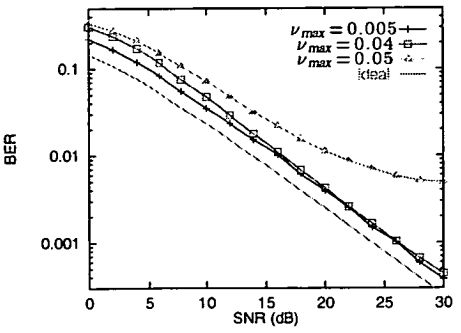


Fig. 6. BER vs SNR for various Doppler frequencies ν_{max} for IIR–PSP.

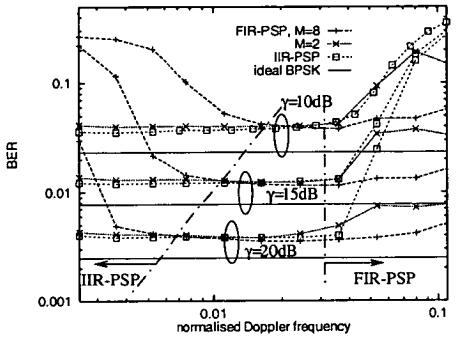


Fig. 7. BER vs normalised Doppler frequency ν_{max} for FIR–PSP and IIR–PSP.

B. PSP using IIR channel estimation

Different results are obtained when an IIR filter is used for channel estimation, shown in Figure 6. It can be observed from Figure 6 that the IIR–PSP performance is excellent for slow fading, under the constraint that $\alpha_{max} \leq 0.5$, in (6). Simulation results for $\alpha_{max} > 0.5$ suggest, that the IIR–PSP suffers from similar stability problems as the FIR–PSP discussed before. In the following α_{max} is set to 0.5. On the other hand, IIR–PSP degrades significantly for Doppler frequencies $\nu_{max} > 0.04$. The observed irreducible bit error rate (IBER) is due to the induced phase lag of the IIR filter, being essentially a low-pass filter. The IIR–PSP is therefore not applicable for fast fading.

C. Hybrid filtering PSP

It has been shown that the FIR–PSP performs superior to the IIR–PSP for fast fading, while the opposite is true for slow fading. Hence in order to design a receiver which can operate in arbitrary fading conditions a hybrid solution becomes attractive. This is shown in Figure 7. It is seen that there is an great overlap between FIR–PSP and IIR–PSP where both estimation filters have virtually the same performance. So, provided that there is a rough estimate of the maximum Doppler spread ν_{max} available, the hybrid receiver switches on changing fading rates between the FIR and IIR estimation filter.

V. Conclusions

A hybrid realisation of per-survivor processing (PSP) was studied, which yields very good performance for a large range of Doppler frequencies. The performance of PSP employing a FIR channel estimation filter is excellent for fast fading channels, while it is not robust for slow fading. The stability problems of FIR–PSP are a rather unexpected result. On the other hand IIR filter channel estimation works very well for slow fading, but it is not suitable for fast fading. Simulation results also showed a large overlap, where the FIR and IIR–PSP have the same performance. The proposed hybrid PSP switches from IIR to FIR filtering dependent on the Doppler frequency and the SNR, requiring only rough estimates of the channel statistics.

References

[1] T. Kailath, "Correlation Detection for Signal Perturbed by a Random Channel," *IRE Trans. Inf. Theory*, vol. IT-6, pp. 361–66, June 1960.

[2] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–78, March 1973.

[3] R. Raheli, A. Polydoros, and C. Tzou, "Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments," *IEEE Trans. Commun.*, vol. COM-43, pp. 354–364, Feb.-Apr. 1995.

[4] M. L. Moher and J. H. Lodge, "TCMP—A Modulation and Coding Strategy for Rician Fading Channels," *IEEE Journal Sel. Areas Commun.*, vol. SAC-7, pp. 1347–55, Sep. 1989.

[5] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Trans. Veh. Tech.*, vol. VT-40, pp. 686–693, Nov. 1991.

[6] P. Y. Kam and P. Sinha, "A Viterbi Type Algorithm for Efficient Estimation of M-PSK Sequences of the Gaussian Channel with Unknown Carrier Phase," *IEEE Trans. Commun.*, vol. COM-43, pp. 2429–33, Sep. 1995.

[7] G. M. Vitetta and D. P. Taylor, "Maximum Likelihood Decoding of Uncoded and Coded PSK Signal Sequences Transmitted over Rayleigh Flat-Fading Channels," *IEEE Trans. Commun.*, vol. COM-43, pp. 2750–58, Nov. 1995.

[8] A. N. D'Andrea, A. Diglio, and U. Mengali, "Symbol-Aided Channel Estimation with Nonselective Rayleigh Fading Channels," *IEEE Trans. Veh. Tech.*, vol. VT-44, pp. 41–48, Feb. 1995.

[9] R. H. Clarke, "A Statistical Theory of Mobile-Radio Reception," *Bell System Technical Journal*, vol. 47, pp. 957–1000, Jul-Aug 1968.

[10] J. H. Lodge and M. L. Moher, "Maximum Likelihood Sequence Estimation of CMP Signals Transmitted over Rayleigh Flat Fading Channels," *IEEE Trans. Commun.*, vol. COM-38, pp. 787–94, June 1990.

[11] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 2nd ed., 1991.

[12] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. Wiley Interscience, 2nd ed., 1998.

[13] F. Adachi, "Adaptive Reception of MPSK Signals in Fast Fading and AWGN Channels," *IEE Electronics Letters*, vol. 32, pp. 1944–45, Oct. 1996.

[14] D. I. Laurenson and G. J. R. Povey, "Channel Modelling for a Predictive RAKE Receiver System," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), The Hague, The Netherlands*, pp. 715–719, Sep 1994.

[15] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, NY, USA, 3rd ed., 1995.