MLSE DIVERSITY RECEIVER STRUCTURES

A thesis
submitted in fulfilment
of the requirements for the Degree of
Doctor of Philosophy
in Electrical and Electronic Engineering
from the
University of Canterbury
Christchurch, New Zealand

Brian Hart, BE(Hons)

University of Canterbury
1996
This thesis presents the design and analysis of MLSE diversity receivers for linearly modulated signals transmitted over known and unknown time- and frequency-selective channels, corrupted by additive Gaussian noise. The extended MLSE receiver structure of Ungerboeck [56] is extended further; for the case of a known but time-varying, frequency-selective channel with diversity. The error event analysis technique of Forney [22] is used to approximate and bound the receiver's BER. The MLSE predictor receiver of Yu and Pasupathy [64] for unknown Rayleigh fading channels is also extended, to the case of Ricean fading, correlated diversity threads, uncertain carrier frequency and phase, and unknown symbol timing. The received signal's second order statistics are needed to compute the predictors, and two methods are proposed that achieve this in the Rayleigh fading channel. The MLSE predictor receiver's BER is bounded, assuming ideal knowledge of the received signal's second order statistics.
Many people have assisted me to complete this thesis, and I acknowledge their help gratefully. There is my supervisor, Professor Des Taylor, who prompted the research, indicated new directions, and whose tireless reviews of interim material I thoroughly appreciate. There are the members of the Comms Lab., with whom I have shared several brainstorming (and hacky-sack) sessions: Robert van Nobelen, Wing Seng Leon, Katharine Holdsworth, Mathew Hebley, Ben Benison, Anthony Griffin, and Nick Baas. Then there are my family and flatmates, who have kept me sane and my feet on the ground. I thank you all.
# Table of Contents

ABSTRACT ................................................................................................................. iii

ACKNOWLEDGEMENTS ........................................................................................... v

NOTATION ................................................................................................................ xiii

1 INTRODUCTION ...................................................................................................... 1
   1.1 General .............................................................................................................. 1
   1.2 Background ...................................................................................................... 2
   1.3 System Goals ................................................................................................... 3
   1.4 Content of the Thesis ...................................................................................... 4
   1.5 Limitations of the Thesis ............................................................................... 7
   1.6 Literature Review ........................................................................................... 7
   1.7 Thesis Contributions ...................................................................................... 12

2 THE TRANSMITTER ............................................................................................. 13
   2.1 Introduction .................................................................................................... 13
   2.2 Modulation ................................................................................................... 13
   2.3 Constellations ............................................................................................... 15
   2.4 Pulse Shapes .................................................................................................. 15
   2.5 Coding .......................................................................................................... 17
   2.6 Summary ....................................................................................................... 19

3 THE CHANNEL ...................................................................................................... 21
   3.1 Introduction ................................................................................................... 21
   3.2 Receiver Translation ..................................................................................... 21
   3.3 Physics of the Multipath Channel .................................................................. 23
   3.4 The Tapped Delay-Line Model ..................................................................... 26
   3.5 The Received Pulse Shape Model ................................................................. 28
   3.6 Covariance Functions for Characterising the Channel .................................. 29
   3.7 A Statistical Description of the Multipath Channel ....................................... 32
   3.8 Complex Gaussian Distributions ................................................................... 37
   3.9 Simulating the Channel ................................................................................ 40
   3.10 Implicit Diversity ........................................................................................ 44
   3.11 The Time- and Frequency- Selective Channel ........................................... 47
   3.12 The Time-Selective Channel ....................................................................... 48
   3.13 The Frequency-Selective Channel ............................................................... 50
   3.14 The Additive White Gaussian Noise Channel ............................................ 51
   3.15 Noise Sources ............................................................................................. 51
   3.16 Summary ..................................................................................................... 53

4 THE RECEIVER ..................................................................................................... 55
   4.1 Introduction ................................................................................................... 55
   4.2 Statistical Detection Theory .......................................................................... 55
# TABLE OF CONTENTS

4.3 Analysing MLSE Receiver Structures ............................................................... 59  
4.4 Reduced Complexity Receiver Structures ......................................................... 61  
4.5 Carrier Recovery .............................................................................................. 63  
4.6 Equalisation ..................................................................................................... 68  
4.7 Pilot Information ............................................................................................... 71  
4.8 Diversity ............................................................................................................. 73  
4.9 Summary ............................................................................................................ 76  

5 DESIGN OF THE EXTENDED MLSE RECEIVER .............................................. 79  
5.1 Introduction ....................................................................................................... 79  
5.2 EMLSE-tf/Diversity Receiver Derivation ........................................................... 80  
5.3 White Noise ...................................................................................................... 84  
5.4 EMLSE-tf/Diversity Receiver Operation ........................................................... 85  
5.5 Maximal Ratio Combining ............................................................................... 86  
5.6 The Frequency-Selective Channel .................................................................... 88  
5.7 The Time-Selective Channel ............................................................................ 89  
5.8 The Awn Channel ............................................................................................. 90  
5.9 Short Pulses ...................................................................................................... 90  
5.10 The Discrete-Time EMLSE-tf ......................................................................... 91  
5.11 Pilot Tones And Pilot Symbols ......................................................................... 92  
5.12 Summary ......................................................................................................... 95  

6 PERFORMANCE EVALUATION OF THE EXTENDED MLSE RECEIVER .... 97  
6.1 Introduction ....................................................................................................... 97  
6.2 Performance Evaluation ..................................................................................... 97  
6.3 A Simple Analytic Example ............................................................................ 104  
6.4 Performance Overview ..................................................................................... 105  
6.5 Transmitter Parameters .................................................................................... 109  
6.6 Channel Parameters ......................................................................................... 110  
6.7 Receiver Parameters ......................................................................................... 113  
6.8 Pilot Tones ...................................................................................................... 115  
6.9 Summary ......................................................................................................... 116  

7 DESIGN OF THE MLSE PREDICTOR RECEIVER ......................................... 117  
7.1 Introduction ....................................................................................................... 117  
7.2 Transmitter Design ........................................................................................... 119  
7.3 Receiver Derivation ........................................................................................... 120  
7.4 Independent and Nearly Independent Threads ............................................... 125  
7.5 The Rayleigh Fading Channel ......................................................................... 126  
7.6 The Yu and Pasupathy Receiver Structure ...................................................... 128  
7.7 The Vitetta, Dam, Holdsworth, and Taylor Receiver Structure ....................... 129  
7.8 Receiver Operation ........................................................................................... 131  
7.9 Pilot Symbols .................................................................................................. 134  
7.10 Estimating The Signal Autocorrelation .......................................................... 135  
7.11 The Blind Receiver ......................................................................................... 139  
7.12 Summary ....................................................................................................... 140
TABLE OF CONTENTS

8 PERFORMANCE EVALUATION OF THE MLSE PREDICTOR RECEIVER STRUCTURE ............................................................................................................ 141
  8.1 Introduction ..................................................................................................... 141
  8.2 Receiver Analysis ............................................................................................ 141
  8.3 A Simple Analytic Example ............................................................................ 144
  8.4 Simulation ....................................................................................................... 147
  8.5 Comparison Overview .................................................................................... 148
  8.6 Transmitter Parameters ................................................................................... 151
  8.7 Channel Parameters ......................................................................................... 153
  8.8 Receiver Parameters ....................................................................................... 157
  8.9 Adaptation ...................................................................................................... 160
  8.10 Summary ...................................................................................................... 162

9 CONCLUSIONS .................................................................................................... 163
  9.1 Achievements .................................................................................................. 163
  9.2 Future Research ............................................................................................... 164
  9.3 Future Systems ................................................................................................ 165

A LINEAR PREDICTION........................................................................................ 169
  A.1 General Properties .......................................................................................... 169
  A.2 Perfect Prediction ........................................................................................... 173

B GAUSSIAN QUADRATIC FORMS AND RESIDUES ........................................ 177

C THE SEQUENCE-AVERAGED PREDICTOR RECEIVER .................................. 183
  C.1 Introduction .................................................................................................. 183
  C.2 The Sequence-Averaged Predictor Receiver .................................................. 186

REFERENCES ........................................................................................................... 193
### TABLE OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A radially-symmetric constellation.</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Block diagram of the transmitter.</td>
<td>19</td>
</tr>
<tr>
<td>3.1</td>
<td>Block diagram of the receiver’s RF/IF chain.</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Continuous time tapped delay line model.</td>
<td>27</td>
</tr>
<tr>
<td>3.3</td>
<td>A sample fading process’ amplitude in the vicinity of the receiver.</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Sample fading process.</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Properties of simulated fading processes.</td>
<td>42</td>
</tr>
<tr>
<td>3.6</td>
<td>The $t$-channel.</td>
<td>48</td>
</tr>
<tr>
<td>3.7</td>
<td>The $\tau$-channel.</td>
<td>49</td>
</tr>
<tr>
<td>3.8</td>
<td>The $f$-channel.</td>
<td>50</td>
</tr>
<tr>
<td>3.9</td>
<td>The AWGN channel.</td>
<td>51</td>
</tr>
<tr>
<td>3.10</td>
<td>Block diagram of the transmitter, channel and an analogue receiver.</td>
<td>52</td>
</tr>
<tr>
<td>4.1</td>
<td>Conventional receiver structures.</td>
<td>62</td>
</tr>
<tr>
<td>4.2</td>
<td>Carrier recovery in the distortionless channel.</td>
<td>62</td>
</tr>
<tr>
<td>4.3</td>
<td>Differential demodulation error mechanisms and BER curves.</td>
<td>67</td>
</tr>
<tr>
<td>4.4</td>
<td>Linear equaliser with decision feedback adaptation.</td>
<td>68</td>
</tr>
<tr>
<td>4.5</td>
<td>Diversity example.</td>
<td>74</td>
</tr>
<tr>
<td>5.1</td>
<td>The EMLSE-$t$-space diversity receiver.</td>
<td>86</td>
</tr>
<tr>
<td>5.2</td>
<td>The time-frequency strip for pilot symbols.</td>
<td>93</td>
</tr>
<tr>
<td>5.3</td>
<td>Fourier transform of the transmitter pulse shape.</td>
<td>93</td>
</tr>
<tr>
<td>6.1</td>
<td>BER-SNR curves for different union bounds</td>
<td>107</td>
</tr>
<tr>
<td>6.2</td>
<td>BER-SNR curves for different union bounds and $f_D T = 0.01$</td>
<td>107</td>
</tr>
<tr>
<td>6.3</td>
<td>BER-SNR curves for various signal constellations.</td>
<td>108</td>
</tr>
<tr>
<td>6.4</td>
<td>BER-SNR curves for different pulse lengths, $H$.</td>
<td>108</td>
</tr>
<tr>
<td>6.5</td>
<td>BER-SNR curves for various excess bandwidths, $\chi$.</td>
<td>109</td>
</tr>
<tr>
<td>6.6</td>
<td>BER-SNR curves for increasing delay spread, $\tau/T$, and $f_D T = 0.001$.</td>
<td>110</td>
</tr>
<tr>
<td>6.7</td>
<td>BER-SNR curves for increasing Doppler spread, and $\tau/T = 0.005$.</td>
<td>111</td>
</tr>
<tr>
<td>6.8</td>
<td>BER-SNR curves for increasing delay spread, $\tau/T$.</td>
<td>111</td>
</tr>
<tr>
<td>6.9</td>
<td>BER-SNR curves for increasing Doppler spread, $f_D T$.</td>
<td>112</td>
</tr>
<tr>
<td>6.10</td>
<td>BER-SNR curves for constant $f_D t = 0.0005$.</td>
<td>112</td>
</tr>
<tr>
<td>6.11</td>
<td>BER-SNR curves for an increasing number of paths, $P$.</td>
<td>113</td>
</tr>
<tr>
<td>6.12</td>
<td>BER-SNR curves for different sampling rates, $r$.</td>
<td>114</td>
</tr>
<tr>
<td>6.13</td>
<td>BER-SNR curves for increasing diversity threads, $D$.</td>
<td>114</td>
</tr>
<tr>
<td>6.14</td>
<td>BER-SNR curves for varying antenna separation, $W$.</td>
<td>115</td>
</tr>
<tr>
<td>6.15</td>
<td>BER-SNR curves comparing ideal CSI and pilot tone CSI.</td>
<td>116</td>
</tr>
<tr>
<td>7.1</td>
<td>Computational flow of the MLSE predictor receiver without estimation.</td>
<td>125</td>
</tr>
<tr>
<td>7.2</td>
<td>Operation of predictor receiver for the $\tau$-channel.</td>
<td>132</td>
</tr>
<tr>
<td>7.3</td>
<td>Computational flow for the MLSE predictor receiver with estimation.</td>
<td>140</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>8.1</td>
<td>BER-SNR curves for different union bounds</td>
<td>150</td>
</tr>
<tr>
<td>8.2</td>
<td>BER-SNR curves for different union bounds, with $f_D T = 0.01$</td>
<td>150</td>
</tr>
<tr>
<td>8.3</td>
<td>BER-SNR curves for various signal constellations</td>
<td>152</td>
</tr>
<tr>
<td>8.4</td>
<td>BER-SNR curves for different pulse lengths, $H$</td>
<td>153</td>
</tr>
<tr>
<td>8.6</td>
<td>BER-SNR curves for different delay spreads, $\tau / T$</td>
<td>154</td>
</tr>
<tr>
<td>8.7</td>
<td>BER-SNR curves for different delay spreads, $\tau / T$, and $f_D T = 0.01$</td>
<td>154</td>
</tr>
<tr>
<td>8.8</td>
<td>BER-SNR curves for different values of the normalised Doppler spread, $f_D T$</td>
<td>155</td>
</tr>
<tr>
<td>8.9</td>
<td>BER-SNR curves for an increasing number of paths, $P$</td>
<td>155</td>
</tr>
<tr>
<td>8.10</td>
<td>BER-SNR curves for different tap autocorrelation models</td>
<td>156</td>
</tr>
<tr>
<td>8.11</td>
<td>BER-SNR curves for various predictor lengths, $B$</td>
<td>157</td>
</tr>
<tr>
<td>8.12</td>
<td>BER-SNR curves for various predictor lengths, $B$, and a Dam signal model</td>
<td>158</td>
</tr>
<tr>
<td>8.13</td>
<td>BER-SNR curves for an increasing sampling rate, $r$</td>
<td>158</td>
</tr>
<tr>
<td>8.14</td>
<td>BER-SNR curves for various $B$, where $B/r$ is kept constant</td>
<td>159</td>
</tr>
<tr>
<td>8.15</td>
<td>BER-SNR curves for increasing diversity, $D$</td>
<td>160</td>
</tr>
<tr>
<td>8.16</td>
<td>BER-SNR curves for various training lengths: method 1</td>
<td>161</td>
</tr>
<tr>
<td>8.17</td>
<td>BER-SNR curves for various training lengths: method 2</td>
<td>161</td>
</tr>
<tr>
<td>A.1</td>
<td>Structure of linear predictor</td>
<td>172</td>
</tr>
<tr>
<td>B.1</td>
<td>Contours of integration</td>
<td>178</td>
</tr>
<tr>
<td>C.1</td>
<td>Operation of sequence-averaged prediction</td>
<td>185</td>
</tr>
<tr>
<td>C.2</td>
<td>Computational flow in the sequence-averaged predictor receiver</td>
<td>190</td>
</tr>
<tr>
<td>C.3</td>
<td>A comparison of the MLSE predictor receiver and the SA predictor receiver</td>
<td>191</td>
</tr>
</tbody>
</table>
NOTATION

ACI - adjacent channel interference
A-D - analogue-to-digital
ASK - amplitude shift keying
ATM - asynchronous transfer mode
AWGN - additive white Gaussian noise
BER - bit error rate
BPSK - binary PSK
CCI - co-channel interference
CLT - central limit theorem
CPM - continuous phase modulation
CSI - channel state information
EMLSE-x - extended MLSE receiver for the x-channel
FIR - finite impulse response
GQF - Gaussian quadratic form
IF - intermediate frequency
i.i.d. - independently and identically distributed (circularly symmetric)
IIR - infinite impulse response
ISI - intersymbol interference
KL - Karhunen-Loéve
LMS - least mean squares
LO - local oscillator
LOS - line of sight
ML - maximum likelihood
MLSE - maximum likelihood sequence estimation
MMSE - minimum mean square error
MRC - maximal ratio combining
MSE - mean square error
PLL - phase-lock loop
PSK - phase shift keying
PSP - per-survivor processing
QAM - quadrature amplitude modulation
QPSK - 4-PSK
RF - radio frequency
RLS - recursive least squares
RTZ - return to zero
SAMMSE - sequence-averaged MMSE
SAMSE - sequence averaged MSE
SNR - signal-to-noise ratio
SVD - singular value decomposition
TDMA - time-division, multiple access
TTIB - transmit tone in-band
US - uncorrelated scattering
VCO - voltage-controlled oscillator
WSS - wide sense stationary
WSSUS - wide-sense stationary, uncorrelated scattering
time-selective channel, \( t \)-channel, fading, time-varying, multiplicative, frequency-flat
frequency-selective channel, \( f \)-channel, dispersive, time-invariant, ISI channel
time- and frequency-selective channel, \( tf \)-channel

\( \bar{x} \) - a baseband quantity
\( x_c \) - a quantity at the carrier frequency
\( \hat{x} \) - an estimated or hypothesised quantity
\( x' \) - a counter related to but different from \( x \); derivative of \( x \)
\( x_R, x_I \) - the real and imaginary (in-phase and quadrature) components of \( x \)
\( \bar{x} \) - the complex conjugate
\( E(x) \) - the expectation
\( x \) - a vector
\( X \) - a matrix
\( x* y \) - convolution
\( \lfloor x \rfloor, \lceil x \rceil \) - round down and up to the nearest integer
\( \lfloor x/y \rfloor, x \mod y \) - quotient and remainder of \( x + y \); \( x, y \) integers
\( |x|, \angle x \) - amplitude and phase
\( \Re\{x\}, \Im\{x\} \) - real and imaginary part

\( \alpha_p \) - uncoded data
\( \beta_p \) - transmitted symbol as a complex phasor
\( \chi \) - excess bandwidth of a raised cosine or square root raised cosine pulse; channel
plus noise
\( \delta \) - a small amount; Dirac delta function; Kronecker delta function
\( \varepsilon_p \) - error phasor
\( \phi \) - phase offset between transmitter and receiver oscillators
\( \gamma_{hr} \) - hypothesised receiver phasor
\( \eta \) - synchronisation parameters
\( \varphi \) - direction of mobile's velocity
\( \kappa \) - a GQF
\( \kappa_{mn} \) - bias term in analysing the MLSE predictor receiver
\( \lambda \) - eigenvalue
\( \lambda_{bp} \) - branch metric
\( \mu_{bp} \) - possibly time-varying forgetting factor
\( \pi \) - pi
\( \theta \) - path arrival angle
\( \rho \) - radial constellation scale factor
\( \rho_k \) - single scatterer attenuation
\( \rho_i \) - path attenuation
\( \rho_p \) - power fraction devoted to pilot tones or symbols
\( \sigma_{fp} \) - Viterbi algorithm state
\( \sigma^2 \) - variance
\( \tau \) - maximum two-sided delay spread
\( \tau_{rms} \) - rms delay spread
\( \omega \) - index of first received sample to be used in a prediction with erroneous predictor tap weights
\( \xi \) - delay variable of convolution
\( \xi_F \) - delay of first path
\( \xi_L \) - delay of last path
\( \psi \) - index of first symbol that affects the pairwise probability of an error event
\( \zeta \) - argument of characteristic function

\( \Delta \) - a change
\( \Delta(t, \xi) \) - CSI estimation error
\( \Phi \) - index of last received sample to contain a pulse tail from an errored symbol
\( \Gamma \) - gamma function
\( \Theta \) - antenna array orientation
\( \Lambda_p \) - path metric
\( \Pi \) - product operator
\( \Sigma \) - summation operator
\( \Omega \) - index of last received sample to be used in a prediction with erroneous predictor tap weights
\( \Psi \) - index of final symbol that affects the pairwise probability of an error event

\( a(t), a_t \) - transmitted signal
\( b_t \) - predictor tap weights
\( c(t-iT, iT), c_{t,ir,ir} \) - received pulse, encompassing transmitter pulse and channel
\( c_k \) - KL coefficients
\( c \) - speed of light
\( d \) - counter for diversity branches (as superscript)
\( e \) - used often as an error
\( e(t-iT, iT), e_{t,ir,ir} \) - received pulse estimation error
\( e(\beta^{v,w} \rightarrow \beta^{v,v,w}) \) - number of bit errors from an error event
\( f \) - frequency
\( f_c \) - carrier frequency
\( f_{c,\text{nom}} \) - estimated carrier frequency
\( f_{0,\text{nom}} \) - residual carrier offset
\( f_D \) - maximum one-sided Doppler frequency
\( f_H \) - frequency location of pilot tone
\( f_s \) - maximum signal frequency
\( g \) - vector of complex Gaussian random variables
\( h(t) \) - transmitter pulse shape
\( i, k \) - counters, usually used for symbol-rate quantities
\( j \) - square root of -1
\( l, m \) - counters, usually used for Nyquist rate quantities such as sampled time
\( m_p \) - matched filter terms from the EMLSE-tf receiver
\( n(t), n_t \) - noise
\( n_v \) - non-random (as superscript)
\( p_r(x) \) - pdf
\( p_t \) - pole of the characteristic function
\( r \) - number of samples per symbol; random (as superscript)
\( \mathcal{r}_t \) - residue
\( s_{\nu,kr} \) - ISI terms from the EMLSE-tf receiver
\( t \) - time
\( t_0 \) - timing error
\( t_d \) - interleaving delay
\( t_{ps} \) - time location of a pilot symbol
\( u \) - error event length
\( v \) - index for transmitted sequence; mobile's speed
\( w \) - index for error event
\( x \) - a dummy; position
\( y(t), y_t \) - received signal
\( z(t, \xi) \) - multipath tf-channel
\( z(t), z_f \) - t-channel

\( A \) - number of points per constellation sector
\( B \) - number of predictor tap weights
\( C_x^y \) - combination
\( D \) - number of diversity branches
\( E \) - length of an error event
\( E \) - inverse MMSE matrix from Cholesky decomposition
\( G \) - matrix of quadratic terms present in a GQF
\( H \) - length of transmitter pulse shape
\( H(f) \) - Fourier transform of pulse shape
\( I_B \) - beginning of transmission
\( I_E \) - end of transmission
\( J_\nu(x) \) - Bessel function of \( \nu \)th order
\( J \) - Jacobian
\( L \) - length of received pulse shape
\( L \) - predictor matrix from Cholesky decomposition
\( M \) - constellation size
\( O \) - polynomial order
\( P \) - number of rotationally symmetric sectors in a constellation; number of independently faded paths in a tf-channel; probability operator
\( P_\zeta(\zeta) \) - characteristic function
\( R_{xx} \) - the autocorrelation of random processes \( x \) and \( y \)
\( S \) - number of branches in a suboptimal predictor receiver
\( S_{xx} \) - power spectral density of random process \( x \)
\( S \) - square root of matrix (from the Cholesky decomposition)
\( T \) - symbol period
\( T(f,t) \) - time variant transfer function
\( T_0 \) - maximum timing uncertainty
\( T_s \) - sampling period
\( W \) - length of the MLSE predictor receiver's hypothesis vector; number of wavelengths between antennae
\( Y \) - vector of past received samples
\( y \) - sample from the received signal, with thread samples interleaved
\( \Psi \) - vector of past samples from the received signal, with thread samples interleaved
1 INTRODUCTION

1.1 GENERAL

The theory and practice of digital communications is relatively recent. It is
tremendously successful however, spawning highly integrated, global telephone and
data networks. In parallel with these fixed wire services, there has been an explosive
growth in the market for mobile communications. The need to support more mobile
voice users and the new data-based services drives the technology towards digital
techniques. There are new problems, requiring novel solutions for future services.

What is the ultimate destination of mobile communications? This author
argues that mobile communications should follow the direction of fixed wire
communications, its more mature cousin. The collected wisdom of decades of fixed
wire network development is presently culminating in the Asynchronous Transfer
Mode (ATM) standard. It is designed to be a universal digital carrier, and accordingly
it offers high bandwidth, a low bit error rate (BER), and a low delay. This author
argues that the various existing and proposed mobile standards are merely stepping
stones, until ATM-like properties are replicated on a mobile communications
network.

However, there are substantial additional obstacles in mobile communications,
whether or not ATM-like networks are considered. The available bandwidth is
fundamentally limited, media access is less straightforward, there is multipath and
shadowing, and the mobile's power consumption is tightly constrained.

The problems can be ameliorated in two ways. First, spectrum regulatory
authorities should ensure that the spectrum is used only when mobility is involved: the
spectrum should not be squandered on fixed-point-to-fixed-point communication,
such as microwave line-of-sight. Second, research is needed on protocols, on high
frequency devices, and on improving the bandwidth and power efficiency of mobile
links through coding, modulation and improved receiver structures. This thesis
synthesises improved receiver structures and analyses their performance.
1.2 BACKGROUND

In many real situations, transmitted signals are distorted before they arrive at the receiver. This distortion can be represented mathematically, where the received signal is some (perhaps random) function of the transmitted signal and interfering signals.

In this thesis, radio transmission has particular importance. The environment is cluttered with hills, buildings, vehicles and trees. Signals radiated from an antenna can be represented validly as rays, since wavelengths at radio frequencies (RF) are small compared to the physical structures. These rays reflect off the structures in all directions, and some arrive in the vicinity of the receiving antenna. Each ray is delayed and attenuated differently. The signal attenuation worsens when the distance between transmitter and receiver increases (usually modelled as an inverse power relationship), when there are large obstacles between transmitter and receiver (shadowing), and when the modulated travelling waves superimpose and create a standing wave interference pattern (multipath propagation). When the environment or either antenna moves, the interference pattern at the receiver’s antenna changes. Accordingly, multipath appears as a time-varying effect. In addition, the receiver’s front-end adds thermal noise, having a Gaussian probability density function (pdf), and there may be transmitters on the same channel, introducing co-channel interference (CCI), or imperfectly filtered transmitters on adjacent channels, causing adjacent channel interference (ACI). Thus the received signal is the distorted transmitted signal plus additive noise, CCI, and ACI.

Other channels have different properties. The analogue telephone channel is dispersive, time-invariant, and has a high signal-to-noise ratio (SNR). Single-mode optical fibres are time-invariant and frequency-flat, but introduce shot noise. The satellite channel is non-linear, due to non-linear power amplification within the satellite.

The purpose of a receiver is to detect the transmitted information as accurately as possible, for a given channel model. Much early research considered the telephone channel and the microwave line-of-sight (LOS) channel. These channels are frequency-selective (f-channels), but the microwave LOS channel also varies slowly in
time. Receivers for these channels were designed under the assumption of a time-
invariant channel, but they incorporate an adaptive channel estimation sub-system or
equaliser to accommodate the slowly changing channel.

More recently, the demand for digital cellular telephony and trunked radio has
prompted interest in the mobile radio channel. The channel can vary swiftly for fast-
moving mobiles, leading to a BER floor when receivers designed for the time-
invariant channel are used. Accordingly, there is interest in better receiver structures
designed specifically for the time-varying nature of the channel.

1.3 SYSTEM GOALS

Since this thesis is concerned with designing improved receiver structures, it is
worthwhile to define first the design objectives of a communications system.

First, there is low cost. A system must be economically attractive to the
system operator and the mobile terminals must be good value to the mobile user.
Since this thesis is concerned mainly with digitally implemented receiver signal
processing algorithms, the low cost objective is manifested as a low complexity goal,
particularly in the mobile terminals.

Second, there is capacity. Both the burgeoning population of mobile users and
the future high data-rate services increase the demands on the finite spectrum. The
efficient use of the spectrum is critical to mobile communications in the future. The
transmitted signals are multiplexed together while retaining separation through
various hierarchies of space, frequency, time and code division. For instance, cellular
time-division multiple access (TDMA) systems are allocated a wide frequency band
(frequency division); the coverage area is divided into cells (space division); each cell
is allocated several channels (frequency division); and each channel is split into a
sequence of packets (time division). Each division scheme should be individually
optimised to improve the total capacity. For instance, appropriate base station siting
can increase the cell density; adaptive channel allocation can optimise the frequency
reuse pattern for a mobile user distribution; bandwidth efficient pulses and large
signal constellations increase the number of available channels; and source encoding reduces the data rate that the system conveys.

Third, there is power efficiency. A system may operate thousands of base station transmitters, and so the cost of the transmitted energy adds up. Power efficiency is considerably more important to the mobile user. Reduced radiated power lowers the battery drain, leading to smaller, lighter weight mobile terminals with longer operating times. The effect of radiation on living tissue is also reduced.

Fourth, there are the BER and probability of outage. A low BER and probability of outage are needed for reliable and intelligible voice. A low probability of outage is essential to reduce the number of retransmissions with data networks. A BER of $\sim 10^{-3}$ to $10^{-4}$ is sufficient for compressed voice. For data, a BER better than $10^{-6}$ is preferred with random errors.

Fifth, there is delay. For conversation, low delay is necessary if the listener is to interrupt the speaker. Low delay is also best for interactive data services and protocols. However, broadcast services and simplex tasks such as file transfer are insensitive to delay.

1.4 CONTENT OF THESIS

Chapters 2, 3, and 4 detail introductory material on a mobile communications system. This provides a concrete foundation for the rest of the thesis. An intuitive appreciation of the problem and a precise mathematical model are developed in parallel. In chapter 2, the transmitter is reviewed briefly. In chapter 3, the multipath channel is examined in detail. Its mathematical representation as a time-varying linear filter is derived from first principles, and special cases are identified. Noise sources are identified. In chapter 4, aspects of statistical detection theory are presented. The sub-systems of a conventional receiver are described, and their degraded performance in the fast fading channel is noted.

In the body of the thesis, statistical detection theory is used to develop optimal receiver structures. In particular, for equiprobable symbols, the technique of
maximum likelihood sequence estimation (MLSE) leads to the optimal receiver structure.

The MLSE receiver structure is interesting for several reasons. First, the MLSE receiver maximises the mean time between error events. Second, at moderate to high SNRs, the MLSE receiver's BER performance is essentially a lower bound on the achievable BER of any receiver in the same channel [22]. Thus this BER performance can be used as a benchmark for lower-complexity receivers. Currently, most researchers compare their proposed receiver structures to existing receivers designed for the time-invariant channel. Naturally their proposed receiver structures offer improved BER performance in the time-varying channel, but of more interest is the performance degradation from the optimal receiver structure to their proposed receiver structures. Third, insight into the proper receiver processing for time- and frequency-selective channels (tf-channels) can be gained. With this insight, it may be easier to design reduced complexity receiver structures.

The complexity of MLSE receiver structures is high, and increases exponentially with the delay span of the intersymbol interference (ISI). For many transmitter/channel combinations, the MLSE receiver structure is currently infeasible. However, digital hardware gets faster and more powerful year by year. "Infeasible" is a nebulous boundary: what is now impractical may not be so in the future. Accordingly this thesis does not dismiss a receiver structure merely because it is complicated. In fact, given the choice between generality and special cases, optimality and mild sub-optimality, or high complexity and reduced complexity, this thesis always pursues general, optimal receiver structures, whatever the cost in complexity. In this way, the thesis' value should be more persistent.

In chapter 5, an MLSE diversity receiver structure is designed, assuming ideal knowledge of the synchronisation parameters and the channel. That is, the received signal is at a known carrier frequency and phase, regularly sampled at known points, and the received signal's distortion in time and frequency is known. Accordingly, only the noise is a random variable, and the receiver structure applies for all tf-channels, where its ideal operation guarantees that it is optimal. The presence of pilot tones or pilot symbols is considered as an aid to gain near-ideal knowledge of the channel. In chapter 6, bounds on the receiver structure's BER performance are...
deduced for fast Rayleigh fading, frequency-selective channels. The analysis employs Gaussian quadratic forms (GQF), and their mathematics is described in Appendix B. Using these bounds, the receiver's sensitivity to various transmitter and channel parameters is examined in chapter 8.

In chapter 7, an MLSE diversity receiver structure is designed assuming no synchronisation and no knowledge of the instantaneous channel. The received signal is centred at some nominally known but uncertain carrier frequency, with symbols transmitted at a known symbol rate but with an unknown transmission beginning. Accordingly, these parameters are unknown but not random variables. However the channel and noise are still random processes, and this leads to a fundamentally different receiver structure from the previous situation. Predictors underpin the channel estimation sub-system, and their properties are detailed in Appendix A. The channel is constrained to be a circularly symmetric complex Gaussian random process (i.e. a Rayleigh or Ricean envelope). The receiver structure is ML only when the first and second order statistics of the complex Gaussian channel are known, and methods to estimate them adaptively are presented for Rayleigh fading, either blindly or with a training sequence. In chapter 8, bounds on the receiver structure's BER performance are deduced for fast Rayleigh fading, frequency-selective channels. Using these bounds and simulation, the receiver's sensitivity to various transmitter and channel parameters is studied.

In earlier research, this author could not devise a viable method for estimating the channel's second order statistics every symbol period, whichever sequence was transmitted. Accordingly, a "sequence-averaged" receiver structure was conceived, where "sequence-averaged second order statistics" are computed each symbol period. Assuming the sequence-averaged second-order statistics are known to the sequence-averaged receiver structure and the second-order statistics are known to the previous MLSE receiver structure, then the sequence-averaged receiver structure is substantially more complex and its BER performance is worse. Computing and using the sequence-averaged second-order statistics is also highly complicated, so the sequence-averaged receiver structure has little importance. It is presented in Appendix C.
The final chapter summarises the achievements presented in the thesis and lists new directions that the research could take.

1.5 LIMITATIONS OF THE THESIS

The thesis considers linear modulations only. With some work, the analytic approaches can be extended to non-linear modulations, including the practically important class, continuous phase modulation (CPM). Although consideration is given to the choice of signal constellations, code design and performance are not addressed. However, extending the receiver structures to coded sequences is straightforward. Transmitter diversity is not considered. It is assumed that feedback from receiver to transmitter on the channel state is too slow to be useful.

The linear $tf$-channel and its special cases are considered, but non-linearities within the channel are not. Bulk path loss due to an inverse power law or shadowing is not considered. In many cases, only channels with Rayleigh envelopes are addressed. Non-stationary channels are not thoroughly dealt with. The arriving rays (multipath) are assumed to be distributed evenly around the receiver in the two horizontal dimensions; the gain of the receiver's antenna is assumed to be uniform in the two horizontal dimensions also. The noise is assumed to be a circularly symmetric complex Gaussian random process, and often it is assumed white also. CCI is not explicitly considered, although many independent CCI sources with approximately equal power can be modelled as Gaussian noise.

Receiver complexity is not constrained unless it grows infinitely large with the sequence length.

1.6 LITERATURE REVIEW

This thesis is based on research in four main areas: channel characterisation, MLSE receiver structures for known channels, MLSE receiver structures for
unknown, complex Gaussian fading channels, and Per-Survivor-Processing (PSP). The historical progression of these four areas are described in turn.

1.6.1 Channel Characterisation

The largest, most general and most fundamental work on characterising linear channels is by Bello [2]. The linear channel has two degrees of freedom. The usual system function employs delay and time-variation as the two degrees of freedom. However, Bello’s paper details a wide range of system functions. Delay is the dual of frequency-selectivity, and time-variation is the dual of Doppler spreading. In the same vein, a wide range of autocorrelation functions of the channel are developed. Amongst others, the special case of Wide-Sense Stationary, Uncorrelated Scattering (WSSUS) is developed. Power series expansion of the system functions are outlined. This thesis assumes a working knowledge of Bello’s Input-Delay-Spread Function, Time-Variant Transfer Function, the WSSUS channel model, and the $f$- and $t$- power series. However, the strength and weakness of Bello’s paper is its generality. While it applies for all linear channels, it says nothing specifically of the mobile radio channel.

A statistical model of the mobile channel was developed by Clarke [13], using a scattering model. The model assumes many horizontally travelling unmodulated rays arriving at the moving receiver with uniformly distributed arrival angles and carrier phases. Assuming sufficient rays, the Central Limit Theorem applies and the received signal has a complex Gaussian pdf. The Doppler autocorrelation and Doppler power spectrum are calculated, and are the standard models in the literature [36, 32]. Jakes’ work is also extensive and goes beyond Clarke’s simpler models [25]. Another popular channel model is an exponentially decaying Doppler autocorrelation [29, 14, 64], although it has challenging properties.

The delay spread is grossly characterised by the length of the delay spread, and more precisely by its profile. Since the delay profile of a mobile channel depends intimately on the terrain, it can vary widely. Standard models used in the literature are a two-tap channel [36, 32, 38], a three-tap channel [49], a uniform, $L$-tap delay profile [11, 26], and an exponential decay model [14, 13, 64].
In practice, the choice of model only influences the computed BER performance, not the design or analysis of the receiver structure. The Clarke autocorrelation model and the uniform \( L \)-tap delay profile are used within the thesis.

1.6.2 MLSE for Known Channels

The design of receiver structures for known, time-invariant channels corrupted by Gaussian noise is one of the oldest fields in digital communications. Nyquist’s work [42] described suitable pulse shapes for AWGN channels to avoid ISI. Matched filtering for dispersive channels developed as estimator-correlator structures. The output of the estimator-correlator structures provides a set of sufficient statistics that can be manipulated into \textit{a posteriori} probabilities, the probability that a sequence was sent given the received signal. Within the introduction of [28], Kailath showed how these structures could be trivially extended to the case of the \( tf \)-channel with perfect CSI.

However, Kailath and his predecessors were satisfied with calculating the \textit{a posteriori} probabilities, that could then be used in an exhaustive comparison sub-system to determine the most probable transmitted sequence. The structure of this comparison sub-system was not defined, but it was presumably brute force. Since the complexity of brute comparison increases exponentially with the length of the sequence, their receivers were impractical.

Forney’s seminal MLSE paper, [22], employed the Viterbi algorithm to implement the exhaustive comparison with finite complexity. Symbol-rate sampled matched filters provide a set of sufficient statistics, but colour the additive noise. A whitening filter removes this constraint, so that the Viterbi algorithm can be validly employed. As a metric, the algorithm is fed with an Euclidean distance between the whitened matched filter output and the hypothesised signals. The MLSE receiver structure maximises the probability that the received signal was indeed received, given a transmitted sequence. For equiprobable symbols, this is identical to the \textit{a posteriori} probabilities of Kailath.

Soon afterwards, Ungerboeck developed an equivalent solution to the same problem [56], the extended MLSE receiver structure. His receiver structure also
employs the Viterbi algorithm, but avoids the whitening filter, and the metrics are simpler to compute. Furthermore, Forney's development employs relatively unusual mathematical tools, such as the chip transform, whereas Ungerboeck's derivation is straightforward. Accordingly, it is Ungerboeck's derivation that is generalised in this thesis to the known \(tf\)-channel, and a thorough knowledge of his paper is assumed. This generalisation has also been studied independently in [4]. However, coloured noise was not considered, nor was the receiver's BER analysed.

1.6.3 MLSE for Unknown Time-Varying Channels

An MLSE receiver structure requires a conditional probability expression before path metrics can be defined. In their seminal paper, [37], Lodge and Moher realised that the Rayleigh fading model of time-selective channels (\(t\)-channels) implies a complex Gaussian channel. Thus a conditional pdf can easily be written for the complex Gaussian channel and noise, and an MLSE receiver structure can be designed. Path metrics are the sum of Euclidean distances between the received and predicted signals. By only addressing constant envelope CPM signals, the multiplicative channel can be decoupled from the transmitted signal without loss of performance, and predicted using data-independent tap weights. The idea that time-varying channels can be predicted, not merely tracked, is of great importance.

A literature review reveals that Kam [29, 30] had already produced a related receiver structure, for slow fading with one sample per symbol. The paper's minimum mean square channel estimates are predictions, although Kam does not note this point.

Dam devised a receiver structure for PSK constellations, working in part from Lodge and Moher's ML ideas [15, 16]. Lodge and Moher's receiver structure assumes the channel's second order statistics are available: after a short training sequence, Dam's receiver can estimate them in an adaptive, decision-directed manner. An analysis of the receiver's BER performance is made, and the fast fading \(t\)-channel's implicit diversity is revealed. However, the scope of the thesis is limited in several respects. Only "textbook" PSK - rectangular, non-overlapping pulses -, white noise, and frequency-flat channels are considered. The metric only exploits the
channel’s correlation over two symbol periods. For predicting $r$ samples per symbol, the receiver employs predictors of order $r, r+1,\ldots, 2r-1$. In an unpublished letter [17], fixed-length predictors are used directly.

Vitetta and Taylor [60, 61, 62] continued this research for the $t$-channel, seeking to decouple the fading process from the transmitted signal, and then predict the fading process. They identified that the predictor tap weights are data-dependent for other than rectangular pulse shapes, but neglected to account for the data-dependent MMSE in the path metrics. Employing sequence-averaged predictor tap weights and neglecting the MMSE’s data-dependence does not lead to a premature BER floor, only a power penalty. Simulations verify that the power penalty due to the sequence-averaged predictors is small when the sampling phase is suitably controlled. The predictor tap weights are calculated from the channel’s second order statistics. They assumed a bank of pre-calculated predictor tap weights, for all the channel conditions.

Another seminal contribution came from Yu and Pasupathy [64]. Instead of decoupling the channel from the data sequence, the receiver structure they derived predicts the complete received signal directly. Individual predictors are required for every possible sequence, but the receiver structure is ML. The second order statistics are assumed to be known. In common with all the literature in this area except for [15], no methods for estimating the second order statistics are presented, and the receiver’s BER performance is evaluated through simulation, not analysis.

1.6.4 Per-Survivor Processing

Per-Survivor Processing is an obvious principle that has been overlooked in the past. Raheli et al [48] formalise it, name it, and offer applications for it. In many cases, data-dependent quantities used in the branch metrics must be estimated by the receiver. Non-PSP algorithms ordinarily delay estimating the quantity until at least the data has been tentatively detected, and only one estimate of the quantity is maintained. The PSP approach maintains many copies of the quantity, one for each hypothesised data sequence or survivor. The copy corresponding to the transmitted sequence is the correct version.
INTRODUCTION

PSP has been successfully applied to the time-invariant channel by Chugg and Polydorus [10]. It has been found to be the appropriate tool for estimating the channel's second-order statistics from the received signal.

1.7 THESIS CONTRIBUTIONS

In summary, this thesis contains several contributions towards advancing the state of the art in receiver design for time- and frequency-selective channels.

First, Ungerboeck's receiver design is extended to known time- and frequency-selective channels. The intuitively satisfying properties of a receiver for the frequency-flat channel are presented. Second, Forney's union bound method for analysing the performance of MLSE receiver structures is applied to the known time- and frequency-selective channel with diversity [66,67,68,69].

Third, Yu and Pasupathy's predictor receiver is extended to perform joint ML synchronisation, equalisation and detection for multiple, correlated diversity threads, given the signal's second order statistics. Two methods for estimating the signal's second order statistics are devised. Fourth, Forney's union bound method for analysing the performance of MLSE receiver structures is generalised to the unknown time- and frequency-selective channel. The problem of catastrophic cycle slips is accommodated within the analysis, and the transmitter is redesigned to reduce the number of bit errors during an error event [70].
2. THE TRANSMITTER

2.1 INTRODUCTION

In this chapter and in the two subsequent chapters, the three elements of a mobile communication system are described: the transmitter, the channel, and the receiver. These three chapters provide introductory material, with the goal of defining a coherent system of notation, a common signal model, an appreciation of the problems in mobile communication, and the motivation for more advanced receiver structures. The overview begins with the transmitter.

The transmitter modulates a sequence of binary information to a form suitable for transmission. The original information may be digitised speech, video, or data, which is source coded to remove redundancy and then despatched to the transmitter. Given that the source coding is efficient, the ones and zeros are equiprobable. The binary information sequence is blocked together into a stream of $M$-ary symbols, \{\alpha\}, each taken from 0, ..., $M$-1. The transmitter first channel encodes \{\alpha\}, in order to protect the information during transit; second it modulates the coded information onto a sinusoidal carrier; and third it translates the modulated signal to the allocated bandwidth.

2.2 MODULATION

Since spectrum regulatory authorities divide the spectrum up by bandwidth (and space too, using national and continental boundaries), the $M$-ary sequence, \{\alpha\}, is ultimately coded, modulated and translated to the carrier frequency, as

$$a_c(t) = \Re\{\tilde{a}(t)\exp(j2\pi f_c(t)t)\}$$

(2.1)

where the amplitude and phase, \tilde{a}(t), and/or the frequency, $f_c(t)$ depend on \{\alpha\}. There are several significant modulation forms, each solving different problems. Fast frequency hopping switches the carrier frequency several times per symbol period.
through $f_c(t)$, to spread the information content across a wider bandwidth. Fading channels exhibit intermittent notches in frequency, so fast frequency-hopping reduces the probability that a symbol is lost in a frequency notch. With many receivers hopping according to orthogonal codes, multiple users can be supported without a hard capacity limit. Code division also permits multiple users in the same bandwidth without fixing their maximum number. This is achieved by multiplying the baseband data signal by orthogonal spreading codes, forming a wideband $\tilde{a}(t)$. CPM is another class of practically important modulations, where $f_c(t) = f_c$ and $|\tilde{a}(t)| = 1$. Since the transmitted signal has a constant envelope, there is no spectral spreading when a non-linear Class C amplifier is used. The Class C amplifier is the most power-efficient. However, by constraining the signal's phase trajectory to traverse a cylinder in time, CPM signals are not as bandwidth efficient as linear modulations, particularly for large constellations.

This thesis only considers linear modulations, since they can produce the most spectrally efficient signals. The carrier is unmodulated, $f_c(t) = f_c$, and

$$\tilde{a}(t) = \sum \beta_i h(t - iT)$$

(2.2)

where $\{\beta\}$ is a sequence of symbol-spaced complex phasors encoding the data, $\{\alpha\}$; and $h(t)$ is the transmitter pulse shape. $\tilde{a}(t)$ can be visualised intuitively as occupying three dimensions: time, real amplitude, and imaginary amplitude. A sequence of pulses, $h(t)$, is spaced along the time axis every $T$ seconds. Each pulse is individually scaled and rotated around the remaining two dimensions by $\{\beta\}$. $\tilde{a}(t)$ is the vector sum of these overlapping pulses. Linear modulations are labelled linear since the modulation of complex phasor sequences obeys superposition.

This thesis employs a mixture of sampled time and continuous time processing. In sampled time, samples are taken every $T_r$ sec, where $T_r = T/r$, with $T$ being the symbol period and $r$ a positive integer. Thus there are $r$ samples per symbol period. Sampled time is denoted by a subscript, $x_i = x(iT_r)$, and an increment in a time subscript always denotes an increment in time of $T_r$. It is for this reason that the symbol-spaced phasor sequence's subscript is $ir$, so that increments in $i$ increment the subscript of $\beta_{ir}$ by $iT_r = T$. 


2.3 CONSTELLATIONS

The complex phasors, \( \{ \beta \} \), are taken from a discrete alphabet, having \( M \) elements. A plot of the alphabet in the complex plane is the signal constellation. Linear modulations include three important constellations: \( M \)-PSK, where \( \beta_m = \exp(j2\pi \frac{m}{M}) \), \( m \in \{0, \ldots, M-1\} \); \( M \)-ASK, where \( \beta_m = m \), \( m \in \{0, \ldots, M-1\} \); and \( M \)-QAM, where \( \beta_m = \left(2m_1 - \sqrt{M} + 1\right) + j\left(2m_2 - \sqrt{M} + 1\right) \), \( m_1, m_2 \in \{0, \ldots, \sqrt{M} - 1\} \).

In the fading channel, the rotational symmetry of the constellation is relevant. Define \( P \) as the number of rotationally symmetric constellation sectors, so that there are \( A = M/P \) points per sector. For \( M \)-PSK, \( P = M \) and \( A = 1 \); for \( M \)-ASK, \( P = 1 \), \( A = M \); for \( M \)-QAM, \( P = 4 \) and \( A = M/4 \).

This thesis also pursues a novel, radially symmetric constellation, comprising \( A \) shells with geometrically decreasing radii, and \( P \) points per shell, so that \( \beta_m = r^m \exp(j2\pi \frac{m}{P}) \), \( m_1 = 0 \ldots A-1, m_2 = 0 \ldots P-1; 0 < r < 1 \). An example is shown in figure 2.1.

2.4 PULSE SHAPES

The pulse shape, \( h(t) \), affects the ISI in the received signal and the transmitted signal's spectrum. Several pulse shapes are important. A rectangular pulse is used in "textbook" PSK, and is defined as

\[
h(t) = \begin{cases} 
1 & 0 \leq t \leq T \\
0 & \text{otherwise} 
\end{cases}
\] (2.3)
Research in [15] and [60] requires a "multi-Nyquist" pulse. For any Nyquist pulse, \( h_{\text{Nyquist}}(t) \), a multi-Nyquist pulse is the superposition of several scaled and shifted Nyquist pulses, as

\[
h(t) = \sum_{i=0}^{r-1} h_{\text{Nyquist}}(tr - iT)
\]  

(2.4)

When sampled at \( t = iT \), the pulse samples are given by

\[
h(IT) = \begin{cases} 
1 & 0 \leq l < r \\
0 & \text{otherwise}
\end{cases}
\]

(2.5)

so there are \( r \) ISI-free sampling points. The bandwidth of \( h(t) \) is \( r \) times the bandwidth of \( h_{\text{Nyquist}}(t) \), although the signal power at the band edges diminishes with increasing \( r \). In the limit as \( r \to \infty \), the multi-Nyquist pulses converge on \( \text{rect}(t/T-1/2) \). The one-sided pulse bandwidth of \( h(t) \) is \( r/2T \) Hz.

These pulses have relatively wide bandwidth. In communications, bandwidth efficient pulses such as the raised cosine pulse,

\[
h(t) = \frac{\sin(\pi t/T) \cos(\chi \pi t/T)}{\pi t/T - 1 - 4\chi^2 t^2 / T^2}
\]

(2.6)

and the square root raised cosine pulse,

\[
h(t) = \frac{1}{T} \text{sinc} \left( \frac{t}{T} \right) + \frac{1}{T} \text{sinc} \left( \frac{t}{T} + \frac{1}{2} \right) \cos \left( \frac{\pi t}{T} + \frac{1}{2} \right) + \frac{1}{T} \text{sinc} \left( \frac{t}{T} - \frac{1}{2} \right) \cos \left( \frac{\pi t}{T} - \frac{1}{2} \right)
\]

(2.7)

have more practical importance. \( \chi \) is the excess bandwidth, \( 0 \leq \chi \leq 1 \), so the one sided bandwidth of both pulses is \((1+\chi)/2T\). The raised cosine pulse decays more swiftly in time than the root raised cosine pulse.

Since an infinite pulse is not implementable, all pulses in this thesis are truncated to be non-zero only over \([0; HT]\), where \( H \) is the pulse length in symbol periods. The raised cosine and square root raised cosine pulses are translated by \( HT/2 \) sec before truncation, to preserve their main lobes.
2.5 CODING

The mapping from the source coded sequence \(\{a\}\) to constellation points, \(\{b\}\) is the subject of coding and modulation. In some sense, coding is the foundation of communications, and coding cannot be neglected in the study of transmitters.

Coding for the AWGN channel is a mature area, particularly block and convolutional codes. Trellis codes have been discovered recently [56], where coding is performed jointly with modulation.

Shannon's channel capacity theorem [51] proves that coding schemes exist to transmit information with an arbitrary low probability of error, assuming the information rate of the source does not exceed the channel capacity. However, the capacity theorem does not constrain the coding delay. In the AWGN channel, errors are induced by the independent noise, and so a substantial BER or SNR improvement can be achieved without significantly increasing the transmission delay.

In the fading channel, the delay problem is much more severe. In effect, the channel capacity is varying in time, according to the depth of the fade. During a deep fade, the channel capacity is very low, but the rate of the information source remains the same. If channel state information (CSI) were available at the transmitter, the information could be buffered, then coding could be used to vary adaptively the transmitted information rate. However, in this thesis, we assume that timely CSI cannot be sent from the receiver to the transmitter.

Accordingly, a successful code must smear the information well beyond the fade duration. In general, this leads to long and complicated codes, with complicated decoding subsystems.

Complexity is normally reduced through a two level code. The outer "code" is an interleaver which distributes the data in time. The inner code may be a powerful, high diversity trellis code [19]. The outer code can be decoded easily with a deinterleaver, while ensuring the powerful inner code is spread far enough to perform properly. The main decoding effort revolves around the inner code. In each case, the information must be spread beyond the duration of the fade, and delay is inevitable.
For instance, even the simplest two-state interleaved trellis code in the frequency-flat, Rayleigh fading channel requires an interleaving delay of at least $t_d$ sec for optimal performance, according to [57]

$$t_d > 1.53/f_D$$

(2.8)

where $f_D = v f_c / c$ is the one-sided Doppler spread; $v$ is the receiver speed; $f_c$ is the carrier frequency; and $c$ is the speed of light. For a receiver operating at 900MHz, travelling at walking pace (4km/h), the Doppler spread is 3.3Hz and the interleaving delay is 0.46 sec. Furthermore, the code's performance is significantly impaired by reducing the interleaving. In channels with significant frequency-selectivity, there is implicit delay diversity, the error rate is less dominated by long, deep fades, and so codes can enhance the system performance. The codes still require the interleaving of equation (2.8) for optimal performance. However, a robust communication system cannot rely on Doppler or delay spread, so this author argues that most coding will be insufficient for future systems that offer ATM-like services: i.e. a common digital transmission system providing high bandwidth, a low error rate, and a low delay.

Coding is still important however. There are many services where delay is irrelevant or permissible, such as broadcast and file transfer. There are many channel models where coding can improve the SNR for a given BER performance, or vice versa. When retransmission is possible, coding can be used to detect errors. The low cost of digital processing means that any improvement due to coding is usually worthwhile.

This thesis largely ignores coding. The MLSE receiver structures and analytic techniques can be extended directly to exploit codes; however this research has not been undertaken here.

For the case of the known channel, the data sequence, $\{x\}$, is mapped directly to the complex phasors, $\{\beta\}$, through a Gray code.
2 THE TRANSMITTER

For the case of an unknown channel in the absence of pilot symbols, pilot tones, or a deterministic component from the transmitted signal, the receiver has no absolute phase reference and its amplitude reference can be unreliable. Accordingly, the transmitter constellation and mapping should be designed with rotational-invariance and amplitude-slip-tolerance. Suitable constellations include an $M$-PSK constellation and the radial constellation of figure 2.1.

The mapping from $\{\alpha\}$ to $\{\beta\}$ has two stages. The $M$-ary data sequence, $\{\alpha\}$, is divided into two fields, to code the constellation's $\log_2 P$ phase bits and $\log_2 A$ sector (amplitude) bits. First, the $\log_2 P$ phase bits select a sector in a rotationally invariant manner, such as differential encoding. Second, the $\log_2 A$ sector bits select a point from that sector, in such a way that the effect of an amplitude slip is diminished. $M$-PSK has one point per sector, so amplitude-slips do not arise. For $M$-QAM constellations, an effective solution is unclear. By comparison, amplitude-slip tolerance can be achieved easily for the proposed radially symmetric constellation. The sector bits Gray-encode the transitions between shells, not the absolute shell. They define the number of shells to increase, with wrap-around if the limit of the constellation is reached.

2.6 SUMMARY

As a summary, a block diagram of the transmitter is shown in figure 2.2.
3 THE CHANNEL

3.1 INTRODUCTION

In this chapter, the properties of time- and frequency-selective multipath channels are explained. A deterministic channel model is initially developed, and two discretised versions of this are deduced. In practice, a statistical model of the multipath channel is accurate and useful, so this is developed too. The properties of special channels are considered: the time-invariant, frequency-selective channel (f-channel); the time-varying, frequency-flat channel (t-channel); and the non-fading, time-invariant, frequency-flat channel (AWGN channel). The sources and properties of noise are described.

The properties of multipath are seen most easily when the received signal (the channel output) is near baseband. Accordingly, the receiver’s translation of the real received signal to a complex signal near baseband is lumped with the channel. This area is described first.

3.2 RECEIVER TRANSLATION

The received signal is centred at a high carrier frequency, yet its bandwidth is relatively small. Therefore there is no need to process the received signal at the Nyquist rate of the bandpass signal, since it can be translated to a lower carrier frequency without information loss. This translation is undertaken by the receiver’s RF and intermediate frequency (IF) sections. The RF section at least is usually analogue. Normally the frequencies of the RF and IF local oscillators (LOs) are tightly controlled in feedback loops, to remove the carrier from the received signal. In this thesis, schemes are sought for estimating the carrier frequency, so free-running oscillators at approximately the carrier frequency are sufficient for the receiver’s RF and IF sections. This eases the RF hardware requirements, but now the IF stage’s output has a residual carrier offset.
In figure 3.1(a), the RF and IF sections are shown as blocks. The output is analogue, so it suits an analogue receiver. Often a $T_s$-sampled version of the received signal is needed by the receiver, and a suitable RF/IF chain is shown in figure 3.1(b). There are many other equivalent hardware arrangements, but the pictured RF/IF chain requires only one low-speed A-D converter and performs digital quadrature demodulation.

The digital receiver translates the bandpass received signal, centred at $f_c$, to a complex signal near baseband in two stages. Since the centre frequency and signal bandwidth may be uncertain, a wide front-end filter is used. It is centred at the carrier’s estimated value, $f_{c,nom} - f_c$, with a two-sided bandwidth of $1/T$. The number of samples per symbol, $r$, is chosen to be large enough that $1/T_s$ passes the signal without distortion. In most instances, the signal is not completely bandlimited, so a distortionless zonal filter cannot be properly defined. However, normally another bandwidth definition (such as the $-40$dB bandwidth) can be used such that the signal distortion is negligible, and sample accordingly. A large error in $f_{c,nom}$ is accommodated by a large $r$. The filtered signal is translated to a low IF frequency,
A free-running LO operating at approximately \( f_{\text{nom}} \cdot 1/2T_r \) suffices. In this way, the A-D converter observes a real signal somewhere within \( f = [0; 1/T_r] \). This signal is sampled at \( 2/T_r \) so that no aliasing arises. The free-running digital oscillator \( \exp(-j2\pi(1/2T_r)t), t = 1T_r/2 \) generates images centred nominally at 0Hz and at \( 1/T_r \). A low pass filter with bandwidth \( 1/2T_r \) retains only the signal around 0Hz. The sampling rate is twice times this signal’s Nyquist rate, so it is 1:2 sub-sampled without loss of information.

### 3.3 PHYSICS OF THE MULTIPATH CHANNEL

In general, mobile receivers exist in a multipath environment. The environment is cluttered with hills, buildings, vehicles and trees. Signals radiated from an antenna can be represented validly as rays, since wavelengths are small compared to the physical structures at RF. These rays reflect off the obstacles (scatterers) in all directions, and some arrive in the vicinity of the receiving antenna. Each ray is delayed and attenuated differently along its path. These modulated travelling waves superimpose and create a standing wave interference pattern. This is a deterministic system, and a precise model of the environment can predict the interference pattern.

Following [13], we consider horizontally travelling rays only. The \( i \)th ray is reflected or refracted by a series of scatterers, indexed by \( k \). Each scatterer attenuates the ray by a real ratio, \( \rho_{ik}, 0 \leq \rho_{ik} \leq 1 \). The attenuation of the \( i \)th ray due to all scatterers is called the path attenuation,

\[
\rho_i = \prod_k \rho_{ik}
\]  

(3.1)

Rays experience different path delays, \( \tau_i \), according to the distance they traverse between transmitter and receiver. The \( i \)th travelling ray, \( a_{\text{tr}}(t, x_1, x_2) \) in the vicinity of the receiver is given by

\[
a_{\text{tr}}(t, x_1, x_2) = \rho_i \Re \left\{ \exp \left( j2\pi f \left( t - \left( \tau_i + \frac{1}{c} \left( x_1 \cos \theta_i + x_2 \sin \theta_i \right) \right) \right) \right\} \right\}
\]  

(3.2)
where $\theta_i$ is the arrival angle of the $i$th path, and is to be assumed constant in the receiver’s vicinity; $c$ is the speed of light; $x_1$ and $x_2$ are the two horizontal directions; and without loss of generality, the receiver’s position is defined as $(x_1, x_2) = (0, 0)$ at time, $t = 0$.

The interfering rays superimpose in the vicinity of the receiving antenna,

$$y_c(t, x_1, x_2) = \sum_i a_{c,i}(t, x_1, x_2)$$

(3.3)

Neglecting noise, this real, bandpass signal is shifted down to a complex signal at approximately baseband by the receiver of figure 3.1, as

$$y(t, x_1, x_2) = y_c(t, x_1, x_2) \exp(-j2\pi f_{c,\text{nom}} t + j\phi)$$

(3.4)

where $f_{c,\text{nom}}$ is the receiver’s best a priori estimate of the carrier frequency; and $\phi$ is the phase offset between transmitter and receiver oscillators. There is also a residual carrier offset, $f_{0,\text{nom}} = f_c - f_{c,\text{nom}}$. After low-pass filtering, this near baseband signal in the vicinity of the receiver’s antenna equals

$$y(t, x_1, x_2) = \sum_i \frac{1}{2} p_i \tilde{a}(t - \left(\tau_i + \frac{1}{c}(x_1 \cos \theta_i + x_2 \sin \theta_i)\right)) \exp\left(j2\pi f_{0,\text{nom}} t + j\phi\right) \times
\exp\left(-j2\pi f_c \left(\tau_i + \frac{1}{c}(x_1 \cos \theta_i + x_2 \sin \theta_i)\right)\right)$$

(3.5)

The receiver antenna is not stationary in mobile radio. It can be modelled reasonably accurately as having a constant speed, $v$, at angle, $\phi$, in the horizontal plane, so the receiver’s position as a function of time is $(x_1, x_2) = (vt \cos \phi, vt \sin \phi)$.

The multipath interference pattern is a function of position, and so appears to change as the receiver moves. The more rapidly the receiver moves, the faster the changes occur. The time variable effectively parameterises the receiver’s path. Thus the translated signal at the receiver’s antenna is given by

$$y(t) = \sum_i \frac{1}{2} p_i \tilde{a}\left(t - \left(\tau_i + \frac{v}{c}\cos(\phi - 0, \theta_i)\right)\right) \exp\left(j2\pi f_{0,\text{nom}} t + j\phi\right) \exp\left(-j2\pi f_c \left(\tau_i + \frac{v}{c}\cos(\phi - \theta_i)\right)\right)$$

(3.6)

Observing that the receiver speed is substantially less than the speed of light, the narrowband transmitted signal, $\tilde{a}(\cdot)$, has virtually the same value at $t - \left(\tau_i + \frac{v}{c}\cos(\phi - 0, \theta_i)\right)$ as at $t - \tau_i$. Accordingly, equation (3.6) can be written as
The channel, \( z(t, \xi) \), is named the Input Delay-Spread Function by Bello. His paper, [2], defines many transforms of this function, of which only the Time-Variant Transfer Function, \( T(f_d, t) \), is used within this thesis. It is defined as

\[
y(t) = \sum \frac{1}{2} \rho_i \bar{a}(t - \tau_i) \exp(j2\pi f_{0,non} t + j\phi) \exp(-j2\pi f_c (\tau_i + \frac{\nu}{c} \cos(\varphi - \theta_i))) = \sum \bar{a}(t - \tau_i) \exp(j2\pi f_{0,non} t (1 - \tau_i) + j\phi) \frac{1}{2} \rho_i \exp(j2\pi f_{0,non} \tau_i) \exp(-j2\pi f_c (\tau_i + \frac{\nu}{c} \cos(\varphi - \theta_i)))
\]

(3.7)

The last line distinguishes between the transmitted signal and the effect of multipath interference. The multipath interference is experienced by the receiver's moving antenna as a time-varying quantity which is labelled the channel. Defining the Doppler frequency as

\[
f_D = \frac{vf_c}{c}
\]

(3.8)

the channel can be written as

\[
z(t, \xi) = \sum \frac{1}{2} \rho_i \exp(j2\pi f_{0,non} \xi) \exp(-j2\pi f_c (\xi + \frac{\nu}{c} \cos(\varphi - \theta_i))) \delta(\xi - \tau_i) = \sum \frac{1}{2} \rho_i \exp(-j2\pi f_{0,non} \xi) \exp(-j2\pi f_c \cos(\varphi - \theta_i)) \delta(\xi - \tau_i)
\]

(3.9)

The complex transmitted signal, including the residual carrier offset, is defined as

\[
a(t) = \bar{a}(t) \exp(j2\pi f_{0,non} t + j\phi)
\]

(3.10)

and so the signal at the receiver's antenna is given by

\[
y(t) = \int_{-\infty}^{\infty} a(t - \xi) z(t, \xi) d\xi
\]

(3.11)

This time-varying channel model was derived assuming a two-dimensional, mobile radio channel with the transmitter and scatterers motionless. In fact, it applies generally to all time-varying linear channels [2], such as the telephone channel, the microwave LOS channel, the optical fibre channel, the submarine channel, or the three-dimensional mobile radio channel with moving terrain. Accordingly, a receiver structure designed for this channel model can in fact be used in a wide range of environments.
3 THE CHANNEL

\[ T(f, t) = \int_{-\infty}^{\infty} z(t, \xi) e^{-j2\pi f \xi} d\xi \]  

(3.12)

so that

\[ y(t) = \int_{-\infty}^{\infty} A(f) T(f, t) e^{j2\pi f t} df \]  

(3.13)

\[ z(t, \xi) \] is the response at time \( t \) to a unit impulse input at time \( t-\xi \).

\[ T(f, t)e^{j2\pi f t} \] is the response at time \( t \) to a tone at frequency, \( f \).

3.4 THE TAPPED DELAY-LINE MODEL

The integral model of equation (3.11) applies for all frequencies. It can be viewed as a densely tapped delay line, where each value of \( \xi \) indexes a tap, and the time-varying tap weight is \( z(t, \xi) \). Figure 3.2(a) is a diagram of the model. Each tap weight is a random process that multiplicatively distorts the delayed transmitted signal. In practice, the transmitted signal and channel tap processes are approximately bandlimited, so taps need not be so densely spaced.

The received signal is ultimately sampled every \( T_r \) sec, where the number of samples per symbol, \( r \), is made large enough to satisfy the signal’s Nyquist rate. In this way, sampling does not discard information if the transmission interval is very long. In practice, the received signal cannot be ideally bandlimited, since the transmitted signal employs pulse shapes truncated to \( H \) symbol periods. However, another bandwidth definition (such as the -40dB bandwidth) can be chosen, and the information loss is negligible.

Therefore, the sampled received signal can be represented without loss of information as

\[ y(\xi T_r) = \int_{-\infty}^{\infty} A(\xi T_r - \xi) z(\xi T_r, \xi) d\xi \]  

(3.14)

The received signal bandwidth is the sum of the bandwidth of the transmitted signal and of the channel tap processes. Accordingly, the transmitted signal can also be constructed from its \( T_r \)-spaced samples, as
Define the tap weights of this tapped delay line as

\[ z_{1,m} = \int_{-\infty}^{\infty} z(IT_r, \xi) \sin c(m - \xi/T_r) d\xi \]  

so that the received signal can be expressed as a discrete convolution,

\[
\sum_{n=-\infty}^{\infty} a_{1-m} z_{1,m} = \sum_{n=-\infty}^{\infty} a_{1-m} \sin c\left(m - \frac{\xi}{T_r}\right) z(IT_r, \xi) d\xi
\]

\[
= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{1-m} \sin c\left((IT_r - \xi)/T_r - m \right) z(IT_r, \xi) d\xi
\]

\[
= \int_{-\infty}^{\infty} a(IT_r - \xi) z(IT_r, \xi) d\xi = y_i
\]

This tapped delay line channel model is drawn in figure 3.2(b). Figure 3.1(b) is a diagram of a digital receiver structure that would employ this discretised channel model.

![Continuous time tapped delay line model](image1)

![Sampled time tapped delay line model](image2)
3.5 THE RECEIVED PULSE SHAPE MODEL

In this section, another channel model is described. In many ways it is a more intuitive and powerful channel model than the tapped-delay line model. Instead of isolating the transmitter pulse shape from the channel, the two are lumped together as the received pulse shape.

The received signal at the receiver’s antenna, equation (3.11), is expanded by substituting equation (2.2) for the transmitted signal, as

\[ y(t) = \int_{-\infty}^{\infty} a(t-\xi)z(t,\xi) d\xi \]

\[ = \sum_i B_i \int_{-\infty}^{\infty} h(t+iT-\xi)z(t,\xi) \exp\left(j2\pi f_{0,\text{nom}}(t-\xi) + j\phi\right) d\xi \]

\[ = \sum_i B_i c(t-iT,iT) \tag{3.18} \]

where the received pulse shape is defined as

\[ c(t-iT,iT) = \int_{-\infty}^{\infty} h(t-iT-\xi)z(t,\xi) \exp\left(j2\pi f_{0,\text{nom}}(t-\xi) + j\phi\right) d\xi \tag{3.19} \]

In discrete time, the analysis is similar,

\[ y_i = \int_{-\infty}^{\infty} a(IT_r,\xi)z(IT_r,\xi) d\xi \]

\[ = \sum_i B_i \int_{-\infty}^{\infty} h((l-i)rT_r-\xi)z(IT_r,\xi) \exp\left(j2\pi f_{0,\text{nom}}(IT_r-\xi) + j\phi\right) d\xi \]

\[ = \sum_i B_i c_{l-i,r} \tag{3.20} \]

where the discretised received pulse shape is defined as

\[ c_{l-i,r} = \int_{-\infty}^{\infty} h((l-i)rT_r-\xi)z(IT_r,\xi) \exp\left(j2\pi f_{0,\text{nom}}(IT_r-\xi) + j\phi\right) d\xi \tag{3.21} \]

If the anti-aliasing/IF filtering is too narrow, its impulse response affects the received pulse shape through a further convolution. The problem becomes complicated to analyse, since the received pulse is at least as long as the IF filter’s impulse response. Low sampling rates have been used [15, 61, 14, 64], but the influence of an IF filter has not been addressed. In effect the anti-aliasing filter is
3 THE CHANNEL

widened without increasing the sampling rate, and then the additional aliased noise is ignored.

These models are intuitive extensions of the linearly modulated $\tilde{a}(t)$. $y(t)$ can be visualised as occupying three dimensions: time, real amplitude, and imaginary amplitude. A sequence of complex received pulses, $c(t,iT)$, is spaced along the time axis every $T$ sec. Each pulse is individually scaled and rotated around the remaining two dimensions by $\{\beta\}$. $y(t)$ is the vector sum of these overlapping pulses. The only novel aspect of this signal model is that the time-varying channel distorts each transmitted pulse differently. Accordingly, the received pulses have an additional parameter, $iT$, to show that they each vary. This model is used extensively within the thesis.

3.6 COVARIANCE FUNCTIONS FOR CHARACTERISING THE CHANNEL

In practice, the path attenuations, delays and angles of arrival are not known to the transmitter or receiver. Instead it is appropriate to describe the channel statistically, by assuming that the environment contains many scatterers. Before investigating these ideas further, general properties of stochastic channels are described. (When the channel cannot be estimated, a knowledge of its second order statistics can be used to predict how the channel evolves.) Even when the channel is known, the channel's second order statistics influence a receiver's BER performance.

Define the channel autocorrelation as

$$R_{zz}(t,\xi, t + \Delta t, \xi + \Delta \xi) = \frac{1}{2} E\{z(t,\xi)\bar{z}\left(t + \Delta t, \xi + \Delta \xi\}\} \quad (3.22)$$

In many channels, including the mobile radio channel, the channel statistics are approximately stationary for a sufficiently long interval that the channel can be called Wide Sense Stationary (WSS) [13]. That is, the autocorrelation properties depend on time differences, not on an absolute time reference. Furthermore, the rays that contribute to $z(t,\xi)$ for each $\xi$ have the same delay. Therefore rays with different delays follow different paths, and so can be reasonably described as uncorrelated. This Uncorrelated Scattering (US) property and the complex Gaussian channel model
implies that the channel tap processes are independent. A Wide Sense Stationary, Uncorrelated Scattering (WSSUS) model is mathematically convenient, since

\[ R_{\infty}(t, \xi, t + \Delta t, \xi + \Delta \xi) = R_{\infty}(\Delta t)R_{\xi}(\Delta \xi) \]  

(3.23)

where \( R_{\infty}(\Delta t) \) is the autocorrelation in time of each tap, and \( R_{\xi}(\Delta \xi) \) is the delay power spectrum.

Unfortunately, the WSSUS model does not model several real channels. The microwave LOS channel has a non-faded direct wave, a slowly faded ground wave and a fast faded sky wave. The statistics of the ground wave and the sky wave are different, a property that cannot be described by the WSSUS model of equation (3.23). Similarly, the “scattering” is correlated in the time-invariant telephone channel, since

\[ R_{\infty}(t, \xi, t + \Delta t, \xi) = \frac{1}{2} E\{z(t, \xi)z(t, \xi + \Delta \xi)\} = \frac{1}{2} z(0, \xi)z(0, \xi + \Delta \xi) \]  

(3.24)

The channel can be divided into random and non-random (mean, deterministic) components. \( z(t, \xi) = (z(t, \xi) - E[z(t, \xi)]) + E[z(t, \xi)] = z'(t, \xi) + z''(t, \xi) \). Then the channel autocorrelation equals

\[ R_{\infty}(t, \xi, t + \Delta t, \xi + \Delta \xi) = \frac{1}{2} E\{z'(t, \xi)z'(t + \Delta t, \xi + \Delta \xi)\} + \frac{1}{2} z''(t, \xi)z''(t, \xi + \Delta \xi) \]  

(3.25)

The channel can be WSSUS only if the first term is WSSUS and the second is zero, or if both terms are non-zero at only one value of \( \xi \) and the second term is time-invariant. Only zero-mean \( f \)-channels and all \( t \)-channels with time-invariant deterministic components are WSSUS.

Therefore a general and accurate model with tractable mathematical properties couples a random, WSSUS component and a deterministic component. \( R_{\infty}(\Delta t) \) is then the tap or Doppler autocovariance in time, and \( R_{\xi}(\Delta \xi) \) is the delay power spectrum of the random components only.

A channel, \( z(t, \xi) \) can be characterised in a gross way by its delay and Doppler spreads. The first and last rays reach the receiver \( \xi_F \) and \( \xi_L \) sec after transmission respectively, so the (two-sided) delay spread is

\[ \tau = \xi_L - \xi_F \]  

(3.26)
An rms definition of the delay spread has more value, since it does not weight highly the hugely delayed, highly attenuated rays. The delay power spectrum is normalised

\[ p_{\xi}(\xi) = \frac{R_{\xi}(\xi)}{\int_{-\infty}^{\infty} R_{\xi}(\xi) d\xi} \]  

(3.27)

for use as a pdf to calculate the rms delay spread, \( \tau_{\text{rms}} \), as

\[ \tau_{\text{rms}}^2 = \int_{-\infty}^{\infty} p_{\xi}(\xi) \xi^2 d\xi - \left( \int_{-\infty}^{\infty} p_{\xi}(\xi) \xi d\xi \right)^2 \]  

(3.28)

assuming there are no deterministic components [2].

However, it is \( \tau \) that is important in MLSE receiver design. A large value leads to increased receiver complexity, so in practice \( \tau \) is chosen large enough to accommodate the significant rays only.

The non-dimensional ratio, \( \tau/T \), is an important measure of the channel's frequency-selectivity. For \( \tau/T = 0 \), the channel is frequency-flat. For \( \tau/T \ll 1 \), the channel can be accurately modelled as linearly frequency-selective. Longer delay spreads must be specifically accommodated by the receiver, through equalisation for instance.

The Fourier transform of the delay power spectrum is the channel's autocovariance in frequency,

\[ S_\nu(\Delta f) = \int_{-\infty}^{\infty} R_{\xi}(\xi) \exp(-j2\pi\nu\xi) d\xi \]  

(3.29)

The channel correlation generally decays with increasing frequency separation. The Fourier transform of the tap autocovariance, \( R_\tau(\Delta t) \), is the Doppler spectrum,

\[ S_\nu(\nu) = \int_{-\infty}^{\infty} R_\tau(\Delta t) \exp(-j2\pi\nu\Delta t) d\Delta t \]  

(3.30)

From physical considerations, the Doppler spectrum is only non-zero for low frequencies. In a motionless multipath channel, the time variation is due solely to the receiver's velocity. The maximum (one-sided) Doppler spread is then \( f_D \). In practice, the actual maximum Doppler spread is somewhat higher, since the environment is not
motionless: there may be trees rustling in the wind, moving vehicles, or walking pedestrians.

The fractional Doppler, $f_D T$, is an important dimensionless product that characterises the degree of time-selectivity. When $f_D T = 0$, the channel is time-invariant. For other values of $f_D T$, the fading is described imprecisely as "slow" or "fast." Slow fading is usually taken to assume that the fading process is as good as constant over a symbol duration, which is largely satisfied for $f_D T < 0.01$. In fast fading channels, the fading process cannot be assumed constant over a symbol period. $f_D T > 0.1$ is generally accepted as fast fading.

Ultimately, the distinction between slow and fast is made from BER considerations. If a receiver is designed under the slow fading approximation, but has a BER floor within typical SNR values, then the fading is in fact fast [8].

3.7 A STATISTICAL DESCRIPTION OF THE MULTIPATH CHANNEL

When the number of paths between transmitter and receiver is large, a statistical description of the channel is appropriate. Then $\rho_k$, $\theta_i$ and $\tau_i$ are random variables, so that $\rho_i$ and $z(t, \xi)$ are also.

When there are many independent scatterers along a single path, the Central Limit Theorem (CLT) applies to $\log_{10} \rho_i = \sum_k \log_{10} \rho_k$. Thus $\log_{10} \rho_i$ is a real, Gaussian random variable and the path attenuation, $\rho_i$, is log-normally distributed.

Clarke considered the case of narrowband fading, with equiprobable arrival angles and carrier phases due to slight differences in path lengths [13]. A complex path attenuation with random phase was constructed, by lumping the real path attenuation and complex carrier phase together, leading to a complex Gaussian channel model. However, in this thesis the path lengths are already explicitly identified, so the carrier phase is the same for all paths at a given delay. This can be seen in equation (3.9) for $t = 0$, where

$$
z(0, \xi) = \frac{1}{2} \exp\left(-j2\pi f_{\text{carrier}} \xi\right) \sum \rho_i \delta(\xi - \tau_i)
$$

(3.31)
$z(0,\xi)$ is only a complex scalar representation of a real Gaussian random process in $\xi$; it is not a complex Gaussian random process. As we shall see, this difficulty disappears when the signal's narrow bandwidth is considered.

The tapped delay line model exploits the signal's narrow bandwidth, and moreover describes everything about the channel that is observable to the receiver. Accordingly, the discretised channel of equation (3.16) is expanded using equation (3.9) as

$$z_{l,m} = \sum_i \frac{1}{2} \rho_i \sin c(\frac{m - \tau_i}{T_r}) \exp(-j2\pi(f_{c,non}\tau_i + l\beta \cos(\varphi - 0)))$$  \hspace{1cm} (3.32)

If all the rays are scattered completely dependently, or if there is only one path, then $z_{l,m}$ has a log-normal distribution. If there are sufficient independent paths for the CLT to apply, then each sample of each tap is a circularly symmetric, complex Gaussian random variable. For $l = 0$,

$$z_{0,m} = \sum_i \frac{1}{2} \rho_i \sin c(\frac{m - \tau_i}{T_r}) \exp(-j2\pi f_{c,non}\tau_i)$$  \hspace{1cm} (3.33)

Unlike equation (3.31), this sum is a circularly symmetric complex Gaussian random process in $\xi$, since it is the sum of independent complex random variables, weighted by the $\sin c(.)$ function.

The channel is also modelled by other pdfs. The important density functions are the Suzuki pdf, the mixed Rice and lognormal distribution [54], and the Nakagami distribution [40]. These are broadly similar to a complex Gaussian distribution with non-zero mean. In this thesis only the circularly symmetric complex Gaussian distribution is pursued, since it accurately models the multipath channel, it can be justified theoretically, and it has a mathematically tractable multivariate distribution.

If all the paths have similar but independent attenuations at random delays, then the sum in equation (3.32) is a zero-mean, circularly symmetric complex Gaussian random variable. If all the paths have similar but independent attenuations at random delays, except for one dominant path (the line-of-sight path), then the sum in equation (3.32) is a circularly symmetric complex Gaussian random variable with a non-zero mean.
Labelling the line of sight path as the 0th path, the mean of the channels taps,
\[ E(z_{j,m}) = z_{i,m}^0 \], equals
\[ z_{i,m}^0 = \frac{1}{2} \rho \sin c(m - \tau_0 / T_r) \exp \left(-j2\pi \left( f_{c,\text{nom}} \tau_0 + f_r T_r \cos(\phi - \theta_i) \right)\right) \] (3.34)

Thus the deterministic component in a mobile radio channel is only time-invariant when the Doppler spread is zero (the receiver is immobile) or the LOS ray arrives at right angles to the receiver’s motion. The channel tap autocovariance,
\[ \frac{1}{2} E\left(\left(z_{i,m} - E(z_{i,m})\right)\left(z_{j,n} - E(z_{j,n})\right)\right) = \frac{1}{2} E(z_{i,m}^r z_{j,n}^r) \], equals
\[ \frac{1}{2} E(z_{i,m}^r z_{j,n}^r) = \sum_{\theta > 0} \sum_{\theta < 0} \frac{1}{2} E(\rho^2 \sin c(m - \tau_i / T_r) \sin c(p - \tau_k / T_r) \exp(j2\pi f_{c,\text{nom}}(\tau_k - \tau_i))) \times E\left(\exp(j2\pi f_r T_r(n \cos(\phi - \theta_k) - l \cos(\phi - \theta_i))\right) \right) \] \] (3.35)

where the path attenuations and delays are assumed to be independent from the arrival angles. The path delays must be treated specially, since their coarse values, \( \tau_i / T_r \), describe the bulk delay properties of the channel, and must be preserved by the model. However, their fine values, \( f_{c,\text{nom}} \tau_i \), can reasonably be considered random variables. This division is equivalent to the random carrier phase assumption of [13]. Accordingly, the first expectation equals zero for \( i \neq k \), and the channel tap autocovariance simplifies to
\[ \frac{1}{2} E(z_{i,m}^r z_{j,n}^r) = \sum_{\theta > 0} \sum_{\theta < 0} \frac{1}{2} E(\rho^2 \sin c(m - \tau_i / T_r) \sin c(p - \tau_i / T_r) \exp(j2\pi f_r T_r(n - l) \cos(\phi - \theta_i)) \right) d\theta_i \]
\[ = J_0(2\pi f_r T_r(n - l)) \sum_{\theta > 0} \frac{1}{2} E(\rho^2 \sin c(m - \tau_i / T_r) \sin c(p - \tau_i / T_r)) \] (3.36)

where the arrival angles are assumed to be uniformly distributed, as
\[ p_0(\theta_i) = \begin{cases} \frac{1}{2\pi} & 0 \leq \theta_i < 2\pi \\ 0 & \text{otherwise} \end{cases} \] (3.37)

If the arrival angles are not uniformly distributed, then a Doppler shift is introduced.

Since the tapped delay line model describes everything about the channel that is observable to the receiver, the statistical properties of \( z(t, \xi) \) can be freely altered, as
long as there is no change to the tapped delay line model’s statistical properties. Real
and imaginary samples of the channel are i.i.d complex Gaussian random variables,
since the factor \( \exp(-j2\pi f_c\tau) \) distributes the contribution from each path
randomly towards the multipath sum, equation (3.32). Accordingly, the sum is
completely described by its mean and complex conjugate autocovariance,
\[ \frac{1}{L} E\left( z_{t,m}^* z_{t,n}^* \right) \]

We propose another model for \( z(t,\xi) \). Each tap, \( \xi \), has a circularly symmetric
complex Gaussian random process as its time varying tap weight, \( z(t,\xi) \). \( z(t,\xi) \)
comprises random and non-random components. The non-random component is
arbitrary, and includes
\[ z^\omega(t,\xi) = \frac{1}{2} \rho_0 \exp\left(-j2\pi f_c\tau_0 + \frac{f_c t}{c} \cos(\psi - \theta)\right) \delta(\xi - \tau_0) \]  \hspace{1cm} (3.38)
as a special case. The random component of each tap weight is independent of the
random component of every other tap weight. The autocovariance of each tap with
itself over time is identical for all taps. Then the random part of the channel is
WSSUS, with autocorrelation
\[ E\left( z'(t,\xi) z'(t + \Delta t,\xi + \Delta \xi) \right) = R_u(\Delta t) R_{z \xi}(\xi) \delta(\Delta \xi) \] \hspace{1cm} (3.39)

The autocovariance in time is arbitrary, and includes the isotropic scattering
model of [13],
\[ R_u(\Delta t) = J_u\left(2\pi f_D |\Delta t|\right) \] \hspace{1cm} (3.40)

Its U-shaped Doppler spectrum is calculated from equation (3.30) as
\[ S_u(f) = \begin{cases} \frac{1}{\pi f_D \sqrt{1 - f^2 / f_D^2}} & |f| \leq f_D \\ 0 & \text{otherwise} \end{cases} \] \hspace{1cm} (3.41)

The delay power spectrum due to the random components equals
\[ R_{z \xi}(\xi) = \sum_i \frac{1}{L} E\left( |z_i|^2 \right) \delta(\xi - \tau_i) \] \hspace{1cm} (3.42)

Using the special case of equation (3.38), the mean of the discretised \( z(t,\xi) \)
equals
Using the special case of equation (3.40), the autocovariance of the discretised $z(t, \xi)$ equals

$$\frac{1}{2} E\left( z'_{\tau_m} z'_{p} \right)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\left( z\left( (n-1)T_r, \xi_1 \right) z\left( nT_r, \xi_2 \right) \right) \sin c\left( m - \xi_1 / T_r \right) \sin c\left( p - \xi_2 / T_r \right) d\xi_1 d\xi_2$$

$$= \int_{-\infty}^{\infty} R_{n} \left( (n-1)T_r \right) R_{n} \left( \xi_1 \right) \sin c\left( m - \xi_1 / T_r \right) \sin c\left( p - \xi_1 / T_r \right) d\xi_1$$

$$= J_{o} \left( 2\pi f_{D} T_r |n-1| \right) \sum_{m=0}^{\frac{1}{2}} E\left( \rho^2 \right) \sin c\left( m - \tau / T_r \right) \sin c\left( p - \tau / T_r \right)$$

Both the mean and autocovariance of the proposed model match identically the channel’s actual mean, equation (3.34), and autocovariance, equation (3.36). The two models cannot be distinguished, but the proposed model is more general and tractable. Accordingly, the proposed model is used for $z(t, \xi)$, from now onwards.

The temporal correlation can be transformed into a spatial correlation by observing $\Delta x = v\Delta t$, according to

$$R_{\alpha} (\Delta x) = J_{o} \left( 2\pi f_{D} \frac{1}{2} |\Delta x| \right)$$

This is the multipath interference pattern’s correlation between two points separated by $\Delta x$. Since the arrival angles are assumed uniformly distributed, the orientation of the two points does not affect the correlation.

Space diversity is considered in this thesis also, so the channel’s autocorrelation across antennae and time is also required. The antenna array comprises $D$ antennae in a row, spaced every $W$ wavelengths of $f_c$ ($c/f_c = v/f_D$). The $d$th antenna traces the path

$$\left( x_1, x_2 \right) = \left( \frac{vdW \cos \theta}{f_D} + vt \cos \varphi, \frac{vdW \sin \theta}{f_D} + vt \sin \varphi \right)$$
where $\theta$ is the angle of the antenna array in the two horizontal dimensions. The random component of the channel experienced by the $d$th antenna at time $t$ and delay $\xi$ is denoted $z^d(t,\xi)$. Therefore the inter-antenna covariance equals

\[
\frac{1}{2}E\left(z^{d_1}(t_1,\xi_1)z^{d_2}(t_2,\xi_2)\right)
= \delta(\xi_1 - \xi_2)R_{\xi}(\xi_1)J_0\left(2\pi \sqrt{\left((d_2 - d_1)W \cos \theta + (t_2 - t_1)f_0 \cos \varphi\right)^2 + \left((d_2 - d_1)W \sin \theta + (t_2 - t_1)f_0 \sin \varphi\right)^2}\right)
= \delta(\xi_1 - \xi_2)R_{\xi}(\xi_1)J_0\left(2\pi \sqrt{\left(d_2 - d_1\right)^2W^2 + (t_2 - t_1)^2f_0^2\cos^2(\theta - \varphi)}\right)
\]

using the spatial autocovariance of equation (3.45).

### 3.8 COMPLEX GAUSSIAN DISTRIBUTIONS

In this section, the properties of complex Gaussian processes are described in more detail. Given an $X \times 1$ column vector, $x$, of samples from a complex Gaussian random process, $x(t)$, with mean $E(x)$, and an $X \times X$ autocovariance matrix $R_{xx} = \frac{1}{2}E\left((x - E(x))(x - E(x))^H\right)$, the multivariate pdf equals

\[
p_x(x) = \frac{1}{(2\pi)^X \det|R_{xx}|} \exp\left(-\frac{1}{2}(x - E(x))^H R_{xx}^{-1}(x - E(x))\right)
\]

In the limit as the sampling density rate is increased, the pdf becomes a probability functional,

\[
p_x(x) \sim \exp\left(-\frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x(t_1) - E(x(t_1))\right)R_{xx}^{-1}(t_1 - t_2)\left(x(t_2) - E(x(t_2))\right)dt_1dt_2\right)
\]

where the inverse kernel, $R_{xx}^{-1}(t)$, is the limiting form of a matrix inverse, and satisfies [58]

\[
R_{xx}^{-1}(t) * R_{xx}(t) = \delta(t)
\]
The envelope of a complex Gaussian random variable, $x$, is $|x| = \sqrt{x_R^2 + x_I^2}$ and the phase is $\angle x = \tan^{-1}(x_I, x_R)$. When the random variable is zero-mean, the phase is uniformly distributed,

$$p_\angle(\angle x) = \frac{1}{2\pi} \quad 0 \leq \angle x < 2\pi$$

and the envelope has a Rayleigh distribution [46],

$$p_{\text{Ray}}(|x|) = \frac{2|x|}{\sigma^2_{|x|}} \exp\left(-\frac{|x|^2}{\sigma^2_{|x|}}\right) \quad |x| \geq 0$$

where $\sigma^2_{|x|} = \mathbb{E}(|x|^2)$. The cumulative distribution function is also important, and is given by

$$F_{\text{Ray}}(|x|) = 1 - \exp\left(-\frac{|x|^2}{\sigma^2_{|x|}}\right) \quad |x| \geq 0$$

When the channel has a non-zero mean, the channel envelope has a Rice distribution [46].

Figure 3.3 depicts a complex Gaussian function, the channel amplitude, $|z(x_1, x_2, \xi)|$, as a function of position, $(x_1, x_2)$, at one delay, $\xi$. The multipath interference pattern is different for each value of delay. Any signal transmitted through the channel is scaled by it, so the signal strength changes smoothly as a function of position. At some positions, the signal is deeply faded.

Figure 3.3: Logarithmic plot of a sample fading process' amplitude in the vicinity of the receiver at the delay, $\xi$, $|z(x_1, x_2, \xi)|$. Nulls occur regularly. The plot was generated by adding 20 sinusoidal plane waves with uniformly distributed arrival angles and carrier phases.
A receiver moves through the multipath interference pattern, transforming a channel tap's spatial dependence into an apparently time-varying random process. Figure 3.4(a) plots a sample complex process, with time as a parameter. The
amplitude variation of the channel as a function of time is shown in figure 3.4(b),
which effectively is a slice taken from figure 3.3. When the receiver moves slowly
through a deep fade, the SNR is low for a long time. This has serious implications for
coding, as indicated by equation (2.8). The phase variation is shown in 3.4(c). The
phase can jump rapidly at the bottom of deep fades, which has serious consequences
for carrier frequency estimation schemes and the detection of phase-encoded signals.

3.9 SIMULATING THE CHANNEL

In simulations, sample channels must be generated with the correct delay
power spectrum and autocorrelation properties. Methods for achieving suitable
results are described here.

The delay power spectrum is characterised grossly by its delay spread, and
precisely by the profile of the delay power spectrum. However, the delay power
spectrum’s shape has only a second order effect on the receiver. In a delay spread
channel, the chief impediment to reliable communications is the possibility of pulses
smearing across adjacent pulses so that an individual pulse cannot easily be isolated
from the pulse “soup.” A model for the delay power spectrum must replicate this
property by generating a wide and dense delay spread, as well as generating a narrow
delay spread. In this thesis, the P-path channel model,

\[ z(t, \xi) = \sum_{p=0}^{P-1} z_p(t) \delta \left( \xi - \frac{p\tau}{P-1} \right) \quad (3.54) \]

is used; where there are P distinct paths, with uniformly-spaced delays and uniformly
distributed mean powers; and \( z_p(t) \) are independent, circularly symmetric complex
Gaussian random processes. A large P ensures a densely spread delay spectrum when
\( \tau \) is large.

The received signal then equals
\[ y(t) = \int_{-\infty}^{\infty} a(t - \xi) z(t, \xi) d\xi \]
\[ = \sum_{p=0}^{P-1} z_p(t) \int_{-\infty}^{\infty} a(t - \xi) \delta(\xi - \frac{pT}{P-1}) d\xi \]
\[ = \sum_{p=0}^{P-1} a(t - \frac{pT}{P-1}) z_p(t) \]  
\[ (3.55) \]

However, \( pT/(P-1) \) is not normally an integral number of sample periods, \( T_r \), so implementing the delay power spectrum is not straightforward in a sampled simulation. The delay can be implemented within the channel block as a filter,

\[ a \left( lT_r - \frac{pT}{P-1} \right) = \sum_{m=-\infty}^{\infty} a((l-m)T_r) \text{sinc} \left( m - \frac{pT}{(P-1)T_r} \right) \]
\[ (3.56) \]

but this is not the best technique since in practice the \( \text{sinc}(\cdot) \) function must be truncated and the effect is not a simple delay. A computationally convenient method convolves the delay with the transmitted pulse shape, and the transmitter generates \( P \) uniformly delayed versions of the transmitted signal, as

\[ a_p(lT_r) = a \left( lT_r - \frac{pT}{P-1} \right) = \sum_{m=-\infty}^{\infty} \beta_p h \left( lT_r - \frac{pT}{P-1} - iT \right) \exp \left( j2\pi f_{\text{nom}} \left( lT_r - \frac{pT}{P-1} \right) \right) \]
\[ (3.57) \]

which can easily be implemented if \( h(t) \) is known for all \( t \).

The individual channel tap processes, \( z_p(t) \) are correlated circularly symmetric complex Gaussian processes, and can be generated by filtering complex white Gaussian noise, then adding a deterministic component. Only the filter design is of interest. Verdin showed that the autocovariance of equation (3.40) can be created by a filter with impulse response [59],

\[ g_r = \frac{J_{1/4} \left( 2\pi f_{D} T_r \right)}{\sqrt{\left| l \right| T_r}} \]
\[ (3.58) \]

and the impulse response at \( l = 0 \) equals

\[ g_o = \frac{\sqrt{\pi f_{D}}}{\Gamma(5/4)} \]
\[ (3.59) \]
Figure 3.5 (a): Polar plot of a process generated by filtering complex white noise by the truncated but unwindowed impulse response of equation (3.58), showing the unpredictable fine structure.

Figure 3.5(b): Polar plot of a process generated by filtering complex white noise by the truncated and windowed impulse response of equation (3.58). The process is easily predicted.

Figure 3.5(c): Power spectral density of the ideal fading process, and one generated by truncating and Hanning-windowing the impulse response of equation (3.58). $f_0 T = 0.1$, $r = 3, 192$ taps.
Since the impulse response is anti-causal and infinite in extent, it must be delayed and windowed. Windowing widens the resultant process' bandwidth, so must be performed carefully.

First, the windowed impulse response must taper to zero at its ends. The fading process is generated by a receiver travelling through superimposed sinusoids. Therefore the fading process must evolve smoothly, however fast the receiver travels. In Chapter 7, a receiver structure is designed that relies on the fading process' highly correlated property to predict it. An impulse response that does not taper to zero at its ends generates a fading process with an unpredictable fine structure (figure 3.5(a),(b)).

Second, the minimum bandwidth of a finite impulse response (FIR) filter is inversely proportional to its length: the narrower the fading bandwidth required, the longer the filter and the more computationally intensive the filtering is. Accordingly [59] adopted a multi-rate filtering approach, using two filters. The first filter generates a wide bandwidth fading process with the correct autocorrelation properties. Having a wide bandwidth, only a short FIR filter is required. The output is zero padded to reduce the bandwidth to the desired value, then low-pass filtered for image rejection. The second filter has an infinite impulse response (IIR), allowing very narrow bandwidths without an undue number of coefficients. Since the IIR filter is not ideal, attenuated images of the desired spectrum remain at high frequencies.

In this thesis, high fading rates are employed predominantly, so single rate filtering is sufficient, using the impulse response of equation (3.58) windowed by a Hanning window (figure 3.5(c)). The sharp peaks at $f_D$ are diminished, but in fact this better matches the three-dimensional scattering model of the physical channel [25].

Another important technique uses an autoregressive model of the channel. Given a $Z \times 1$ column vector, $z_p$, of samples from the $p$th tap process, $z_{p,i}$, the $Z \times Z$ autocovariance matrix is given by $R_{z,z} = \frac{1}{T} E\left((z_p - E(z_p))(z_{p} - E(z_p))^H\right)$. For WSSUS channels, the autocovariance matrices are identical for all $P$ channel processes. The autocovariance's inverse can be Cholesky decomposed as

$$R_{z,z}^{-1} = L^H \Sigma L$$

(3.60)
where $L$ is a $Z\times Z$ lower triangular matrix with unity on the leading diagonal; and $E$ is a $Z\times Z$ real diagonal matrix. The Cholesky decomposition is a "special" decomposition in that the $x$th row entries of $L$ comprise the parameters to generate an autoregressive process of order $x-1$, and the $x$th diagonal entry of $E$ is the inverse mean square power of the process noise [23]. That is, the fading process can be generated by the recursion,

$$z_t = \sum_{s=1}^{x-1} L_{(s, x-s)} z_{t-s} + e_t \sqrt{E_{xx}} \quad (3.61)$$

where the process noise, $e_t$, is a sequence of unit variance, white i.i.d complex Gaussian random variables. Thus the autoregressive model introduces an unpredictable fine structure, due to $e_t$. The process noise's mean square power can be made small enough to neglect if the process order is chosen large enough.

The autoregressive method has the advantage of computational simplicity. However, there are three disadvantages. Unlike the stable FIR filter of equation (3.58), the IIR filter can be computationally unstable for very narrowband processes. The fading process' autocovariance is correct over $Z$ samples, but is unconstrained otherwise. Thus the fading process' autocovariance and spectrum are slightly incorrect. Finally, an autoregressive model of the fading process is assumed in deriving the receiver structure of Chapter 7, so it is inappropriate to employ a process that is in fact autoregressive in simulations.

Some researchers simulate the fading process with a first order autoregressive model [29, 14, 64]. The process noise power is significant, so the simulated receiver structures exhibit an error floor.

### 3.10 IMPLICIT DIVERSITY

Channels with zero-mean complex Gaussian statistics are particularly difficult to communicate over, since from equation (3.53), there is a non-negligible probability of an arbitrarily deep fade. For instance, the probability of a channel tap fading by $-40$dB is approximately $10^{-4}$. To sustain the low BERs required by modern
communication systems, a high mean SNR is needed to prevent the signal strength dropping close to or below the noise floor too frequently.

The fading problem can be ameliorated through diversity, which will be discussed more fully in section 4.8. The same signal is sent in several different ways, with a much reduced probability that all copies are deeply faded.

In fact, fast fading and frequency-selective channels already provide diversity-like effects, known as *implicit Doppler diversity* and *implicit delay diversity*. The diversity is described as implicit, since it is an intrinsic property of the channel, not the product of deliberate design. In this section, the channel's implicit diversity is discussed.

In channels with delay spread, the receiver observes multiple scaled "echoes" of the transmitted signal. The time-varying component of the multipath is independent across delay from equation (3.39), so each echo is scaled differently. A suitably designed receiver structure can identify and exploit these independently faded echoes, thus achieving a diversity-like effect.

There is a dual property in Doppler spread channels. Information is distributed in time by the transmitter pulse shape, from equation (2.2). In a slow fading channel, fade durations are longer than the pulse duration. Accordingly, there is a high probability that information transmitted during the fade is wholly corrupted by additive noise. By comparison, fade durations are shorter than the pulse duration in a fast fading channel. The pulse shape is faded independently along its length, yet the fading process must attenuate the whole pulse before the information is lost. This also achieves a diversity-like effect.

The Karhunen-Loève (KL) expansion is an important tool for characterising and explaining implicit diversity. Each information phasor, $\beta_n$, modulates the pulse shape, $h(t)$, which is distorted by the channel into the received pulse, $c(t-iT,iT)$. If the probability of incorrectly detecting $\beta_n$ reduces, then the $i$th received pulse must provide the implicit diversity, $c(t-iT,iT)$ is zero for $t < \xi_r$ and $t \geq \xi_l+HT$. Accordingly, the $i$th received pulse is expanded over the interval, $[\xi_r ; \xi_l+HT)$, according to
where \( \{ \Phi_k(t) \} \) is a set of orthogonal functions (basis functions, eigenfunctions), which equal zero outside the interval and have unit energy within the interval; and \( \{ c_k \} \) is a set of coefficients, given by

\[
c_k = \int_{\tau_f}^{\zeta_f+HT} c(t-iT,iT) \overline{\Phi_k(t)} \, dt =
\]

(3.63)

Assuming the channel has no deterministic components, \( c(t-iT,iT) \) is a zero-mean, circularly symmetric complex Gaussian random process and the coefficients are zero-mean circularly symmetric complex Gaussian random variables. In the KL expansion, the basis functions satisfy

\[
\lambda_k \Phi_k(t) = \int_{\tau_f}^{\zeta_f+HT} R_{\tau\tau,jj}(t_1-iT,t_2-iT) \Phi_k(t_2) \, dt_2
\]

(3.64)

where the received pulse autocorrelation equals

\[
R_{\tau\tau,jj}(t_1-iT,t_2-iT) = \frac{1}{2} E \left( c(t_1-iT,iT) \overline{c(t_2-iT,iT)} \right)
\]

\[
= R_{ii}(t_1-t_2) \int_{-\infty}^{\infty} R_{\xi\xi}(\zeta) \mathcal{H}(t_1-\zeta-iT) \overline{\mathcal{H}(t_2-\zeta-iT)} \, d\zeta
\]

(3.65)

Calculating the basis functions from equation (3.64) ensures that the complex coefficients are uncorrelated,

\[
\frac{1}{2} E(c_k c_{k'}) = \lambda_k \delta_{kk'}
\]

\[
\frac{1}{2} E(c_k c_{k'}) = 0
\]

(3.66)

where \( \{ \lambda_k \} \) are the eigenvalues.

Equation (3.62) represents the received pulse as the superposition of an infinite number of orthogonal pulses. Thus the information phasor, \( \beta_{\tau\tau} \), is independently transmitted, infinitely many times: i.e. infinite diversity. From equation (3.66), \( \lambda_k \) is the mean power of \( c_k \).

As shown by Dam [15], most eigenvalues have negligible power compared to the noise power introduced by typical receiver front-ends, and thus have negligible effect on the receiver’s BER performance. At very low SNRs, the noise dominates all eigenvalues, and the BER is approximately 0.5. When the SNR reaches
approximately 0dB (the exact value depends on the signal constellation and the channel), the first eigenvalue appears and the BER reduces by a factor of 10 for every 10dB increase in SNR. As the SNR increases further, more eigenvalues exceed the noise power, and the BER reduces by $10^p$ for every 10dB increase in SNR, where $D$ is the number of eigenvalues above the noise floor. Ordinarily, only one or two diversity orders appear above the noise floor of typical receivers [15].

When the channel is not known ideally, it can be predicted. However, the predictor tap weights must balance accurate prediction against noise enhancement. Accordingly, predictor receivers do not usually achieve the BER-SNR gradient of $10^p$ times BER improvement for every 10dB increase in SNR.

In the very-slow fading, frequency-flat channel, there is no delay or Doppler diversity. The received pulse autocorrelation equals

$$R_{rr}(t_1-iT, t_2-iT) = h(t_1-iT)\hat{h}(t_2-iT)$$

and there is only one basis function, $\Phi_x(t) = 1/\tau HT$. Accordingly, this channel has no implicit diversity.

### 3.11 THE TIME- AND FREQUENCY-SELECTIVE CHANNEL

This is a general model for all linear channels. It applies when both the (inverse) Doppler spread and variation in transit delays for different paths are significant compared to the symbol period. A sample of the transfer function, equation (3.12), is plotted in figure 3.6.

Equation (3.11) can be regarded as a projection operation, where the one dimensional (1d) transmitted signal probes the 2d channel. Only a 1d output, the received signal, is available. Clearly, knowledge of the channel is an important step in detecting the transmitted information.

This deconvolution problem is commonly encountered in imaging, where a 3d object can only be observed as a 2d projection. The 3d object is reconstructed using a series of 2d views from around the object. More accurately, the object occupies 4d: the three spatial dimensions and time, but only varies in the three spatial dimensions.
Thus the 3d image can be reconstructed from the three available dimensions: two spatial dimensions and time.

However, this approach cannot be used generally in communications, since repeated observations of the channel cannot be made (time is one of the channel’s dimensions). Accordingly, channels cannot be reconstructed and coherent detection is impossible if their Doppler and delay spread satisfy

$$f_D \tau > \frac{1}{2}$$

(3.68)

since $f_D$ is a one-sided Doppler spread and $\tau$ is the two-sided delay spread. These channels are called *overspread*. The submarine channel is one example [43]. However, the second-order statistics of these channels can be computed [27], so coherent demodulation and differential decoding are possible using a predictor-based receiver structure.

In the mobile radio channel, equation (3.68) is not satisfied, and the channel can be reconstructed. Pilot tones and symbols can be transmitted with the information-bearing signal, and provide an absolute reference for coherent detection.

![Logarithmic plot of the channel's amplitude transfer function, as a function of time and frequency.](image)

**Figure 3.6:** Logarithmic plot of the $tf$-channel's amplitude transfer function, as a function of time and frequency.

### 3.12 THE TIME-SELECTIVE CHANNEL

The standard channel model for low-rate transmissions to mobile terminals is the time-selective, frequency-flat channel model. This is a specialisation of $z(t, \xi)$ where
\[ z(t, \xi) = z(t) \delta(\xi) \]  
(3.69)
as shown in figure 3.7(a).

In the mobile radio channel, the reflectors tend to be nearby. For example,  
buildings form the reflectors in a central business district, and beyond a few hundred  
metres, the paths are too attenuated to be significant. In this situation, the variation in  
delay between all the received paths is small compared to the relatively large symbol  
period, \( \tau \ll T \). Effectively the channel is memoryless, and therefore cannot and does  
not perform any filtering. It is frequency-flat. Furthermore, vehicularly mounted  
mobiles may be travelling at 100\(km/h\) or above, so that \(f_0T\) is relatively large.

These properties are idealised in a channel that is frequency flat, but fading in  
time. From equation (3.12), the time-variant transfer function equals

\[ T(f, t) = z(t) \]  
(3.70)
where figure 3.7(b) presents a sample \(t\)-channel. Time-selectivity is multiplicative  
distortion by a time-varying random process, as

\[ y(t) = a(t)z(t) \]  
(3.71)

Figure 3.7 (a): Densely tapped delay line model of the \(t\)-channel.

Figure 3.7(b): Logarithmic plot of the \(t\)-channel’s amplitude transfer function, as a function of time and frequency.
3 THE CHANNEL

3.13 THE FREQUENCY-SELECTIVE CHANNEL

The standard channel model for high-rate transmissions to fixed terminals is the time-invariant, frequency-selective channel model. This is a specialisation of \( z(t, \xi) \) where

\[
z(t, \xi) = z(0, \xi)
\]  

(3.72)

as shown in figure 3.8(a).

In high-rate communications, the symbol periods are so short that the path length differences do not need to be large before \( \tau \sim T \). The fractional Doppler is small, since for fixed terminals, the Doppler spreads are due only to the terrain moving: there may be trees rustling in the wind, moving vehicles, or walking pedestrians.

These properties are idealised in a channel that is time-invariant, but fading in frequency. From equation (3.12), the time-variant transfer function equals

\[
T(f, t) = T(f, 0) = \int_{-\infty}^{\infty} z(0, \xi) \exp(-j2\pi ft) d\xi
\]  

(3.73)
where figure 3.8(b) presents a sample $f$-channel. The transmitted signal is convolved with the channel's impulse response, as

$$ y(t) = a(\xi) * z(0,\xi)_{k-1} $$

(3.74)

### 3.14 THE ADDITIVE WHITE GAUSSIAN NOISE CHANNEL

The basic channel model studied in communications is the additive white Gaussian noise (AWGN) channel model. This is a specialisation of $z(t,\xi)$ where

$$ z(t,\xi) = \delta(\xi) $$

(3.75)

as shown in figure 3.9. The channel has no random components. From equation (3.12), the time-variant transfer function equals

$$ T(f, t) = 1 $$

(3.76)

### 3.15 NOISE SOURCES

Up until this section, the presence of interfering signals has been ignored. In practice, a receiver observes the distorted transmitted signal plus co-channel interference, adjacent channel interference, and noise. It is these interfering signals that ultimately prevent the receiver from perfectly recovering the transmitted information.

![Figure 3.9: Densely tapped delay line model of the AWGN channel](image)
In cellular systems, frequencies are reused regularly, albeit not in adjacent cells. Signals transmitted in one cell can propagate to another cell using the same frequencies. This interference is called co-channel interference (CCI). Typically there are many interfering transmitters: a small set of dominant interferers from the nearby cells, and many more weak interferers from more distance cells [12]. The weak interferers can be lumped together as Gaussian noise by the CLT.

However, the dominant interferers cannot be easily dismissed. First, the interfering signal has the same structure as the signal of interest, so a receiver can easily detect the interferer when the desired signal is faded out and the interferer is not. Second, increasing the transmitted signal power improves the signal-to-interference ratio in one cell, but worsens the problem in other cells. Accordingly, the problem of CCI cannot be removed by simply increasing the transmitted power level. Diversity (including coding) is required when dense frequency-reuse patterns are employed.
An optimal receiver structure for this signal model must jointly detect the signal of interest and the dominant interferers, to extract the signal of interest \[12\]. In this thesis, the problem of dominant interferers is not considered.

Thermal noise arises in the receiver's antenna and front-end electronics (low noise amplifiers and mixers). This noise is a real Gaussian random process. Quadrature demodulation converts it into complex Gaussian noise, \( \bar{n}(t) \), having independent real and imaginary components, with equal mean power. Its power spectral density is uniform (up to the bandwidth of the analogue componentry). The thermal noise and the weak interferers are conveniently modelled as additive white Gaussian noise, with a two-sided bandpass noise spectral density of \( N_0/2 \). The complex baseband’s noise autocorrelation equals

\[
\frac{1}{2} E[n(t)\bar{n}(t+\Delta t)] = N_0\delta(\Delta t) \tag{3.77}
\]

The signal-to-noise ratio (SNR) when no power is allocated for channel sounding is defined as the ratio of mean bit energy to noise spectral density,

\[
\frac{E_b}{N_0} = \frac{\frac{1}{2} E \left( \int_{-\infty}^{\infty} |\beta_n h(t-iT-\xi)|^2 dt \right)}{N_0 \log_2 M}
\]

\[
= \frac{\frac{1}{2} E \left( \int_{-\infty}^{\infty} h(t-iT-\xi) \frac{\partial}{\partial \xi} h(t-iT-\xi) d\xi \right)}{N_0 \log_2 M} \tag{3.78}
\]

### 3.16 SUMMARY

The mobile communications channel is difficult to communicate over. The signal is regularly and severely attenuated by the multipath. Additive noise can mask the faded signal, preventing reliable transmission. Moreover, each transmitted pulse is smeared out in time, and the smearing is different each time.
These processes have been represented mathematically in this chapter, as a time-varying convolution called the channel. Two canonical channel models have been described. The channel's statistical properties have been studied: its pdf, mean, and autocovariance in time and in delay. Special cases of the channel have been considered: time-invariant, frequency-flat, and non-fading, distortionless channels.

A block diagram of the transmitter, channel and the RF/IF part of the receiver is shown in figure 3.10.
4 THE RECEIVER

4.1 INTRODUCTION

Before presenting the novel receiver structures of this thesis, the theory and practice of receiver design and analysis are presented in this chapter. Two directions are pursued: the theoretical basis of receiver design, namely statistical detection theory; and the limitations of existing receiver sub-systems in fast fading channels. The former area is pursued in chapters 5 and 7 to derive more optimal receiver structures. The latter area is only a sampling from the literature, to motivate new receiver structures. It is not an exhaustive summary.

4.2 STATISTICAL DETECTION THEORY

The optimum receiver selects the most probable transmitted sequence, using all available information fully. The \textit{a posteriori} probability is the probability that the symbol sequence, \( \{ \alpha \} \) was transmitted, given that \( y(t) \) was received, as

\[
p_{a \mid y(t)}(\{\alpha\} \mid y(t))
\]

The statistically optimum receiver computes \textit{a posteriori} probabilities for all transmitted sequences, then chooses the sequence with the greatest \textit{a posteriori} probability. This receiver structure is called maximum \textit{a posteriori} probability (MAP).

Using Bayes theorem, the \textit{a posteriori} probability can be rearranged as

\[
p_{a \mid y(t)}(\{\alpha\} \mid y(t)) = \frac{p_{a \mid y(t)}(\{\alpha\}, y(t))}{p_{y(t)}(y(t))} = \frac{p_{y(t) \mid a}(y(t) \mid \{\alpha\}) p_{a}(\{\alpha\})}{p_{y(t)}(y(t))}
\]

Ultimately the probability expression is used for decision making, so the denominator, \( p_{y(t)}(y(t)) \), can be discarded, since it is common for all hypothesised sequences. A goal of communications is maximising the information rate, so source coding (e.g. Huffman coding, arithmetic coding) is often employed. The symbol sequence is
approximately white, with *equiprobable* symbols. Accordingly, maximising the *a posteriori* probabilities is equivalent to maximising the conditional probabilities,

$$p_{y(t)|\alpha}(y(t)|\{\alpha\})$$

(4.3)

for all symbol sequences. This is an MLSE structure, and is optimal when the symbols are equiprobable. As written, the conditional probabilities are computed at the end of transmission, whereas a recursive algorithm to compute the conditional probabilities is preferred since transmission may never stop.

A continuous time version of the derivation in [37] is used. The transmission interval begins at $I_G$ sec and ends at $I_C$ sec. These times may be finite or infinite. The received signal in the $i$th symbol interval is defined as the chip,

$$y_r(t) = \begin{cases} y(t) & iT \leq t < (i+1)T \\ 0 & \text{otherwise} \end{cases}$$

(4.4)

The signal up to time $(i+1)T$ is the chip history,

$$Y_r(t) = \begin{cases} y(t) & I_a \leq t < (i+1)T \\ 0 & \text{otherwise} \end{cases}$$

(4.5)

so the conditional probability can be expanded by repeated application of Bayes theorem, as

$$p_{y(t)|\alpha}(y(t)|\{\alpha\}) = \prod_{i=[I_r/I]}^{[I_r/I]} p_{y_r(t)|Y_{i-1},\alpha}(y_r(t)|Y_{i-1}|r,\{\alpha\})$$

(4.6)

Since the logarithm function is one-to-one and monotonic, choosing the transmitted sequence with maximum log-likelihood is equivalent to choosing the transmitted sequence with maximum conditional probability. The log-likelihood is defined as

$$\ln p_{y(t)|\alpha}(y(t)|\{\alpha\}) = \sum_{i=[I_r/I]}^{[I_r/I]} \ln p_{y_r(t)|Y_{i-1},\alpha}(y_r(t)|Y_{i-1}|r,\{\alpha\})$$

(4.7)

where the product has been reduced to a sum. The sequence with the largest log-likelihood function or metric is the maximum likelihood sequence, and it is selected by the receiver. The sequence of complex phasors, $\{\beta\}$, can reconstruct the symbol sequence, $\{\alpha\}$, so it is sufficient for a receiver to maximise the log-likelihood over $\{\beta\}$ instead, where the revised metric equals
\[
\ln p_{y(t)|\beta}(y(t)|\beta) = \sum_{r=\lfloor I_b/T \rfloor}^{\lfloor I_b/T \rfloor + 1} \ln p_{x_r(t)|\beta}(y_r(t)|\beta)
\]

(4.8)

With linear modulations, the information phasor, \( \beta \), does not arrive at the receiver until \( t = kT + \xi \), so the partial sum, \( \sum_{r=\lfloor I_b/T \rfloor}^{k-1} \ln p_{x_r(t)|\beta}(y_r(t)|\beta) \), depends on the transmitted sequence only up to \( \beta_{k-1} \). At the \( k \)th symbol interval, there are \( M^{k-\lfloor I_b/T \rfloor + 1} \) distinct metrics, and in general this number grows exponentially with the transmission duration. Thus choosing the ML sequence involves searching for the best metric through an ever expanding tree.

The \( k \)th log-likelihood, \( \ln p_{x_k(t)|\beta}(y_k(t)|\beta) \), at the \( k \)th symbol period is labelled the branch metric. The running total of branch metrics, from \( \lfloor I_b/T \rfloor \) to \( k \), is labelled the path metric. A sequence of transmitted symbols is called a path, since it defines the branches taken through the tree.

"Per-Sequence-Processing" is the reason for the exponentially increasing complexity. In the general communications problem, the optimal receiver structure has no \textit{a priori} knowledge of the channel. However, knowledge of the channel and other parameters is necessary to compute the branch metrics, and these are progressively estimated. The estimation is normally data-dependent, since the transmitted signal must be deconvolved from the received signal before the channel is revealed. Thus the estimated channel and the branch metric depend on the whole symbol sequence history. The number of branch metrics increases exponentially in time.

Furthermore, the log-likelihoods of equation (4.8) are difficult or impossible to compute when all random processes are considered. The transmitter carrier oscillator, the receiver carrier oscillator, the transmitter symbol rate oscillator and the receiver symbol rate oscillator all introduce random phase noise. The multipath channel has a random number of paths, with a randomly time-varying path attenuations, delays, and arrival angles. The receiver's motion is random. When all the individual pdfs are known, it is mathematically prohibitive to construct the joint pdf. When the pdfs are not known, it is impossible.
Thus the MLSE receiver structure is not implementable, except when the communication system can be described by a simple statistical model, and either the transmission interval is short or the tree search simplifies to a trellis search. Trellis searches arise when no data-dependent quantities need to be computed and the branch metric is a function of a finite number of code states and transmitted symbols.

One example is the transmission of uncoded data through a time-invariant channel corrupted by white noise, when the channel, the carrier's frequency and phase, the symbol rate oscillator's frequency and phase, and the beginning of transmission are completely known [22, 56]. The received pulse shape extends over \( L \) symbol periods. The branch metric is a function of the hypothesis vector, \( \{ \beta_{(i-L+1)r}, \ldots, \beta_{ir} \} \). There are only a finite number, \( M^d \), of hypothesis vectors, which can be mapped to the \( M^d \) branches of an \( M^d-1 \) state trellis. At the \( i \)th symbol period, the trellis' state is controlled by the first \( L-1 \) symbols, \( \{ \beta_{(i-L+1)r}, \ldots, \beta_{(i-1)r} \} \). The last symbol, \( \beta_{ir} \), specifies which of the \( M \) branches are selected. There are \( M \) paths arriving at each state in the \((i+1)\)th symbol period.

[22] and [56] demonstrate that the path metric can be constructed as the sum of independent branch metrics. Accordingly, the exhaustive comparison required to compute the maximum likelihood sequence can be performed iteratively, before the end of transmission. Since a path's metric beyond the \( i \)th symbol period is independent from its path metric before the \( i \)th symbol period, it is sufficient for each state to retain only the path with the best metric from the \( M \) arriving paths. Thus each symbol period, the MLSE receiver extends \( M^d-1 \) surviving paths in \( M^d \) ways, one for each hypothesis vector. Immediately, these paths are pruned back to the best \( M^d-1 \) surviving paths. This is the Viterbi algorithm [22]. Ideally, the algorithm makes no decisions until the end of transmission ("ideal Viterbi"); however, the path histories require linearly increasing storage, and the decisions are delayed too long. In practice, the decision delay is truncated to some fixed value.
4.3 ANALYSING MLSE RECEIVER STRUCTURES

The BER of linear receiver structures is relatively straightforward to compute, since symbols are processed independently and so their errors are independent too. However, this is not the case for non-linear receiver structures, since consecutive errors are not usually independent.

The exact BER can be computed as follows. Given a transmitted sequence, the joint pdf of all hypothesised sequences' path metrics is calculated. This can be viewed geometrically as a density function in a multidimensional space. Each hypothesised sequence is assigned its own (positive only) axis. All the path metrics at the end of transmission can be written as a coordinate vector, specifying a point in this multidimensional space. The value of the density function at this point expresses the likelihood of computing that set of path metrics.

The space can be divided into decision regions. Points within the same decision region share the same largest path metric, and so detect the same maximum likelihood sequence. In fact, a hypothesised sequence's decision region encloses the points closer to its axis than any other.

When the ML sequence is detected instead of the transmitted sequence, there are a number of bit errors - unless the ML sequence is the transmitted sequence. Each region of the joint pdf is weighted by this number, divided by the total number of bits in the transmitted sequence. Then the bit error rate due to the transmitted sequence is calculated by repeatedly integrating over all the weighted joint pdf's dimensions. The overall BER is then this quantity, averaged across all transmitted sequences.

Clearly this exact method has little value since it is difficult to compute. The number of path metrics is increasing exponentially with the transmission length, so the jointly pdf gets very complicated and the number of integrations gets very large.

Forney's union bound technique is more straightforward to compute [22]. The probability that the sequence with the largest path metric is not the transmitted sequence can be upper bounded by the probability that any error sequence has a larger path metric than the transmitted sequence's path metric. Thus a joint pdf is not
needed, only the pdf of the path metric difference, for all possible transmitted and error sequences.

Furthermore, there is no need to compute the pdf over the whole transmission interval. An error sequence follows the same states as the transmitted sequence until the first error. Errors follow until the two sequences merge at a common state again. This sequence of errors is called an error event. Any useful communications system has a low BER, so the error events are normally short compared to the mean time between them. Accordingly, they can be considered independent, so their probability can be calculated by only considering the pdf of the path metric difference in the vicinity of the error event.

Some notation is needed. The actual transmitted sequence is denoted by \( \{ \beta^u \} \). Potential error events are written as \( \{ \beta^{u,v,w} \} \). The superscript \( u \) denotes the length of the error event under consideration. The superscript \( v \) enumerates each distinct transmitted sequence in the vicinity of the length \( u \) error event. Each transmitted sequence can be confused with several others, so the error sequences are enumerated by a further index, \( w \). When an error occurs, the ML sequence is one of the error sequences, \( \{ \beta^{u,v,w} \} \).

The probability that the sequence, \( \{ \beta^u \} \), is transmitted is labelled by \( P(\beta^u) \). The probability that an error sequence has a better metric than the transmitted sequence (the pairwise probability of error) is denoted by \( P(\beta^v \rightarrow \beta^{u,v,w}) \). In general, the pairwise probability of error depends on the correct symbols in the vicinity of the error event as well as the actual erroneous symbols. The number of bit errors that arise from the error event is written \( e(\beta^v \rightarrow \beta^{u,v,w}) \).

An upper bound on the BER can be deduced from a union bound over all error events. Since this is an infinite sum, it must be truncated. The truncated bound is a credible upper bound if at least the dominant error events are considered; the bound is tight if these error events are relatively disjoint.

Thus the BER bound is the union bound of the dominant error events, averaged across the transmitted sequences in the vicinity of the error event,
The form of an error sequence is \( \{ \beta^{u,v,w} \} = \{ \beta^{u,v} \exp(j\theta^{u,v,w}) + \epsilon^{u,v,w} \} \), where the sequences \( \{ \epsilon^{u,v,w} \} \) and \( \{ \theta^{u,v,w} \} \) specify the particular error sequence, and are constrained so that \( \{ \beta^{u,v,w} \} \) is also an allowed sequence. For an error event extending from the \( i \)th to the \( (i+u-1) \)th symbol period, \( \epsilon_{k,v,w}^{u,v,w} \) is zero for \( k < i \) and for \( k > i+u-1 \). When the data is not encoded rotationally-invariantly, \( \theta_{k,v,w}^{u,v,w} \) is always zero; otherwise \( \theta_{k,v,w}^{u,v,w} \) is zero for \( k < i+u \) and is constant for \( k \geq i+u \). This remaining phase offset allows the error event to end when phase lock between transmitter and receiver is lost (cycle slip), since the rotationally invariant code prevents further bit errors. By constraining \( \{ \theta^{u,v,w} \} \) to be zero until the end of the error event, the sequences, \( \{ \epsilon^{u,v,w} \} \) and \( \{ \theta^{u,v,w} \} \), uniquely describe an error event. Therefore the error sequence can be written as

\[
\{ \ldots \beta_{(i-2)}^{u,v}, \beta_{(i-1)}^{u,v}, \beta_{i}^{u,v} + \epsilon_{i}^{u,v,w} \ldots \beta_{(i+u-1)}, + \epsilon_{(i+u-1)}^{u,v,w}, \beta_{(i+u)}^{u,v} \exp(j\theta_{(i+u)}^{u,v,w}) \ldots \} \tag{4.10}
\]

4.4 REDUCED COMPLEXITY RECEIVER STRUCTURES

Since the complexity of the joint MLSE receiver increases exponentially in time and the necessary log-likelihoods cannot easily be computed, reduced complexity systems are employed. In a basic receiver with linear detection, carrier frequency, carrier phase, channel estimation, channel equalisation and symbol timing are acquired by separate sub-systems, as in figure 4.1(a). The symbol-rate oscillators at transmitter and receiver are assumed to be sufficiently precise and stable that the symbol rate is known \textit{a priori} at the receiver. The more sophisticated scheme of figure 4.1(b) uses a fractionally spaced equaliser for joint carrier phase recovery, symbol timing estimation and channel equalisation.
Many successful algorithms exist for each of these tasks, for the channels of most historical interest, namely the AWGN channel and $f$-channel. However, the performance of many of these algorithms degrades substantially in the fast fading channel. In the next two sections, some of these difficulties are described, to motivate the development of new receiver structures, explicitly designed for the fast $t_f$- and $t_t$-channels. Simple received signal models are used for illustrative purposes, since the same or worse problems appear when more sophisticated signalling formats and channel models are used.
4.5 CARRIER RECOVERY

The carrier frequency and phase can be recovered using several different methods: an $M$th power law device and phase-locked loop (PLL); the Costas loop; the decision feedback loop; or by transmitting a pilot tone [46].

Using an $M$th power law device is a straightforward method of carrier frequency and phase recovery. The non-linear device introduces the $M$th harmonic of the carrier, which is filtered by a PLL. An $M$-ary frequency divider recovers a carrier with correct frequency and phase, up to an $M$-ary phase ambiguity. All carrier recovery schemes share this problem. Accordingly, data is either differentially encoded and detected, or a training sequence is employed to resolve the ambiguity. Figure 4.2 is a block diagram of the system. This assumes a BPSK constellation, the rectangular pulse of equation (2.3), bandpass transmission, and a distortionless (time-invariant, frequency-flat and noiseless) channel, as

$$y_c(t) = \Re\{b_{i,j}^{(l)}\exp(j2\pi f_c t)\} \quad (4.11)$$

When the PLL has acquired the carrier frequency but there is a residual phase offset, $\phi$, in the distortionless channel, the VCO is driven by a dc control signal, $\frac{1}{2} \sin(\phi)$, towards phase coherency. The loop bandwidth should be small to minimise the effect of noise on the estimated carrier phase, $\phi$.

In the time-varying channel, the sub-system’s operation differs. Keeping the other signal parameters the same, the received signal equals

$$y_c(t) = \Re\{b_{i,j}^{(l)}z(t)\exp(j2\pi f_c t)\} \quad (4.12)$$

The PLL’s properties can be best visualised by disconnecting the VCO output from the multiplier’s input, and applying $\sin(4\pi f_c t + \phi)$ instead. With this idealised arrangement, the VCO is driven by a time-varying control signal, $\frac{1}{4} \Re\{z^2(t)\exp(-j\phi + j\frac{s}{2})\}$. The loop filter’s bandwidth must be widened to pass the control signal undistorted, so considerably more noise is present in the control signal and the carrier estimation is degraded. More seriously, the $z^2(t)$ factor causes the frequency and phase of the VCO output to range widely, even when there is no
frequency offset. Thus this scheme cannot recover the carrier accurately when $f_0$ gets large.

In fact, even the ML receiver structure cannot maintain phase lock between the transmitter and receiver in a fast fading channel without pilot information or a deterministic channel component [15]. A training sequence ensures phase lock initially, but this is quickly lost. Periodically the signal is attenuated near or below the noise floor for longer than a symbol period, so the complex phasors cannot be detected reliably and the smooth evolution of the received signal is disguised. When the signal strength recovers, the received signal,

$$y(t) = \sum \left\{ \beta_\nu \exp\left[ j2\pi \frac{P}{T} \right] \right\} \left( c(t - iT, iT) \exp\left[ -j2\pi \frac{P}{T} \right] \right)$$

(4.13)

could be due to any one of $P$ rotationally symmetric sequences, $\beta_\nu \exp\left( j2\pi \frac{P}{T} \right)$, distorted by $P$ matching received pulse shapes, $c(t - iT, iT) \exp\left( -j2\pi \frac{P}{T} \right)$, where $p \in \{0, ..., P - 1\}$. Therefore the data must be encoded in a rotationally invariant manner, such as differential encoding over an $M$-PSK constellation, $\beta_\nu = \beta_{(p-1), \nu} \exp\left( j\frac{2\pi}{M} \alpha_\nu \right)$. Differential demodulation or differential detection is used to recover the data.

With differential demodulation in the noisy AWGN channel, using a BPSK constellation, and the rectangular pulse of equation (2.3), the received signal equals

$$\bar{y}(t) = \left\{ \beta_{(\nu/T), r} + n(t) \right\} \exp(j\phi)$$

(4.14)

after carrier frequency acquisition. The receiver multiplies the received signal by its delayed, complex conjugate, then integrates over the symbol period, as

$$\int \bar{y}(t) \bar{y}^*(t_1 - T) dt_1$$

(4.15)
The integral's value at the end of the symbol period is the phase-encoded transmitted data plus signal-dependent noise and noise-noise terms. At high SNR, the noise-noise term is small compared to the signal-dependent noise terms. In this region there is a 3dB penalty compared to ideal coherent detection, due to the signal-dependent noise term, $\beta_{v,T} \int P(t) dt$. This arises since the signal is not correlated with a noise-free reference.

With coherent demodulation and differential detection in the noisy AWGN channel, the receiver first coherently demodulates and detects the differentially encoded phasor sequence, $\{\beta\}$. In general these decisions have a $P$-ary phase ambiguity. Second, the data sequence, $\{\alpha\}$, is differentially detected by computing $\exp(j \frac{2^P}{P} \alpha_{j}) = \beta_{v,T} \beta_{v(t+1)}$, for $M$-PSK. At high SNR, the errors are sparsely distributed, so an error in $\beta_{v}$ usually leads to two errors: one in $\alpha_{v}$ and one in $\alpha_{(v+1)}$. Accordingly, the asymptotic BER of differentially detected $M$-PSK is twice that of coherent detected $M$-PSK.

In fact, the idea of differential demodulation can be extended to surmount the need for carrier frequency recovery also [3]. Given a noise-free bandpass signal after quadrature demodulation and image rejection filtering,

$$y(t) = \beta_{v,T} \exp(j(2\pi f_{0,low} t + \phi))$$

(4.16)

double differential demodulation removes the carrier frequency from the received signal without introducing a phase offset,

$$\int_{[v,T]} y_{v}(t) y_{v}^{*}(t_1 - T) y_{v}(t_1 - 2T) dt_1 = T \beta_{v,T} \bar{\beta}_{v,T} \beta_{v,T+1} \beta_{v,T-1}$$

(4.17)

With suitable coding at the transmitter, the data sequence, $\{\alpha\}$, can easily be recovered from this product of complex phasors, $\beta_{v,T} \bar{\beta}_{v,T} \beta_{v,T+1} \beta_{v,T-1}$. However, the noise performance is degraded further, and the method only generalises easily to the case of full-Nyquist pulse shapes, sampled by the receiver at their peaks, and transmitted over the $\tau$-channel.
Differential demodulation in the time-varying channel performs differently. It is visualised most easily in sampled time, where the received signal is filtered by a zonal filter to limit the noise bandwidth, then sampled at the sub-Nyquist rate of $r = 1$ samples per symbol. Full Nyquist pulses are used, and the sampling phase matches the ISI-free points. Then the received signal equals

$$\tilde{y}_r = (\beta_r z_r + n_r) \exp(j\psi) \quad (4.18)$$

The differential demodulator computes

$$\int_{\tau}^{(\tau+1)} y(t) \tilde{y}(t-T) dt \approx \tilde{y}_r \tilde{y}_{(\tau-1)r} = (\beta_r z_r + n_r) \left( \tilde{\beta}_{(\tau-1)r} \tilde{z}_{(\tau-1)r} + \tilde{n}_{(\tau-1)r} \right) \quad (4.19)$$

which simplifies to

$$\int_{\tau}^{(\tau+1)} y(t) \tilde{y}(t-T) dt \approx \exp(j\alpha_r \frac{2\pi}{M}) z_r \tilde{z}_{(\tau-1)r} \quad (4.20)$$

at high SNR. This comprises the data and channel’s differential phase. Since the detector observes the sum of these phases, it errors when the channel’s differential phase is large. There is a cycle slip. The problem arises because the differential demodulator uses the past symbol as a phase reference for the current symbol.

Since the channel’s differential phase gets large more often as the fading rate increases, the error rate due to cycle slips increases with the fading rate. These cycle slips are a second error mechanism, separate from the additive noise. Accordingly, the BER decreases with reducing noise power at low SNR, until the number of cycle slip errors becomes significant. Reducing the noise power does not improve the BER further. This BER floor is sometimes called an “irreducible error floor,” but this is a misnomer, since the error floor is due to the simple receiver structure, not to the channel. More sophisticated receiver structures can substantially reduce or eliminate this error floor, and it these structures that are pursued in this thesis.

These two error mechanisms are presented in figure 4.3. In figure 4.3(a), additive noise causes an error during a fade, when the instantaneous SNR is low. In figure 4.3(b), no noise is present, but the channel undergoes a rapid phase change during the fade. The differential demodulator is confused and an error arises. In figure 4.3(c), the differential receiver’s BER is plotted against SNR for various fading rates. The BER floors are worst for fast fading.
Figure 4.3(a): Additive noise usually causes errors when the instantaneous SNR is low. The noisy and noiseless channels have different phases at the bottom of the fade, so an error arises.

Figure 4.3(b): There are rapid phase changes at the bottom of fades even in the absence of noise. These also introduce errors.

Figure 4.3(c): BER for QPSK in the Rayleigh fading channel with various fading rates [31], using differential demodulation.
4.6 EQUALISATION

Linear equalisation is used successfully in the $f$-channel, and in the slowly time-varying channel with decision-aided adaptation [47]. The detected symbols have a high probability of correctness, so they can be treated as a known training sequence. A MMSE equaliser adapts its tap weights so as to minimise the error between the equalised output and the detected symbol. The Least Mean Squares (LMS) or Recursive Least Squares (RLS) algorithms are normally used to adapt the tap weights [23]. A diagram is presented in figure 4.4.

![Figure 4.4: Linear equaliser with decision feedback adaptation.](image)

However, this structure is unsuccessful in the fast fading channel. The tap gains are chosen to equalise the channel’s local first order statistics: that is, the channel transfer function, $T(f,t)$, in the immediate past. In the fast fading channel it is necessary to equalise the predicted channel transfer function, since the channel transfer function earlier than $t-T$ can vary considerably from the transfer function at $t$. Inappropriate equalisation leads to a BER floor. In fact, the receiver needs to know how the channel evolves in time in order to predict the channel transfer function; thus the receiver requires the channel’s second order statistics.

The RLS algorithm can be construed as performing prediction, with poorly chosen, exponentially weighted tap weights. This interpretation underlines the deficiencies of conventional equalisers in the fast fading channel. The RLS algorithm is considered for a complex baseband received signal applied to the fractionally-spaced
spaced equaliser. A sampled, $M$-PSK signal with rectangular pulses in the frequency-flat channel is assumed,

$$\tilde{y}_i = \beta(t_{i,r})z_i + n_i$$  \hfill (4.21)

In very slow fading, this channel model closely resembles the AWGN channel. Accordingly, a conventional linear equaliser is not inappropriate, although equalisation is not actually needed. The symbol-rate sampled output of the $r$-tap equaliser is chosen to approximate the desired data phasor,

$$\beta_r = \left[ \begin{array}{c} \tilde{y}_{i-1} \\ \vdots \\ \tilde{y}_r \\ \vdots \\ \tilde{y}_{i-1} \\ \tilde{y}_r \end{array} \right] + e_r$$  \hfill (4.22)

where $\mathbf{b} = [b_1, ..., b_r]^T$ are the equaliser tap weights and $e_r$ is the error in the $i$th symbol due to non-ideal equalisation. A symbol-rate sampler and hard decision device follow the equaliser, so this criterion tends to minimise the number of errors. The RLS algorithm attempts to minimise $\sum_{r=0}^{i} \mu^{i-r} |e_r|^2$ by adjusting the equaliser tap weights. $0 < \mu \leq 1$ is an exponential forget factor. The MMSE equaliser tap weights are given by the normal equations [23],

$$\mathbf{b} = (\mathbf{Y}^H \mu \mathbf{Y})^{-1} \mathbf{Y}^H \mu \mathbf{\beta}$$  \hfill (4.23)

where the $(i+1)\times(i+1)$ forget factor matrix, $\mu$, equals diag($\mu_i, ..., \mu^0$); the $(i+1)\times r$ matrix of received samples, $\mathbf{Y}$, equals

$$\mathbf{Y} = \left[ \begin{array}{c} \tilde{y}_{i-1} \\ \vdots \\ \tilde{y}_0 \\ \vdots \\ \tilde{y}_{i-1} \\ \tilde{y}_r \end{array} \right]$$  \hfill (4.24)

and the $(i+1)\times 1$ column vector of data phasors, $\mathbf{\beta}$, equals $[\beta_0, ..., \beta_r]$. Transmission is assumed to start at $i = 0$.

Equation (4.23)'s first factor resembles an inverse time-averaged autocorrelation matrix. For notational simplicity, the time-averaging of the white noise is assumed perfect, so the inverse expression can be expanded as
Y^H \mu Y = \begin{bmatrix}
\sum_{k=0}^{I} \mu^{-k} \left( z_{(k+1)r-1} \tilde{z}_{(k+1)r-1} + E \left( |n_{(k+1)r-1}|^2 \right) \right) & \cdots & \sum_{k=0}^{I} \mu^{-k} z_{br} \tilde{z}_{(k+1)r-1} \\
\vdots & \ddots & \vdots \\
\sum_{k=0}^{I} \mu^{-k} z_{(k+1)r-1} \tilde{z}_{br} & \cdots & \sum_{k=0}^{I} \mu^{-k} \left( z_{br} \tilde{z}_{br} + E \left( |n_{br}|^2 \right) \right)
\end{bmatrix}

\text{(4.25)}

The remaining factors of equation (4.23) can be expanded as

Y^H \mu \beta = \begin{bmatrix}
\sum_{k=0}^{I} \mu^{-k} z_{(k+1)r-1} \\
\vdots \\
\sum_{k=0}^{I} \mu^{-k} z_{br}
\end{bmatrix}

\text{(4.26)}

where again the noise time-averaging is assumed to be ideal. Noise has only been retained in the signal model so that the matrix, equation (4.25), is invertible in the time-invariant channel.

Equation (4.26) exposes the inadequacy of the forget factor in a fast fading channel. A forget factor of \( \mu = 1 - \delta, \delta \ll 1 \), retains approximately \( 1/\delta \) symbols. If the fading rate is too fast, \( f_DT \gg \delta \), then the sums in equation (4.26) average out the fading process as well as the noise. Only in slow fading channels, \( f_DT \ll \delta \), is the channel sufficiently constant in the recent past that time-averaging is appropriate.

Even for moderate fading, \( f_DT < \delta \), there are problems. The sums in equation (4.26) have the same mathematical structure as linear predictors. The predictor tap weights, \( \mu^{i-k} \), should be specially calculated to predict the current fading process from the distant samples. However, these exponentially weighted taps are unlike the process’ MMSE predictor tap weights; the exponentially weighted taps can only track the process. In a fast fading channel, they can camouflage the available channel information, so that the inverse matrix factor cannot achieve anything.

In fast fading, frequency-flat channels, the differential demodulator is superior to the equaliser described here, since the immediately past symbol is used as for channel state information, not a weighted sum of more distant symbols.
4.7 PILOT INFORMATION

Two techniques have been used in the fading channel to improve the receiver performance. In this section, the application of pilot tones and pilot symbols to the frequency-flat channel is described.

Receiver structures make decisions by comparing the received signal to a number of hypothesised signals, constructed from a hypothesised channel and symbol sequence (when hard, symbol by symbol decisions are made, the hypothesised channel and symbol sequence merely define decision boundaries). Therefore an estimate of the channel at time, \( t = iT \), is required for the decision at time, \( t = iT \), which then defines the channel at time, \( t = iT \). This apparently circular requirement for CSI before it is available can be met through predicting it from past received samples, or by transmitting pilot information which can be isolated before detection. The pilot information is transmitted at the same time as the transmitted signal, and is distorted in a similar manner. It consumes a fraction of the transmitted power, so the SNR definition of equation (3.78) is scaled commensurately.

In the frequency-flat channel, one tone characterises the whole channel for the tone’s duration. Since the frequency-flat model is only an approximation, the best arrangements are one tone in mid-band [63, 39, 18, 1], or two tones placed symmetrically at the band edges [52]. The pilot tones can also be used for carrier frequency and phase recovery. Tones are Doppler spread by the channel, but they must be orthogonal from the data-bearing signal at the receiver. Accordingly, a spectral null in the transmitted signal of at least \( 4f_D \) is required when the tone is placed in mid-band. The spectral null can be achieved through pulse shapes [46], coding [18], or the transmit tone in band (TTIB) technique of band-splitting, frequency translation and recombination [18]. When two tones are employed, they must be positioned at least \( 2f_D \) outside the transmitted signal’s bandwidth and no nulls are needed.

The optimal filter for recovering a pilot tone is the Weiner filter, which at high SNR is approximately a zonal filter with the same bandwidth as the Doppler spread tone. Since this bandwidth varies, the receiver must adaptively control the zonal
filter’s bandwidth [34]. The optimal bandwidth balances the amount of additive noise with the Doppler spread tone’s distortion.

Some pilot tone papers describe a calibration sub-system [52, 18]. It creates a waveform with an amplitude inversely proportional to that of the received pilot tone without affecting the phase. When this modified tone is used to demodulate the received signal, the demodulated output has a much reduced dynamic range. Although this is desirable from an implementation standpoint, it is far from the proper signal processing. There is significant noise amplification during fades, so a sequence estimating receiver structure cannot compute meaningful path metrics.

Pilot tones increase the peak to average power ratio and the dynamic range of the transmitted signal. Therefore power amplification is less efficient [6].

Pilot symbols avoid these problems. Known symbols are regularly time multiplexed with the data symbol stream. By pseudo-randomly varying the symbols, there are no spectral lines. A demultiplexor at the receiver isolates the pilot symbols. The fading process is estimated by filtering the zero-padded pilot symbol sequence with the same zonal filter as used for pilot tones [34]. A multi-pass approach can be used, where this fading process is used to detect the remaining symbols. These can then be treated as pilot symbols, and the fading process estimate can be refined. The improved fading process estimate is then used to detect the symbols again.

In some implementations, a square-root Nyquist pulse shape is used at the transmitter, and the demultiplexor works on the symbol-space sampled matched filter output. However, the matched filter output is only appropriate in slow fading. A better technique uses a full Nyquist pulse shape, so one sample per symbol is an ISI-free estimate of the channel. In this way, pilot symbols can be used for all fading rates, except when the Nyquist rate of the fading process exceeds the fastest pilot symbol rate.

The use of pilot tones or pilot symbols removes the P-ary phase ambiguity of fading channels, permitting coherent detection. The improved BER usually outweighs the power allocated to the pilot information. Pilot information also offers a simple way to estimate the channel, and remove the BER floor.
4.8 DIVERSITY

The second method for improving the BER is diversity. The channel models show that any signal can be faded below the noise floor. The only reliable way to ensure accurate transmission is to send the same signal in $D > 1$ different ways: at different frequencies, repeated at different times, or received by different antennae. Coding is a powerful form of time diversity. Each distinct signal copy present at the receiver is called a thread.

The diversity receiver observes $D$ threads. The $d$th thread, $d = 1, ..., D$, is the transmitted signal distorted in time and frequency by the channel, modelled generally by a possibly non-zero mean complex Gaussian, time-varying, filter, $z^d(t, \xi)$. Gaussian noise, $n^d(t)$, is added, resulting in the $D$ threads, $y(t) = [y^1(t) ... y^D(t)]^T$.

The benefits of diversity are maximum when the threads are independently faded: clearly if the threads are dependently faded, then one thread entering a fade increases the probability that the other threads are also faded. If the signal copies are faded independently, then the BER decreases by $10^D$ for every $10\text{dB}$ increase in SNR. Thus diversity can significantly reduce the SNR required to achieve a given BER. Figure 4.5 shows graphically how diversity can reduce the effect of fades.

Frequency diversity arises from the properties of multipath. The phase of the $i$th path depends on the product $f_i \tau_i$. Therefore the arriving paths superimpose differently at different frequencies, so the fades appear at different locations. A system designer can ensure maximum diversity by transmitting the same signal at different frequencies over uncorrelated channels, since an uncorrelated Gaussian channel signifies independent fading. Accordingly, the channel’s correlation in frequency, equation (3.29), is needed.

The channel correlation generally decays with increasing frequency separation, but it is significant over an interval, the channel’s coherence bandwidth [2]. It is inversely proportional to the delay spread. When the delay spread is large, the channel decorrelates quickly, so independent fading can be achieved with signals closely spaced in frequency. However, the required bandwidth increases linearly with the diversity order, so the system is not bandwidth efficient.
A better scheme pools the bandwidth from several users, then divides it into frequency bins. Each user fast frequency hops its signal across the bins, so the faded and non-faded bins are shared out equally between users. Orthogonal hopping codes can be used to isolate users in the AWGN channel, but orthogonality is not preserved in multipath channels, where the signals are distorted differently. Accordingly, a receiver detects a signal from the desired user plus interfering signals from all other users. The system is limited by the number of users, not noise, and therefore it is outside the scope of this thesis.

Another suitable scheme requires timely channel state feedback. The base station has several channels at different frequencies, each time-divided between multiple users. The channels are spaced sufficiently in frequency that they are independently faded. The mobiles are widely spaced, so they are also independently faded. When a mobile or the base station detect that the mobile's channel is entering a fade, the base station transfers the mobile to a second channel at a second frequency. In effect this is a form of selection diversity, but the channel selection is negotiated at a protocol level. Therefore the fading must be relatively constant over hundreds of symbol periods. In principle, no additional bandwidth is required. Bandwidth is made
made available on the second channel by swapping another mobile from the second channel to the first channel at the same time. Since the mobiles are independently faded, the first channel is normally non-fading to the second mobile.

Pure time diversity repeats the same signal $D$ times, so it is also bandwidth inefficient. However, time diversity can be achieved without bandwidth expansion, through coding. Redundancy is added so that an error event only occurs if several symbols are detected erroneously. Joint coding and modulation achieves the necessary redundancy through expanding the constellation.

However, coding is hampered by the need for the interleaving of equation (2.8). The channel's coherence time is the dual of its coherence bandwidth. The channel correlation generally decays with increasing time difference, but it is significant over an interval, the channel's coherence time. This is inversely proportional to the Doppler spread. To obtain maximum diversity, a system designer must ensure that the time repetitions or codes are smeared beyond the channel's coherence time for the maximum diversity benefit, so delay is essential. Interleaving is one option. However, when the channel and receiver are immobile (for instance, a wireless PC in an empty office), the coherence time is unbounded. In these situations coding does not provide diversity, and a receiver located in a fade performs poorly.

One technique that does not reduce the system capacity or introduce delay is receiver space diversity. The receiver has $D$ physically separated antennae. A system designer maximises the diversity by ensuring that the antennae receive independently faded signals. Since the channel is assumed Gaussian, the independence criterion simplifies to an uncorrelated criterion. Thus the optimum antenna spacing occurs at zeros of the channel's spatial autocorrelation function; from equation (3.45), this occurs at 0.383 wavelengths. Often antennae are spaced at one quarter wavelength intervals, so an array of $D$ antennae in a line is $(D-1)c/4f_c \sim 0.08$ m in length at a 900 MHz carrier frequency with two antennae. This volume is not unreasonable, and decreases with increasing carrier frequency. However, the receiver does require $D$ antennae and RF chains.
4.9 SUMMARY

In the past three chapters, a communications system has been described. The transmitter and channel properties have been studied. Statistical detection theory for this system has been presented, as well as the limitations of various existing receiver structures. Underlying these topics is the main purpose of the three chapters: namely developing a consistent, realistic and general signal model as a grounding for subsequent chapters. This process is finished here by addressing symbol timing. The final signal model is summarised.

The receiver detects one symbol every symbol period. However, it does not know a priori when transmission begins, or the transit delay, $\xi_f$. In practice, the beginning of transmission is characterised by a ramping signal level, as the received signal comprises first noise, then the first symbol's pulse tail, then finally the sequence of main pulse lobes. By observing this power variation, a receiver calculates a crude estimate of the beginning of transmission and the symbol timing. The error in this estimate is the timing error, $t_0$. Define $T_0$ to be large enough so that the timing error, $t_0$, lies within the interval [0; $T_0$]. Although the interval [-$T_0/2$; $T_0/2$] is a more natural for the timing error, this unduly complicates subsequent notation. The former interval is used, since clearly the two intervals are equivalent.

With this definition, a pulse transmitted at time $t$ stretches over $[iT; iT+HT]$. The channel extends the pulse to $[iT+\xi_f; iT+HT+\xi_d]$. The receiver's time origin is decoupled from the transmitter's time origin since it is acquired from the ramping signal level at the beginning of transmission. Therefore the received pulse appears to the receiver to lie within $[iT; iT+HT+t+T_0]$. Define

$$L = \left[ H + \tau/T + T_0/T \right]$$

(4.27)

This is the length in symbol periods of the received pulse, accounting for the transmitter pulse shape width, delay spread, and timing uncertainty. Thus the generalised linear modulation notation of equation (3.18) can be written as

$$y''(t) = \sum_{i=-L}^{L} b_i c^d(t-iT,T) + n''(t)$$

(4.28)
where the received pulse shape is defined as

\[ c^d(t - iT, iT) = \int_{-\infty}^{\infty} h(t - iT - t_0 - \xi) z^d(t - t_0, \xi) \exp(j2\pi f_{0,\text{non}}(t - t_0 - \xi) + j\phi) d\xi \]  

The noise is also time-shifted by \( t_0 \), but since it is assumed strictly stationary, the time shift need not be reflected in the notation.

In discrete time,

\[ y^d_i = \sum_{i_0 = -L+1}^{L} \beta_{\nu} c^d_{i-i_0, j} + n^d_{i0} \]  

where the received pulse shape is defined as

\[ c^d_{i-i_0, j} = \int_{-\infty}^{\infty} h((m - kr)(iT) - t_0 - \xi) z^d((m - k \nu)(iT) - t_0 - \xi) \exp(j2\pi f_{0,\text{non}}((m - k \nu)(iT) - t_0 - \xi) + j\phi) d\xi \]  

The autocorrelation of the received signal equals

\[ \frac{1}{2} E(\mathbf{y}^d_i \mathbf{y}^d_i^\dagger \mathbf{f}_{0,\text{non}}, \Phi, t_0) = \sum_{\nu = -L+1}^{L-1} \sum_{\nu = -L+1}^{L-1} \beta_{\nu} E(c^d_{i-i_0, j} \mathbf{c}^d_{i-i_0, j}^\dagger \mathbf{f}_{0,\text{non}}, \Phi, t_0) + \frac{1}{2} E(n^d_{i0} \mathbf{n}^d_{i0}^\dagger ) \]  

where the noise autocorrelation is governed by a zonal anti-aliasing filter with bandwidth \( 1/2T \),

\[ \frac{1}{2} E(n^d_{i_0} \mathbf{n}^d_{i_0}^\dagger ) = \begin{cases} N_0/T, & d_1 = d_2 \\ 0, & \text{otherwise} \end{cases} \]

assuming independent noise between threads. The received pulse autocorrelation is

\[ \frac{1}{2} E(c^d_{i-i_0, j} \mathbf{c}^d_{i-i_0, j}^\dagger \mathbf{f}_{0,\text{non}}, \Phi, t_0) = \int_{-\infty}^{\infty} h((I - ir)(iT) - t_0 - \xi) h((m - k r)(iT) - t_0 - \xi) \exp(j2\pi f_{0,\text{non}}((I - ir)(iT) - t_0 - \xi) + j\phi) d\xi \]

\[ \times \exp(j2\pi f_{0,\text{non}}((I - r)(iT) - (\xi_1 - \xi_2)) \]  

and the channel autocorrelation in the absence of deterministic components equals
\[
\frac{1}{2} E\left( z^{d_1} (mT_r - t_0, \xi_1) \bar{z}^{d_2} (mT_r - t_0, \xi_2) \right) \\
= \delta(\xi_1 - \xi_2) R_{\xi_1}(\xi_1) J_0\left(2\pi \sqrt{\frac{(d_1 - d_2)^2 W^2 + (m - l)^2 f_D^2 T_r^3}{2(d_2 - d_1)(m - l)Wf_DT_r \cos(\theta - \varphi)}} \right)
\]
5 DESIGN OF THE EXTENDED MLSE RECEIVER

5.1 INTRODUCTION

This chapter develops an MLSE diversity receiver for the time- and frequency-selective channel corrupted by additive Gaussian noise, when linear signal constellations are employed. Ungerboeck's derivation of the Extended MLSE receiver for the purely frequency-selective channel is extended to the more general channel \([56]\) with diversity\(^1\). Accordingly the new receiver is known as the "Extended MLSE diversity receiver for the time- and frequency-selective channel," or EMLSE-tf diversity receiver. Suffixes other than -tf denote other channel models. The channel need not be complex Gaussian or WSSUS, and the noise need not be white.

The new receiver structure assumes perfect knowledge of the carrier, symbol timing, and channel state information (CSI); all are available before detection. Only the additive noise is unknown. However, the receiver can also be used with imperfect estimation subsystems, at the cost of optimality and a poorer BER.

High quality CSI is needed by any high performance receiver, so it is not an unduly restrictive requirement. As will be discussed, the time- and frequency-selective channel can be estimated from a comb of pilot tones or time-isolated pilot symbols. In the time-invariant channel, a training sequence is sufficient. In the frequency-flat channel, only one pilot tone is needed. Pilot symbols do not need to be time-isolated if full-Nyquist pulse shapes are used. No CSI is required for the AWGN channel.

There are several highlights in this chapter. A finite-complexity diversity receiver is derived that is ML for all linear channel models and sources of diversity, as long as ideal CSI is available. The new receiver structure offers insight into matched

\(^1\) A subset of this research was also conducted in parallel but independently by Bottomley and Chennakeshu \([4]\).
filtering and ML diversity receiver processing for the \( tf \)-channel and its special cases. This receiver structure is of theoretical interest, since it is optimal (it maximises the mean time between error events). At moderate and high SNRs, its BER can be used as a benchmark, since it is essentially a lower bound on the performance of practical systems, which either lack ideal CSI or are not ML. In the next chapter, bounds on the new receiver's BER for ideal CSI and pilot tone CSI are derived for a fast Rayleigh fading channel with multiple independently faded paths.

### 5.2 EMLSE-\( tf \)DIVERSITY RECEIVER DERIVATION

The MLSE receiver searches all allowed phasor sequences in the transmission interval and chooses the one with maximum likelihood. The interval is assumed long enough that its endpoints, \( I_B \) and \( I_E \), can be assumed \( \pm \infty \) without penalty.

The receiver observes a vector of \( D \) signals, \( y(t) = [y^1(t), ..., y^D(t)] \), where \( D \) is the diversity order. Each entry comprises a signal term and a noise term. By conditioning on the signal terms, the received signal vector equals a deterministic component plus circularly symmetric complex Gaussian noise. Its pdf is given by equation (3.48). Conditioning on the signal terms is equivalent to assuming that the receiver has perfect synchronisation and CSI before detection. The channel, \( z^d(t, \xi) \), may be correlated between threads, but we shall see that this has no influence on the receiver structure for ideal CSI.

To condense the notation, the vector, \( \eta \), is defined as a list of the synchronisation parameters plus the \( D \) sequences of received pulses,

\[
\eta = \{f_{0, \text{nom}}, \phi, t_0, c^1(t, iT), ..., c^D(t, iT)\} \quad (5.1)
\]

The \( d \)th vector, \( \eta^d \), is defined as the synchronisation parameters and the sequence of received pulses at the \( d \)th antenna,

\[
\eta^d = \{f_{0, \text{nom}}, \phi, t_0, c^d(t, iT)\} \quad (5.2)
\]

Thus the log-likelihood of \( y(t) \), conditioned on the phasor sequence and the vector, \( \eta \), matches the log-likelihood of the Gaussian noise vector, \( \eta(t) = [n^1(t) ... n^D(t)]^T \).
\[ \ln p_{y|\beta, \eta}(y|\beta, \eta) = \ln p_n(n) \quad (5.3) \]

For space diversity, each thread's noise comes from different antennae and front-end electronics. For frequency diversity, the noise occupies different bandwidths. Therefore the noise is assumed independent between threads. Then the log-likelihood of equation (5.3) is the sum of each thread's log-likelihood, so that

\[ \ln p_{y|\beta, \eta}(y|\beta, \eta) = \sum_{d=1}^{D} \ln p_n(n^d) = \sum_{d=1}^{D} \ln p_{y|\beta, \eta}(y^d|\beta, \eta^d) \quad (5.4) \]

By conditioning on \( \eta \), the statistical properties of its components do not affect the pdf. The receiver is ML for all non-ideal carrier and symbol-rate oscillators and all linear channels, including \( f_1, \tau, f \) and AWGN channels. When ideal CSI is not available, the performance of this receiver still leads to a lower bound on a practical receiver's performance.

The only random process in the single thread is the complex Gaussian noise, so from equation (3.49), the log-likelihood of a single thread equals

\[ \ln p(y^d|\beta, \eta^d) = -\int \int \left[ \bar{y}^d(t_1) - \sum_i \beta_i c^d(t_1 - iT, iT) \right] R_{\text{fil}}^{-1}(t_1 - t_2) \left( y^d(t_2) - \sum_k \beta_k c^d(t_2 - kT, kT) \right) dt_1 dt_2 \]

\[ \quad - \int \int \left[ \bar{y}^d(t_1) - \sum_i \bar{\beta}_i c^d(t_1 - iT, iT) \right] R_{\text{fil}}^{-1}(t_1 - t_2) \left( y^d(t_2) - \sum_k \bar{\beta}_k c^d(t_2 - kT, kT) \right) dt_1 dt_2 \quad (5.5) \]

This log-likelihood can be transformed into a path metric, \( \Lambda(y^d|\hat{\beta}, \eta) \), by replacing the transmitted phasor sequence, \( \{\beta\} \), with the hypothesised phasor sequence, \( \{\hat{\beta}\} \); interchanging the order of summation and integration; and neglecting the term, \( -\int \int \bar{y}^d(t_1) R_{\text{fil}}^{-1}(t_1 - t_2) y^d(t_2) dt_1 dt_2 \), since it is independent of the hypothesised phasor sequence. Therefore the path metric equals

\[ \Lambda(y^d|\hat{\beta}, \eta) = \sum_i 2R \left[ \bar{\beta}_i, m_{i^d} \right] - \sum_i \sum_{k} \bar{\beta}_k s_{i^d, k} \hat{s}_{i^d, k} \quad (5.6) \]

where the matched filter term, \( m_{i^d} \), and the ISI term, \( s_{i^d, k} \), are defined as
\[ m_{ir}^d = \int \int y^d(t_1) R^{-1}_{mr}(t_1 - t_2) \mathcal{E}^d(t_1 - iT, iT) dt_1 dt_2 \]
\[ = y^d(t_1) * R^{-1}_{mr}(t_1) * \mathcal{E}^d(-t_1, iT) \]  
\[ (5.7) \]
\[ s_{ir,kr}^d = \int \int \mathcal{E}^d(t_1 - iT, iT) R^{-1}_{mr}(t_1 - t_2) \mathcal{E}^d(t_2 - kT, kT) dt_1 dt_2 \]

As with Ungerboeck's EMLSE-\( f \) receiver [56], there is no need for a noise-whitening filter. Substituting equation (5.6) into (5.3) leads to

\[ \Lambda(y, \hat{\eta}) = \sum_{i} 2 \Re \left\{ \hat{\eta}_i m_{ir} \right\} - \sum_{i,k} \hat{\eta}_i s_{ir,kr} \hat{\eta}_k \]
\[ (5.8) \]

where \( m_{ir} \) and \( s_{ir,kr} \) are the sum of all the threads' \( m_{ir}^d \) and \( s_{ir,kr}^d \) terms respectively,

\[ m_{ir} = \sum_{d=1}^{D} m_{ir}^d \]
\[ (5.9) \]
\[ s_{ir,kr} = \sum_{d=1}^{D} s_{ir,kr}^d \]

The symbol-spaced sequence, \( m_{ir} \), is a set of sufficient statistics for the detection of \( \{ \beta \} \). The conjugate symmetry of \( s_{ir,kr} \), \( s_{ir,kr} = \bar{s}_{kr,ir} \), can exploited by properly grouping the double summation of equation (5.8) to obtain a simpler form of the metric,

\[ \Lambda(y, \hat{\eta}) = \sum_{i} 2 \Re \left\{ \hat{\eta}_i m_{ir} - \sum_{k} \hat{\eta}_k s_{ir,kr} \hat{\eta}_k \right\} - \left| \hat{\eta}_i \right|^2 s_{ir,ir} \]
\[ (5.10) \]

Exhaustive comparison is needed to find the maximum metric and thus the maximum likelihood sequence. Provided that the evolution of the path metric is Markovian with a finite number of states, then the task is undertaken efficiently by the Viterbi algorithm.

As long as the ISI term, \( s_{ir,kr} \), in equation (5.9), extends over only a finite number of symbols, \( L \), then \( s_{ir,kr} = 0 \) for \( |k - i| \geq L \), and the path metric defined in equation (5.10) further simplifies to

\[ \Lambda(y, \hat{\eta}) = \sum_{i} 2 \Re \left\{ \hat{\eta}_i m_{ir} - \sum_{k=i-L,s}^{i-1} \hat{\eta}_k s_{ir,kr} \hat{\eta}_k \right\} - \left| \hat{\eta}_i \right|^2 s_{ir,ir} \]
\[ (5.11) \]
It can be written in recursive form, where the path metric at the \(i\)th symbol period, \(A_{(i)} (y, \tilde{\beta}, \eta)\), equals the path metric at the \((i-1)\)th symbol period, plus a branch metric, \(A_{(i-1)} (y, \tilde{\beta}, \eta)\), as

\[
A_{(i)} (y, \tilde{\beta}, \eta) = A_{(i-1)} (y, \tilde{\beta}, \eta) + \lambda_{(i)} (y, \tilde{\beta}, \eta)
\]

\[
= \left( \sum_{s=0}^{L} \text{Re} \left[ \hat{\beta}_{s} m_{s} - \sum_{k=1}^{L-1} \hat{\beta}_{s} s_{s,k} \hat{\beta}_{s} \right] - \hat{\beta}_{s}^2 s_{s,0} \right) + \left( \sum_{s=0}^{L} \text{Re} \left[ \hat{\beta}_{s} m_{s} - \sum_{k=1}^{L-1} \hat{\beta}_{s} s_{s,k} \hat{\beta}_{s} \right] - \hat{\beta}_{s}^2 s_{s,0} \right)
\]

(5.12)

The path metric, \(A_{(i-1)} (y, \tilde{\beta}, \eta)\), is a function of the hypothesised phasors, \((\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{L})\). The branch metric is a function of the \(L\) hypothesised phasors, \((\hat{\beta}_{(i-L+1)}, \ldots, \hat{\beta}_{i})\). A state,

\[
\sigma_{(i-1)} = (\hat{\beta}_{(i-L+1)}, \ldots, \hat{\beta}_{i})
\]

(5.13)

is associated with each surviving path metric, \(A_{(i-1)} (y, \tilde{\beta}, \eta)\). When the path metric is extended to the \(i\)th symbol period, the branch metric, \(A_{(i)} (y, \tilde{\beta}, \eta)\), is computed. It depends only on the path metric’s state and the new phasor, \(\beta_{i}\). Thus the path metric evolution is Markovian, since it depends only on the previous state. By restricting the ISI length to \(L\) symbols, the number of states is finite, at \(M^{L-1}\).

Thus the Viterbi algorithm can be used to find the maximum likelihood sequence. It processes a partially-connected trellis, with \(M^{L-1}\) states, and \(M\) branches per state. Every \(L\)-phasor hypothesis vector, \(\hat{\beta}_{(i-L+1)}, \ldots, \hat{\beta}_{i}, \hat{\beta}_{i}\), labels a state and a branch.
5.3 WHITE NOISE

For the important special case of white noise with the complex baseband autocorrelation, $R_{nw}(\Delta t) = N_0 \delta(\Delta t)$, the receiver structure is particularly simple. Since the metric of equation (5.10) is used for comparisons only, the scale factor of $N_0$ can be neglected. Then using $R_{nw}^{-1}(\Delta t) = \frac{\delta(\Delta t)}{N_0} \propto \delta(\Delta t)$ simplifies $m_r$ and $s_{r,kr}$ to

$$m_r = \sum_{d=1}^{n} \int y^d(t_1) \bar{c}^d(t_1 - iT, iT) dt_1 = \sum_{d=1}^{n} y^d(t_1) * \bar{c}^d(-t_1, iT)_{t_1 = t}$$

$$= \beta y \sum_{d=1}^{n} \int c^d(t_1 - iT, iT) dt_1 + \sum_{k \neq i} \beta_{r,k} s_{r,kr} + \sum_{d=1}^{n} \int c^d(t_1 - iT, iT) dt_1$$

$$s_{r,kr} = \sum_{d=1}^{n} \int c^d(t_1 - iT, iT) c^d(t_1 - kT, kT) dt_1$$

$m_r$ can be considered as the symbol-rate sampled output, $t = iT$, of the sum of matched filters for the time- and frequency-selective channel (MF-tf). The MF-tf completely removes the phase distortion (due to the complex received pulse) of the phasor sequence in each thread. Furthermore the peak symbol power to expected noise power ratio of the $i$th symbol is maximised by summing the signal weighted according to its instantaneous received pulse power, $|c^d(t - iT, iT)|^2$. It is important to understand that this in itself does not deal with ISI, and in fact may make it worse. The complete EMLSE-tf receiver of equation (5.10) deals with the past ISI by cancelling or equalising it by means of $s_{r,kr}$. This avoids an error floor.

An intuitive appreciation of the path metric is not easily gained, since the term,

$$- \sum_{d=1}^{n} \int \int y^d(t_1) R_{nw}^{-1}(t_1) y^d(t_2) dt_1 dt_2,$$

has been neglected. By substituting $R_{nw}^{-1}(\Delta t) = \frac{\delta(\Delta t)}{N_0} \propto \delta(\Delta t)$ into the original log-likelihood expression, equation (5.5), over all threads, the modified path metric is seen to be a simple Euclidean distance,

$$\Lambda(y, \beta, \eta) = - \sum_{d=1}^{n} \int y^d(t_1) - \sum_{d'=1}^{n} \beta_{r,d'} c^d(t_1 - iT, iT) \int dt_1$$

(5.15)
The actual path metric is merely a mathematical transformation of this metric. When the noise is white, the condition that \( s_{i_r,k_r} \) be zero for \( |k-i| \geq L \) is satisfied by restricting the pulse length, \( H \), delay spread, \( \tau \), and maximum timing offset, \( T_0 \), to be finite. Then \( L \) is given by equation (4.27).

However, the telephone channel is best modelled as having an IIR impulse response, so an MLSE receiver must use a sub-optimal, truncated impulse response. Although the truncation error can be made negligible by choosing \( \tau \) large enough, the number of states grows large very rapidly. Accordingly, in most instances equalisation is more suited to the telephone channel than MLSE.

### 5.4 EMLSE-\( tf \) DIVERSITY RECEIVER OPERATION

The receiver structure can be explained loosely as follows. The \( i \)th phasor scales the \( i \)th received pulse. To recover this phasor, the receiver employs a filter matched to its received pulse. This matched filter optimally captures all the available information from the received pulse, but it achieves this at the expense of ISI: the matched filter output also contains filtered pulse tails from adjacent symbols. To construct a Euclidean distance (or some mathematical transformation thereof), the receiver must estimate not only the contribution from the filtered \( i \)th received pulse, but also the filtered pulse tails from the adjacent symbols. Since the MLSE receiver considers all possible sequences, it must hypothesise all \( M^L \) combinations of the \( i \)th and adjacent symbols. This requirement leads to multiple states and the Viterbi algorithm.

During the \( i \)th symbol period, the receiver forms one value of \( m_{i_r} \) and \( L \) values of \( s_{i_r,k_r} \) according to equation (5.9) or (5.14). The \( M^L \) branch metrics are calculated according to equation (5.11), and then applied to a standard Viterbi processor. The EMLSE-\( tf \) space diversity structure is shown in figure 5.1.

In a practical receiver, ideal CSI is unavailable, so an estimate of the CSI, \( \hat{z}^d(t,\xi) \), replaces \( z^d(t,\xi) \) in the \( m_{i_r} \) and \( s_{i_r,k_r} \) expressions. Strictly speaking, the receiver is optimal for ideal CSI only. The CSI estimate may be written as,
\[ \hat{z}^d(t, \xi) = z^d(t, \xi) + \Delta^d(t, \xi) \]  
(5.16)

where \( \Delta^d(t, \xi) \) is the CSI estimation error. The estimated received pulse equals

\[ \hat{c}^d(t - iT, iT) = \int_{-\infty}^{\infty} \tilde{h}(t - iT - t_0 - \xi) \hat{z}^d(t - t_0, \xi) \exp(j2\pi f_{0, nom}(t - t_0 - \xi) + j\phi) \, d\xi \]  
(5.17)

and the estimation error in the received pulse shape is defined as

\[ e^d(t - iT, iT) = \int_{-\infty}^{\infty} \tilde{h}(t - iT - t_0 - \xi) \Delta^d(t - t_0, \xi) \exp(j2\pi f_{0, nom}(t - t_0 - \xi) + j\phi) \, d\xi \]  
(5.18)

so that

\[ \hat{c}^d(t - iT, iT) = c^d(t - iT, iT) + e^d(t - iT, iT) \]  
(5.19)

For ideal CSI, the estimation error in the received pulse is zero; for pilot-based systems, it is non-zero.

5.5 MAXIMAL RATIO COMBINING

The EMLSE-tf diversity receiver is the optimum structure for a very general channel model. In this and subsequent sections, its structure in simplified channels is described. In many cases, we find that the structure is already known.
When square-root Nyquist pulses are transmitted over the time-invariant, frequency-flat channel, corrupted by white noise, and received by a diversity receiver that recovers the carrier frequency, the Maximal Ratio Combiner (MRC) is the optimal receiver structure [5]. The thread phases are aligned; they are weighted by their signal strength; the summed outputs are matched filtered; then the receiver makes hard, symbol-by-symbol decisions using a Euclidean metric.

The MRC receiver structure can be deduced from the EMLSE-(f diversity receiver. In the time-invariant, frequency-flat channel, \( z^d(t,\xi) = z^d(0)\delta(\xi) \) so the received pulse is time-invariant also,

\[
\bar{z}^d(t - iT, iT) = \bar{z}^d(t - iT) = h(t - iT - t_o)z^d(0)\exp(j\phi)
\]

(5.20)

where the thread phase is \( \angle z^d(0)\exp(j\phi) \) and the thread signal strength is \( |z^d(0)|. \)

The EMLSE-(f diversity receiver's matched filter computes

\[
m_d = \left[ \sum_{d=1}^{D} \left( y_d(t)z^d(0)\exp(-j\phi) - h(t - t_o) \right) \right]_{t=T}
\]

(5.21)

showing the phase alignment, the threads weighted according to their signal strength, the summing, and the matched filtering. Since square root Nyquist pulses are used, there is no ISI, and

\[
s_{\nu,\nu} = \sum_{d=1}^{D} z^d(0)\exp(j\phi)z^d(0)\exp(-j\phi)\int_{-\infty}^{\infty} h(t_i - iT)h(t_i - kT)dt_i
\]

(5.22)

so \( L = 1. \) Therefore the trellis has one state with \( M \) branches, and the Viterbi algorithm reduces to a hard decision device. The branch metric equals

\[
\lambda_{\nu,\nu} \left( \hat{y}, \hat{\nu}, \eta \right) = 2\Re\left( \hat{\beta}, m_{\nu} \right) - |\hat{\beta}_{\nu}|^2 s_{\nu,\nu}
\]

(5.23)

which can be transformed into a Euclidean distance by scaling the branch metric by \( s_{\nu,\nu}, \) then adding \(-|m_{\nu}|^2, \) as
\[
\lambda_n \left( y \bar{\beta}, \eta \right) - \left| m_n \right|^2 + \left( 2 \Re \left( \bar{\beta}_n m_n \right) - \left| \bar{\beta}_n \right|^2 s_{ir, lr} \right) s_{ir, lr} \\
= -\left| m_n - \bar{\beta}_n s_{ir, lr} \right|^2
\]  

(5.24)

These modifications have no influence on the decision making since both \( s_{ir, lr} \) and \(-\left| m_n \right|^2 \) are independent of the hypothesised phasor sequence, and \( s_{ir, lr} \) is positive real.

### 5.6 THE FREQUENCY-SELECTIVE CHANNEL

Since it is Ungerboeck’s derivation in the \( f \)-channel that spawned the EMLSE-\( tf \) diversity receiver, it is unsurprising that the EMLSE-\( tf \) diversity receiver reduces to the EMLSE-\( f \) receiver in the time-invariant channel without diversity. Since there is only one thread, the thread superscripts are discarded. The received pulse is time-invariant,

\[
\tilde{c}(t - iT, iT) = \tilde{c}(t - iT) = \int_{-\infty}^{\infty} h(t - iT - t_0 - \xi) z(0, \xi) \exp(j\phi) d\xi
\]  

(5.25)

and so too are the matched filter and ISI values, yielding

\[
m_n = \int_{-\infty}^{\infty} y(t_1) \tilde{c}(t_1 - iT) dt_1
\]  

(5.26)

\[
s_{ir, kr} = s_{ir, i\cdot} = \int_{-\infty}^{\infty} \tilde{c}(t_1 - iT) \tilde{c}(t_1 - kT) dt_1
\]

That \( s_{ir, kr} \) depends only on symbol differences, \( i-k \), can be demonstrated by making the substitution, \( t_2 = t_1 - iT \), so that

\[
s_{ir, kr} = s_{ir, i\cdot} = \int_{-\infty}^{\infty} \tilde{c}(t_2) \tilde{c}(t_2 + (i - k)T) dt_2
\]  

(5.27)
5.7 THE TIME-SELECTIVE CHANNEL

The slowly time-varying channel is very similar to the AWGN channel, so many researchers [18, 52, 7, 20] have used square-root Nyquist filters at the transmitter and receiver. However, in fast fading, the pulse distortion due to the multiplicative channel leads to non-root-Nyquist pulses, ISI, and thus a BER floor. There is a rich literature on this topic, but it ignores the principles of statistical detection theory and sufficient statistics.

By specialising the EMLSE-tf diversity receiver, the proper matched filter for the t-channel is deduced (MF-t). The received pulse is given by

$$\tilde{c}(t - iT, iT) = h(t - iT - t_0)z(t) \exp(j\phi)$$  (5.28)

Then the matched filter computes

$$m_r = \int_{-\infty}^{\infty} y(t_1)\tilde{z}(t_1) \exp(-j\phi)\overline{h}(t_1 - iT - t_0)dt_1$$

$$= \left( y(t)\tilde{z}(t) \exp(-j\phi) \right) \overline{h}(-t - t_0)_{t = iT}$$  (5.29)

Thus a set of sufficient statistics for the t-channel in white noise is calculated by first multiplying the received signal by the conjugated multiplicative fading process and carrier phase, and second by filtering the signal with the time-reversed, conjugated transmitter pulse shape. The first step distinguishes the MF-t from the conventional matched filter used in the AWGN channel.

The fast fading process distorts the transmitted pulse along its length. The MF-t exploits all the information along the pulse’s length by aligning the received pulse’s phase, weighting the pulse according to the depth of the fade, then adding the pulse at each time instant according to the pulse strength. In this way, it behaves similarly to the MRC receiver structure, where the diversity is in time rather than across threads.

The ISI is time-varying over $L$ symbol periods,

$$s_{r,kr} = \int_{-\infty}^{\infty} h(t_1 - iT - t_0)h(t_1 - kT - t_0)\tilde{z}(t_1)^2 dt_1$$  (5.30)
5.8 THE AWGN CHANNEL

The optimal receiver structure for the AWGN channel is well known. A square-root Nyquist pulse is transmitted, and it is processed at the receiver by a filter with an impulse response equal to the time-reversed, complex-conjugated transmitter pulse shape. This result can also be derived as a special case of the EMLSE-tf diversity receiver. The received pulse equals

$$\tilde{c}(t - iT, iT) = h(t - iT - t_o) \exp(j\phi)$$

and the matched filter computes

$$m_n = \int_{-\infty}^{\infty} y(t) \exp(-j\phi) \overline{h}(t - iT - t_o) dt$$

$$= \left( y(t) \exp(-j\phi) \right) \overline{h}(-t - t_o)$$

There is no ISI,

$$s_{n,n'} = \int_{-\infty}^{\infty} \overline{h}(t - iT) h(t - kT) dt$$

$$= \delta_{nn'} \int |h(t)|^2 dt$$

so it is straightforward to transform the branch metric into a Euclidean distance,

$$\lambda_n(\overline{\beta}, \eta) = -|m_n|^2 + \left( 2 \Re \{ \overline{\beta}_n m_n \} - |\overline{\beta}_{n,n'}|^2 \right) s_{n,n'}$$

$$= -|m_n - \overline{\beta}_{n,n'} s_{n,n'}|^2$$

using the same steps as in equation (5.24). The Viterbi algorithm reduces to a hard decision device.

5.9 SHORT PULSES

The final special case applies when the delay spread pulses are only one symbol long, even allowing for timing uncertainty. This situation does not arise in mobile communications, where bandwidth efficient pulses are used almost exclusively. However, wired links are not constrained in the same way, and
bandwidth inefficient modulation schemes such as Return to Zero (RTZ) coding are employed.

There can be no ISI, the branch metric reduces to a Euclidean distance,
\[ \lambda_n(y, \beta, \eta) = -|m_\eta|^2 + \left(2 \Re \left( \frac{\beta^{\*} m_\eta}{|s_{r,ir}|} \right) s_{r,ir} \right)^2 \]
(5.35)
and the receiver makes hard, symbol-by-symbol decisions.

For this case, and for the cases described by equations (5.24) and (5.34), the receiver is linear, so exact BERs can be easily calculated. The EMLSE-tf diversity receiver's decisions have the same geometric interpretation as the decisions of a receiver for the AWGN channel, except that the constellation points and decision boundaries vary with the depth of the fade, according to \( s_{r,ir} \).

### 5.10 THE DISCRETE-TIME EMLSE-tf

In the preceding sections, a continuous time signal representation has been used, whereas most modern receivers are digital and require sampled-time metrics. The more natural and intuitive continuous time result is pursued initially, since the underlying signals are themselves continuous. More importantly, it is dangerously easy to neglect the effect of the IF filter or to undersample the received signal.

The sampling rate of \( 1/T_r \) is chosen high enough that the \( \text{rect}(f T_r) \) anti-aliasing filter introduces negligible signal distortion. Then the samples taken at \( t = l T_r \) obey \( y_l^d = y^d(lT_r) \) and \( c_{r,ir}^d = c^d((l-ir)T_r, irT_r) \). The integrals of equation (5.14) can be replaced by summations, resulting in

\[
\begin{align*}
    m_\eta &= \sum_{d=1}^{D} \sum_{l=0}^{[s_{r,kr}] - 1} y_l^d \bar{c}_{r,ir}^d \\
    s_{r,kr} &= \sum_{d=1}^{D} \sum_{l=0}^{[s_{r,kr}] - 1} \bar{c}_{r,ir}^d \bar{c}_{r,kr}^d
\end{align*}
\]
(5.36)
where the factor of \( T_r \) is neglected, since the metrics are used for comparison only.
5.11 PILOT TONES AND PILOT SYMBOLS

In the AWGN and frequency-selective channels, it makes little sense to send pilot information. The channel changes so slowly that it can be tracked; that is, past CSI applies to the future with little or no modification. The tracking algorithm is really performing prediction: by using a very long history, it applies past information to the future. In a channel with large time variations, this philosophy breaks down. The channel is varying so swiftly that only recent CSI conveys any information. In fast fading channels, the prediction algorithm weights the freshest channel snapshot so highly that there is noise enhancement. Therefore we consider pilot information, since in fast fading, the power penalty due to pilot information is normally compensated by its higher quality. The application of pilot tones and pilot symbols to frequency-flat channels has already been discussed. Here, we propose a way of extending these ideas to the general channel.

The Time-Variant Transfer Function, \( T(f,t) \), equation (3.12), covers all time and all frequencies. For mobile communications, the frequency range is limited, but the channel may be used for communication over an arbitrarily long time interval. Therefore we are interested in characterising \( T(f,t) \) over the infinite strip \( [-\infty; \infty] \times [-\infty; \infty] \), where \( f \) is the maximum signal frequency.

A pilot tone characterises \( T(f,t) \) at some frequency, \( f_{pt} \), for all time. That is, it offers a noisy estimate of \( T(f_{pt},t) \). A pilot symbol characterises \( T(f,t) \) at some time, \( t_{ps} \), for \( [f_s; f_d] \). That is, it offers a noisy estimate of \( T(f,t_{ps})H(f) \), where \( H(f) \leftrightarrow h(\xi) \) is the Fourier transform of the transmitted pulse. These ideas can be represented graphically, as in figure 5.2, where a comb of pilot tones and a sequence of pilot symbols are considered. The channel response presented by \( T(f,t) \) can be characterised by a grid of sampling points, taken at least as fast as the Nyquist rate for each dimension. From Figure 5.2, either a comb of pilot tones or a sequence of pilot symbols can satisfy this requirement, since both provide a superset of a grid. The full picture is obtained by interpolation of the pilot information in the appropriate dimension. The Nyquist rate for the time and frequency dimensions is governed by the maximum Doppler spread, \( f_D \), and delay spread, \( \tau \).
The EMLSE-\textit{tf} diversity receiver derivation is conditioned on the channel, so the channel must be estimated first. Therefore the pilot tones and symbols must be separated from the data-bearing signal. If this is not undertaken, joint data and channel estimation is required. It is considerably more complex and requires a different approach altogether. In fact the next chapter's receiver structure can be used for joint data and channel estimation when pilot symbols are transmitted.
Pilot tones are frequency multiplexed with the signal, so that even after Doppler spreading, tones can be extracted independently from the signal by bandpass filters. The transmitter pulse shape introduces suitable spectral nulls around the pilot tone frequencies, and cannot easily be square-root Nyquist. Since the pilot tones only sample the Time-Varying Transfer Function, $T(f,t)$, in frequency over a finite frequency range, they must be spaced much closer than the frequency variation's Nyquist rate (i.e. $1/\tau$ pilot tones per Hz).

In the frequency domain, the transmitter pulse spectrum can be viewed either as $x$ evenly spaced frequency lobes, or one main lobe minus $x-1$ nulls, as in figure 5.3. The nulls must be at least $4f_D$ Hz wide to accommodate Doppler spreading. For channels exhibiting only gain-slope, two tones are sufficient (one at each band edge), and no nulls are needed.

We do not advocate using more than two pilot tones with this technique. The narrow bandwidth of the nulls means they widen the total transmitter pulse shape in time. The receiver has many states. Furthermore, this approach is wasteful of bandwidth, since every frequency lobe is modulated with the same information.

A better scheme individually modulates the $x$ frequency lobes, as a multiple subcarrier scheme. Pilot tones are placed much more closely than $1/\tau$, and the subcarriers are positioned between. The CSI is communally calculated by interpolating the pilot tones, but each subcarrier is individually detected. Reference [65] describes a related idea.

The dual of the pilot tone is roughly the pilot symbol. Pilot symbols must be time multiplexed with the signal symbols, so that even after the delay spread, the pilot and signal symbol symbols can be extracted independently. For full Nyquist pulses at the transmitter in the frequency-flat fading channel, this applies automatically.

Where there is delay spread, a pilot symbol should be embedded in a period of no signal transmission, in the same way that a pilot tone is embedded in a signal null. This time isolation ensures the delay-spread pilot symbol information can be extracted conveniently. However, the pilot tone is a delta function in frequency, whereas the pilot symbol's received pulse is non-zero for several symbol intervals. Many symbol
slots must be wasted around pilot symbols for easy CSI estimation (in the next chapter, the need for time-isolation is removed).

Neither pilot tones nor pilot symbols can deal with fast fading and a large delay spread. Pilot tones must be spaced at least as close as $1/\tau$, and are Doppler spread to $2f_D$ Hz. Therefore all the spectrum is occupied by pilot information if

$$f_D \tau \geq \frac{1}{2} \quad \text{(5.37)}$$

Even if the pulse shape $\delta(t)$ is employed, pilot symbols suffer the same problem. They must be spaced closer than $1/2f_D$ sec apart, yet the pulse is smeared over $\tau$.

This is the overspread problem of equation (3.68). It reveals one significant deficiency with the EMLSE-t/f diversity receiver. In overspread and nearly overspread channels, it becomes impossible to estimate the channel. The assumption of ideal CSI is unrealistic, and the BERs achieved by the EMLSE-t/f diversity receiver cannot be achieved by any real system. The EMLSE-t/f diversity receiver is unsuitable as a benchmark for overspread and nearly overspread channels.

5.12 SUMMARY

An MLSE receiver for the time- and frequency-selective channel with diversity has been derived, assuming ideal CSI and linear signal constellations. The general receiver structure has been specialised by considering more restricted channel models. In many cases, these receiver structures are already known.

The MLSE receiver structure provides insight into matched filtering for more general channel models than are normally considered. The matched filter for the frequency-selective channel (MF-f) is well known, and has been used inappropriately in time-varying channels. The correct structure has been deduced from the MLSE receiver structure derivation, as the MF-t/f and the MF-t.

The receiver can operate satisfactorily using non-ideal CSI, but it is not strictly speaking optimal. A method to estimate the channel has been presented, through either a comb of pilot tones, or a sequence of pilot symbols.
6 PERFORMANCE EVALUATION OF THE EXTENDED MLSE RECEIVER

6.1 INTRODUCTION

In this chapter, the BER performance of the EMLSE-tf diversity receiver is analysed. The analysis is applied to a range of transmitter, channel and receiver parameters. A series of figures presents the parameters’ influence on the receiver’s BER.

6.2 PERFORMANCE EVALUATION

The EMLSE-tf diversity receiver’s BER is a benchmark for the achievable BER rate at moderate and high SNRs, for all channel pdfs. However only channels with circularly symmetric complex Gaussian statistics can be easily analysed. In this section, a fast Rayleigh fading, frequency-selective channel in white noise is considered, using the bounding technique of section 4.3. The techniques can be extended to coded transmission and Ricean fading channels. Even without channel coding, the channel’s Doppler and delay spread lead to implicit diversity and an improved BER.

Since the channel is known, coherent detection is possible, and the data symbols, $\alpha_n$, are mapped through a Gray code to the complex phasors, $\beta_n$. Cycle slips cannot occur, so an erroneous sequence can be written as $\{\beta^{y,v} + \epsilon^{y,v}\}$,

$$\{\ldots \beta^{y,v}_{2r}, \beta^{y,v}_{1r}, \beta^{y,v}_{0} + \epsilon^{y,v}_{0}, \ldots, \beta^{y,v}_{(n-1)r} + \epsilon^{y,v}_{(n-1)r}, \beta^{y,v}_{nr} \ldots\}$$  \hspace{1cm} (6.1)

where without loss of generality, the beginning of the error event is aligned at $i = 0$. If too many symbols in a row are detected correctly within an error event, the receiver returns to the correct state, and so there are in fact two distinct error events.

With some manipulation, the pairwise probability of error for each error sequence, $P(\beta^{y,v} \rightarrow \beta^{y,v'})$, can be calculated from equation (5.11) in the form...
Employing equation (5.36) in equation (6.2), the pairwise probability of error then reduces to

\[
P(\beta^{u,v} \rightarrow \beta^{u,v,w}) = \left\{ \begin{array}{c}
2 \sum_{d=1}^{D} \sum_{i=0}^{n-1} \sum_{l=0}^{r_{ir}} \sum_{k=\max\{0,r_{kr}\}} \left( \sum_{d=1}^{D} \sum_{i=0}^{n-1} \sum_{l=0}^{r_{ir}} \sum_{k=\max\{0,r_{kr}\}} -P(\beta^{u,v} \rightarrow \beta^{u,v,w})
\end{array} \right) > 0 \quad (6.3)
\]

The left hand side of the inequality is a Gaussian quadratic form, so the probability it exceeds zero can be calculated easily. For other time-varying channel models, a joint pdf is needed for a quadratic form in the complex Gaussian noise and the correlated, non-Gaussian channel. This analysis is not pursued in this thesis.

The pairwise probability of error depends explicitly on the transmitted phasors, \( \{\beta_{[l+1]}^{u,v}, \ldots, \beta_{[r+1]}^{u,v}\} \), and implicitly on the hypothesised phasors, \( \{\beta_{[l]}^{u,v,w}, \ldots, \beta_{[r]}^{u,v,w}\} \), through the error phasors \( \{v_{u,v}^{w}\} \). An upper bound on the receiver's BER can be deduced from a union bound of error events, as

\[
BER \leq \sum_{n=1}^{M_{-1}^{L-1}} \sum_{v=4}^{M_{-1}^{L-1}} \sum_{\beta_{[l]}^{u,v}, \beta_{[r]}^{u,v,w} : k_{2u,k} \neq 0, \beta_{[r]}^{u,v,w} \neq 0, \beta_{[r]}^{u,v,w} \neq 0} \frac{P(\beta^{u,v} \rightarrow \beta^{u,v,w}) e(\beta^{u,v} \rightarrow \beta^{u,v,w})}{\log_2 M} \quad (6.4)
\]
which is the union over all error events between all transmitted sequences and all possible error sequences, weighted by their transmission probability,

\[ P(\beta^{u,r}) = \frac{1}{M^{u+2L-1}}, \tag{6.5} \]

the pairwise error probability, and the number of bit errors they introduce. Error events can begin in any symbol interval, during which \( \log_2 M \) bits are sent.

In practice, only the dominant short error events are considered, where up to \( E \) symbols may be in error. The range of the first summation in equation (6.4) is \( u = 1, \ldots, E \). If the fading is fast and the SNR high, then the dominant error events are short, so the truncated bound can validly neglect long error events. However, with slow fading, most errors occur in the deep fades, when the instantaneous SNR is very low. If the fading is very slow then the deep fade lasts hundreds of symbols at reasonable SNRs, and so do the error events. Thus the upper bound is tight and easily calculated only for fast fading and high SNR. The union of \( E = 1 \) error events due to the nearest neighbours of the errored symbol is asymptotically correct at high SNRs.

In slow, frequency-flat fading, a different technique is adopted. The fading is approximately constant over the error event, so the BER can be computed as a function of the fading amplitude. The mean BER is then this BER weighted according to the fading amplitude pdf. Accordingly, this technique can be applied to channel pdfs other than complex Gaussian.

For non-ideal CSI, the probability depends on the sequence of \( u+2L-1 \) transmitted symbols, and \( u \) errored symbols. The BER bound is tedious to calculate, since it requires \( \sum_{u=1}^{E} M^{u+2L-1}(M-1)^u \) pairwise probabilities. For ideal CSI, equation (6.3) reduces to

\[ P(\beta^{u,r} \rightarrow \beta^{u,r,w}) = P \left( \sum_{d=1}^{D} \sum_{w=0}^{w+1} \sum_{l=0}^{l+1} \left( \sum_{t=0}^{t+1} \sum_{j=t+1}^{j+1} \sum_{k=0}^{k+1} \sum_{l=0}^{l+1} \sum_{m=0}^{m+1} \sum_{n=0}^{n+1} \sum_{v=0}^{v+1} \sum_{w=0}^{w+1} \sum_{y=0}^{y+1} \sum_{z=0}^{z+1} \right) \right) > 0 \tag{6.6} \]
which depends implicitly on the transmitted and hypothesised symbol sequences, 
\[ \{ \beta_{0}, ..., \beta_{M-1} \} \] and \[ \{ \beta_{0}, ..., \beta_{M-1} \} \], through the allowed values of \( \{ e^{v,N} \} \). An upper bound on the receiver’s BER can be deduced from a union bound of error events, as

\[
\text{BER} \leq \sum_{n=1}^{E} \sum_{\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0}} \frac{p(\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0}) e(\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0})}{\log_{2} M} \tag{6.7}
\]

where the transmitted sequences’ transmission probabilities equal

\[
P(\beta_{0}^{n}) = \frac{1}{M^{n}} \tag{6.8}
\]

Neglecting constellation-specific symmetries, only \( \sum_{n=1}^{E} M^{n}(M-1)^{W} \) pairwise probabilities must be calculated. When geometrically uniform constellations are employed, such as M-PSK, the BER bound need not be averaged across all transmitted sequences; it is sufficient to consider merely one sequence (the “all zeros” sequence). Then

\[
\text{BER} \leq \sum_{n=1}^{E} \sum_{\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0}} \frac{p(\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0}) e(\beta_{0}^{n} \rightarrow \beta_{0}^{\text{w} \cdot 0})}{\log_{2} M} \tag{6.9}
\]

The pairwise probabilities of equations (6.3) and (6.6) are now computed. Define \( g \) as a vector of the germane circularly symmetric complex Gaussian random variables,

\[
g = \begin{bmatrix} c \\ n \\ e \end{bmatrix} \tag{6.10}
\]

where \( c, n \) and \( e \) are vectors of the received pulse samples, \( c_{i-j,r,j}, \) noise samples, \( n_{i}^{d}, \) and received pulse error samples, \( e_{i-j,r,j}, \) from equation (6.3) or (6.6), respectively.

For non-ideal CSI, \( c, n \) and \( e \) are \( D(L+u-1)L \times 1 \), \( D(L+u-1)N \times 1 \) and \( D(L+u-1)L \times 1 \) column vectors, respectively. They are sorted by \( d, l \) and \( i \) in some
computationally convenient fashion. A suitable sorting function for both \( c^d_{u-r,p} \) and \( e^d_{u-r,p} \) is
\[
x = d \left( (L + u - 1)Lr \right) + IL - \left\lfloor \frac{l}{r} \right\rfloor + i + L - 1,
\]
where the inverse sorting functions are
\[
i = \left( x \mod ((L + u - 1)Lr) \right) \mod L + \left\lfloor \frac{x \mod ((L + u - 1)Lr)}{L} \right\rfloor / r - L + 1,
\]
\[
d = \left\lfloor \frac{x}{(L + u - 1)Lr} \right\rfloor
text{and}
\[
l = \left\lfloor \frac{x \mod ((L + u - 1)Lr)}{L} \right\rfloor / L.
\]

For ideal CSI, \( c \) is a \( DuLr \times 1 \) column vector and \( e \) is empty. A suitable sorting function for both \( c^d_{u-r,p} \) and \( e^d_{u-r,p} \) is
\[
x = duLr + l + (L - 1)r, \text{ where the inverse sorting functions are}
\]
\[
d = \frac{x}{(uLr)},
\]
\[
l = \left\lfloor \frac{x \mod (uLr)}{L} \right\rfloor + \left\lfloor \frac{x \mod (uLr)}{L} \right\rfloor / L.
\]

The additive noise samples, \( n^d_i \), are always sorted by thread and then chronologically.

Define \( \kappa^{u,v,w} \) as the left-hand-side of the inequality in equation (6.3) or (6.6), so it can be written explicitly as a Gaussian quadratic form,
\[
\kappa^{u,v,w} = g^H G^{u,v,w} g \tag{6.11}
\]
where the kernel, \( G^{u,v,w} \), is a Hermitian symmetric matrix, defined by equation (6.3) or (6.6). The covariance matrix of \( g \), \( \mathbf{R}_{gg} \), is given by
\[
\mathbf{R}_{gg} = \frac{1}{2} E \left( gg^H \right) = \begin{bmatrix}
\mathbf{R}_{cc} & 0 & 0 \\
0 & \mathbf{R}_{nn} & 0 \\
0 & 0 & \mathbf{R}_{ee}
\end{bmatrix} \tag{6.12}
\]

For ideal CSI, \( \mathbf{R}_{ee} \) equals zero and all the power is allocated to the data-bearing signal. Accordingly, entries in the received pulse and noise autocorrelation matrices are given by equations (4.32) and (4.33).

When pilot tones are used to estimate CSI, they are isolated from the data-bearing signal by bandpass filters. The data-bearing signal is not filtered, since the transmitter pulse shape is constructed so that the Doppler spread signal and pilot tones do not overlap; this is left to the MF-if. Therefore the additive noise is also white when frequency-isolated pilot tones are transmitted.
The tones are assumed to be spaced closely enough for perfect CSI reconstruction [2]. The estimation error then comes from the noise within the tones' bandwidths. $R_{ee}$ depends on the number and spacing of the pilot tones. Here two tones are cotransmitted with the baseband signal, $\exp(\pm 2\pi f_{pt}t)$. The presence of pilot tones placed symmetrically around the carrier frequency can be exploited by the receiver to recover the carrier, up to a $\phi = 0^\circ, 180^\circ$ phase ambiguity. The pilot tones are recovered ideally at an intermediate frequency using filters with $\text{sinc}(.)$ impulse responses. The data-bearing signal and two tones are translated separately to baseband. The tones are labelled by

$$\hat{T}^d(\pm f_{pt}, t) = \frac{1}{\sqrt{2}} T^d(\pm f_{pt}, t) + N^d(\pm f_{pt}, t)$$

where $T^d(\pm f_{pt}, t)$ is the $d$th thread's time-variant transfer function, evaluated at $f = \pm f_{pt}$. The power is evenly distributed between the two tones. $N^d(\pm f_{pt}, t)$ is the noise from the same frequency bands as the pilot tones as a function of time. The CSI is reconstructed from the tones in a polynomial expansion, as

$$\hat{z}^d(t, \xi) \approx \frac{\hat{T}^d(f_{pt}, t) + \hat{T}^d(-f_{pt}, t)}{\sqrt{2}} \delta(\xi) + \frac{\hat{T}^d(f_{pt}, t) - \hat{T}^d(-f_{pt}, t)}{2\sqrt{2}\pi f_{pt}} \delta'(\xi)$$

where $\delta'(\cdot)$ is the derivative of the Dirac delta function. It acts as the derivative operator under integration. This arrangement uses the $d$th pilot tone for CSI in the $d$th thread, which is optimal only when the antennae receive independently faded signals. When the channel process is correlated between threads, then each thread's CSI estimate should be a MMSE linear combination of all $D$ received pilot tones.

However, the performance gain does not warrant the additional complexity since the threads are approximately independent by design (proper antenna spacing for space diversity; widely separated carrier frequencies for frequency diversity; sufficiently delayed signals for time diversity).

Assuming perfect carrier recovery, $f_{0,\text{nom}}$ is 0Hz and the estimation error in the received pulse, equation (5.18), equals
6 PERFORMANCE EVALUATION OF THE EXTENDED MLSE RECEIVER

\[ e_{t-\nu,rr} = \frac{N_d\left(f_{\nu}, t-t_0\right) + N_d\left(-f_{\nu}, t-t_0\right)}{\sqrt{2}} h\left(t-iT-t_0\right)e^{j\phi} + \frac{N_d\left(f_{\nu}, t-t_0\right) - N_d\left(-f_{\nu}, t-t_0\right)}{2\sqrt{2}\pi f_{\nu}} h'\left(t-iT-t_0\right)e^{j\phi} \] (6.15)

and

\[ \frac{1}{2} E\left(e_{t-\nu,rr}^* e_{t-\nu,rr}^* \right) = \begin{cases} \left(\frac{h'((l-ir)T'_{\nu}-t_0)\bar{h}'((m-kr)T'_{\nu}-t_0)}{4\pi^2 f_{\nu}^2}\right) d_1 = d_2 & \text{otherwise} \\ 0 & \text{otherwise} \end{cases} \] (6.16)

A fraction of the transmitted power, \( p_z \), \( 0 \leq p_z < 1 \), is allocated to the pilot tones. This scales the SNR by the ratio \( 1/(1-p_z) \); it scales the received pulse autocorrelation by \( 1-p_z \); and the noise spectral density within the pilot tone band by \( (1-p_z)/p_z \).

With these definitions, \( R_{gg} \) is completely described. The characteristic function of a Gaussian quadratic form, \( K^{u,v,w} \), is given by

\[ P_{\xi}^{u,v,w}(\zeta) = \frac{1}{\text{det}\left|I - j2\zeta R_{gg} G_{u,v,w}^*\right|} \] (6.17)

From Appendix B, the pairwise probability of error can be calculated as

\[ P(\beta^{x,y} \rightarrow \beta^{x',y}) = 1 - \sum_{i,z([p]) \neq 0} \prod_{k=1}^{i} \left(1 - \frac{1}{p_i/p_k}\right) \] (6.18)

where \( p_i \) is the \( i \)th pole of equation (6.17). This equation assumes that the poles of equation (6.17) are simple; this is satisfied unless the diversity branches are ideally independent. This more complex case is also considered in Appendix B.
6.3 A SIMPLE ANALYTIC EXAMPLE

The previous analysis appears complicated, so one simple case is presented here to demonstrate that the underlying mathematics is straightforward. The transmitter parameters are BPSK \( \{\beta\} \in \pm 1 \), and a rectangular pulse of \( H = 1 \) symbol periods. A frequency-flat channel without carrier or timing offset is assumed, so that there is only \( P = 1 \) path, \( \tau / T = 0 \), and the received pulse is \( L = 1 \) symbol period long. The Bessel function autocorrelation model of equation (3.40) is used, with \( f_0T \) open. The receiver takes \( r = 1 \) sample per symbol and has ideal CSI. The receiver has \( M^{L-1} = 1 \) state, so it makes hard decisions. The only error event is one symbol long, as

\[
\{ \beta^{x_r} \} = \{ \beta_{-2}, \beta_{-1}, \beta_0, \beta_1, \ldots \}
\]

so \( u = 1 \). The BER expression of equation (6.9) simplifies to

\[
BER = \frac{1}{2} \sum_{r=1}^{3} \sum_{\text{dominant error events}} P\left( \beta^{x_r} \rightarrow \beta^{x_{r,r}} \right) = P\left( \beta^{x_r} \rightarrow \beta^{x_{r,r}} \right)
\]

since \( P\left( \beta^{x_r} \right) = \frac{1}{2} ; \ e\left( \beta^{x_r} \rightarrow \beta^{x_{r,r}} \right) = 1 ; \ \log_2 M = 1 ; \) and the pairwise probability of \( 0^\circ \) being mistaken for \( 180^\circ \) matches the pairwise probability of \( 180^\circ \) being mistaken for \( 0^\circ \). Considering the latter case, \( e_{ir} = \{ ... , 0, -2, 0, ... \} \).

From equation (4.34), the received pulse autocorrelation equals

\[
\frac{1}{2} E\left( c_{0,0} \bar{c}_{0,0} \right) = E_b
\]

and the noise autocorrelation is given by equation 4.33. Then

\[
R_{xx} = \begin{bmatrix} E_b & 0 \\ 0 & \frac{\sigma_n^2}{2} \end{bmatrix}
\]

The pairwise probability of error equals

\[
P\left( \beta^{x_{r,r}} \rightarrow \beta^{x_{r,r}} \right) = P\left( \left( \bar{e}_o^{x_{r,r}} n_o \bar{c}_{0,0} + e_o^{x_{r,r}} \bar{n}_o \bar{c}_{0,0} \right) - \bar{e}_o^{x_{r,r}} e_o^{x_{r,r}} \bar{c}_{0,0} \bar{c}_{0,0} > 0 \right)
\]

\[
= P\left( -2n_o \bar{c}_{0,0} - 2\bar{n}_o c_{0,0} - 4\bar{c}_{0,0} c_{0,0} > 0 \right)
\]

so
The poles of equation (8.8) obey

\[ \det \begin{bmatrix} -2 \varphi \frac{2E_b}{N_0 T} & -2E_b \\ -2 \frac{N_0}{T} & 0 \end{bmatrix} = 0 \] (6.25)

so that

\[ p_1 = \frac{j}{4N_0 / T} \left( -1 + \sqrt{1 + \frac{E_b}{N_0 / T}} \right), \quad p_2 = \frac{j}{4N_0 / T} \left( -1 - \sqrt{1 + \frac{E_b}{N_0 / T}} \right) \] (6.26)

The first pole is positive imaginary and the second pole is negative imaginary.

The pairwise probability of error given by equation (6.18) equals

\[ P(\beta^{x,y} \rightarrow \beta^{x',y'}) = 1 + \frac{1}{(1 - p_1 / p_2)} = \frac{p_1}{p_1 - p_2} = \frac{1}{2 \left( \frac{E_b}{N_0 / T} + 1 \right) + 2 \left( \frac{E_b}{N_0 / T} \right)^2 + \frac{E_b}{N_0 / T}} \] (6.27)

which is also the BER, by equation (6.20). At low SNR, the BER tends to 0.5; at high SNR, the BER is inversely proportional to the SNR.

### 6.4 PERFORMANCE OVERVIEW

The following sections provide an overview of the EMLSE-tf diversity receiver's performance for a range of transmitter, channel and receiver parameters.

The communication system has many degrees of freedom. The transmitter is influenced by the constellation size, \( M \), the constellation geometry (M-PSK, M-QAM), the pulse shape, \( h(t) \), its length, \( H \), its excess bandwidth, \( \chi \), and the encoding. The channel is characterised by the fading pdf, the SNR, the fading rate, \( f_D T \), the delay spread, \( \tau / T \), the Doppler spectrum, the delay spectrum, the cross-thread correlation (for space diversity, the antenna spacing, \( W \), and orientation, \( \theta - \phi \)), the symbol timing, \( t_0 \), and the residual carrier offset, \( f_{\text{nom}} \). The MLSE predictor receiver structure is affected by the number of threads, \( D \), the number of samples per symbol, \( r \), and the decision delay of the Viterbi processor.
A comprehensive set of curves is impractical, and is not attempted. Instead a baseline set of parameters is selected, and a series of BER-SNR graphs are plotted when one parameter only is modified. Thus the parameter's influence is isolated.

The baseline transmitter uses QPSK, a square-root raised cosine pulse, $h(t)$, truncated to $H = 1.5$ symbol periods, with $\chi = 50\%$ excess bandwidth. This is realistic, except for the truncated pulse shape which is too bandwidth inefficient. This choice aligns the baseline parameters with those used in chapter 8, where the computational time is substantial for longer pulses.

There is considerable implicit delay and Doppler diversity in the baseline channel. A substantial delay spread is used, $\tau/T = 0.5$. The Doppler spread is wide, with $f_D T = 0.1$. Thus the product, $f_D \tau$, equals 0.05, or one tenth of the overspread limit. There is no frequency offset or timing error, so the received pulse occupies $L = 2$ symbol periods. The channel comprises $P = 3$ equally spaced paths with equal mean power. The standard Doppler autocorrelation model of equation (3.40) is used.

The baseline receiver has one antenna only. The receiver takes $r = 3$ samples per symbol. The anti-aliasing filter's one-sided bandwidth is $1/2T_r$. When the anti-aliasing filter's bandwidth is wide enough, it has no effect on the data-bearing signal. When the anti-aliasing filter's bandwidth is insufficient, the received pulses are lengthened by the duration of the anti-aliasing filter's impulse response. The number of ISI states in the receiver becomes large, and the receiver's analysis is unduly complicated. Since the number of samples per symbol is twice the Nyquist rate for the transmitted pulse (before truncation), it is reasonable to assume that the anti-aliasing filter does not distort the data-bearing signal. However, a large residual carrier offset can shift the received signal to the edge of the anti-aliasing filter and beyond, leading to signal distortion and an impaired BER. By neglecting the anti-aliasing filter's distortion, the influence of the residual carrier offset on the BER cannot be computed, and so no figures are plotted for this parameter.
By strictly defining the received pulse length, $L$, to include the uncertainty in the beginning of transmission, the receiver's symbol timing does not affect its BER. According, no figures are plotted for this parameter.

Except for figures 6.1 and 6.2, the analytic curves presented in the following sections are the one-symbol, nearest neighbour error events. These curves are asymptotically correct at high SNRs. Figure 6.1 demonstrates that the dominant error events are indeed the one-symbol, nearest neighbour error events in fast fading, $f_D T = 0.1$. The other curves consider many more error events; but these converge quickly on the one-symbol, nearest neighbour curve. However, this effect is less pronounced in
slower fading, as shown by figure 6.2. The union bounds do converge on the one-symbol, nearest neighbour curve, but not until a higher $E_b/N_0$. In slower fading, the mean fade duration is longer, so longer error events are more likely. Thus the union bound is loose until the mean fade duration is diminished by an increase in $E_b/N_0$.

It is noteworthy that curves computed for BPSK and QPSK and $f_cT = 0$ using only the one-symbol, nearest neighbour error events match exactly the results derived under a slow fading approximation [46]. Therefore the one-symbol, nearest neighbour error BER curves are likely to be tight BER estimates at both low and high $E_b/N_0$.

Figure 6.3: BER-SNR curves for various signal constellations. The remaining parameters take their baseline values.

Figure 6.4: BER-SNR curves for different pulse lengths, $H$. The remaining parameters take their baseline values.
6.5 TRANSMITTER PARAMETERS

In mobile communications, the available spectrum is limited, so there is considerable interest in the performance of spectrally efficient modulation schemes. Employing linear modulations and large constellations is an effective scheme. In figure 6.3, the influence of the constellation on the BER is depicted. The two most spectrally inefficient schemes, BPSK and QPSK have identical performance. 8-PSK and 16-QAM have similar BER curves, but there is a 3dB penalty for their high spectral efficiency. For 64-QAM, there is an additional 3dB penalty. Accordingly, the larger constellations are reasonable, but there is a power penalty.

The baseline transmitted pulse is a square root raised cosine pulse with $\chi = 50\%$ excess bandwidth, truncated to $H = 1.5$ symbol periods. However, the truncation is not particularly realistic in practice, since the short pulse's spectrum has unreasonably high sidelobes. The choice of a short pulse is consistent with the baseline parameters used in chapter 8, where the analysis' complexity is large for longer pulses. From figure 6.4, it is clear that BER curve for $H = 1.5$ symbol periods is close to the curves for $H > 1.5$. The BER curve for $H = 0.5$ symbol periods exhibits different behaviour. The delay spread received pulses do not overlap, so they are easier to distinguish. Accordingly, the BER for this case is lower.
In figure 6.5, the sensitivity of the BER curves to the pulse's excess bandwidth is examined. The pulses are truncated so severely that they have similar shapes, whatever the excess bandwidth. This is reflected in the BER curves, which are all very similar.

6.6 CHANNEL PARAMETERS

The most interesting aspect of the EMLSE-\(f\) diversity receiver is its ability to exploit the implicit diversity from the Doppler and delay spread. In figure 6.6, the fading is slow, so only implicit delay diversity is available. As the normalised delay spread, \(\tau/T\), increases, the BER curves steepen earlier. For \(\tau/T = 0.05\), the second diversity order appears at approximately 30dB, whereas the second diversity order appears at approximately 10dB for the larger delay spread, \(\tau/T = 0.5\). Figure 6.7 is almost identical to figure 6.6, yet now it is the delay spread that is small and the Doppler spread is varied. The implicit Doppler spread does not steepen the BER curve's gradient (it provides no diversity) until the corresponding eigenvalue exceeds the noise power.

![Graph](image-url)
Figures 6.8 and 6.9 also reflect similar properties. The faster fading and longer delay spread steepen the BER curve's gradient. Doppler and delay spread can be accommodated simultaneously.

Figure 6.10 examines the BER as a function of symbol rate, assuming the cell radii, maximum receiver speed and carrier frequency are fixed (i.e. $f_0 T$ is a constant). The performance is especially good for very low ($f_0 T = 0.1$, $\nu T = 0.005$) or very high symbol rates ($f_0 T = 0.001$, $\nu T = 0.5$), corresponding to high implicit Doppler or delay diversity. Systems can explicitly exploit this feature. By sending signals on many low rate sub-carriers (whose symbol rate is approximately equal to or below the Doppler
spread), in principle the implicit Doppler diversity can be utilised. Similarly, signals can be time-division multiplexed together into a very high rate signal (whose symbol period is approximately equal to or above the Doppler spread) to exploit the implicit delay diversity, as in the RAKE receiver [45].

The baseline set of parameters has $P = 3$ independently faded, uniformly delayed paths with equal mean power. From figure 6.11, this choice has little effect on the BER. The $P = 1$ curve shows a substantially inferior BER; since the channel has only one tap, it is frequency-flat and there is no implicit delay diversity.

![Figure 6.9: BER-SNR curves for increasing Doppler spread, $f_DT$. The remaining parameters take their baseline values ($\tau/T = 0.5$).](image1)

![Figure 6.10: BER-SNR curves for various delay and Doppler spreads. The Doppler-delay product is kept constant, $f_DT = 0.0005$. The remaining parameters take their baseline values.](image2)
6.7 RECEIVER PARAMETERS

Since no IF filtering is performed, the sampling rate has little influence on the BER, as can be seen in figure 6.12. However, for $r = 1$, there are only two samples of the received pulse. At best these are independently faded, so the receiver cannot achieve more than second order diversity. Accordingly at high $E_b/N_0$, the $r = 1$ BER curve drops one hundred times for every 10dB increase in $E_b/N_0$; whereas the other curves are steeper.

The EMLSE-If receiver's BER performance is poor when there is no implicit delay or Doppler diversity: the BER is inversely proportional to the $E_b/N_0$. The constant of proportionality depends on the signal constellation. Accordingly, very low BERs can only be achieved at practical SNRs through explicit diversity (including coding). Figure 6.13 plots the improvement available from multiple antennae, arranged in a line and spaced at quarter-wavelength intervals. The second antenna provides the greatest $E_b/N_0$ reduction for a given BER. Further antennae offer improvements, but the returns diminish with increasing $D$. This illustrates that receiver space diversity is a powerful method to accomplish power efficiency.

![Figure 6.11: BER-SNR curves when the number of independently faded paths, $P$, varies. The remaining parameters take their baseline values.](image)
Receiver space diversity has two practical disadvantages. The receiver requires $D$ antennae and $D$ RF chains, and the volume occupied by the $D$ antennae can be prohibitive in mobiles. Figure 6.14 addresses the consequences of sub-optimal antenna spacing. $W$ wavelengths separate the receiver's two antennae. The gain is substantial even when the separation is only 0.01 wavelengths. The rate of improvement reduces as the antenna separation nears the autocorrelation function's first zero, $W = 0.383$. The BER for $W = 0.25$ is almost as good as at $W = 0.383$. 

Figure 6.12: BER-SNR curves for different sampling rates, $r$. The remaining parameters take their baseline values.

Figure 6.13: BER-SNR curves for increasing diversity threads, $D$. $W = 0.25, \phi = 0^\circ$. The remaining parameters take their baseline values.
6.8 PILOT TONES

Channel state information can be obtained through pilot tones. There is a performance loss due to the power allocated to the pilot tones, the additive noise present with the pilot tones at the receiver, and the interpolation error from estimating the channel by a polynomial expansion in frequency. Since the pilot tones are assumed to be spaced sufficiently closely that the latter effect is negligible, it is the first two sources of degradation that are considered here. The transmitter pulse,

$$h(t) = \begin{cases} \frac{2}{3(0.95)} \left(1 - \cos \left(2\pi \frac{t}{0.95T}\right)\right) & 0 \leq t < 0.95T \\ 0 & \text{otherwise} \end{cases}$$

is used, which is short enough to avoid ISI for the $\tau/T = 0.05$. The pilot tones are located at $f_{pr} = \pm 5/T$. The power fraction, $\rho_p = \sqrt{2f_D T / (1 + \sqrt{2f_D T})}$, is allocated to the pilot tones, which is asymptotically optimum for $r = 1$ sample per symbol [9]. The receiver takes $r = 10$ samples/symbol. Figure 6.15 indicates that pilot tone systems can closely approach the performance of ideal CSI, particularly for slow fading.
6.9 SUMMARY

The BER of an MLSE receiver in the Rayleigh fading time- and frequency-selective channel with diversity has been bounded, assuming ideal CSI and linear signal constellations. A series of figures has been presented, showing the sensitivity of the EMLSE-ff diversity receiver's BER to different transmitter, channel and receiver parameters. Given ideal CSI, the receiver can successfully detect large constellations over extremely fast fading and highly delay spread channels. Pilot tones provide a good approximation to ideal CSI, particularly in slow fading. Given multiple antennae, the required $E_b/N_0$ for reliable communications can be significantly reduced.
7 DESIGN OF THE MLSE PREDICTOR RECEIVER

7.1 INTRODUCTION

An MLSE receiver structure was derived and analysed assuming perfect knowledge of the synchronisation and channel parameters in chapters 5 and 6. While this approach leads to a receiver structure of some theoretical interest, it is unsuited to a practical system in the absence of pilot tones or symbols. Accordingly a fundamentally different approach must be adopted to design an MLSE receiver structure when CSI is unavailable. In particular, only channels with complex Gaussian distributions allow receivers to be easily designed.

Two avenues can be pursued. The first avenue assumes no knowledge of the instantaneous channel, but perfect knowledge of the second order statistics of the received pulse, given by equation (4.32) and noise, given by equation (4.33). The second avenue assumes no knowledge of the channel or of its second order statistics, the "blind" detection problem. The blind MLSE receiver has exponentially increasing complexity.

The first avenue forms the main content of this chapter. As discussed in section 1.6.3, the design of an MLSE receiver with known second order statistics has already been addressed by [30, 15, 61, 64]. There are several remaining research issues. First, existing work implicitly assumes ideal carrier acquisition and symbol timing. Second, the receivers' BER performances are evaluated through simulation, whereas an analytic technique is preferred in addition. Third, the received pulse and noise's second order statistics are assumed to be known perfectly, whereas an estimation scheme is required for any practical implementation. Only [15] has demonstrated a successful estimation scheme, albeit for a simpler signal model. Fourth, only Rayleigh channels have been considered, whereas a Ricean $df$-channel model is superior since it includes virtually all channels of practical interest. Fifth, the problem of code design and decoding for the MLSE receiver structure has been avoided, other than by [61] for a simpler channel model.
In this and the next chapter, the first four issues are tackled. A generalised receiver structure is designed that jointly performs MLSE for the case of an unknown carrier frequency and phase, unknown symbol timing, and an unknown slow or fast Ricean fading, frequency-selective channel, corrupted by white noise. The mild constraint is made that the carrier frequency, phase and symbol timing change relatively slowly. The receiver employs predictors to estimate the received signal, and forms a weighted Euclidean distance between the predicted and received samples. Two \textit{ad hoc} strategies are proposed for estimating the Rayleigh fading channel's second order statistics. The first method has relatively low complexity but requires a long training sequence, its tracking performance is poor, and it is potentially unstable. Accordingly it has limited application. The second method can be used within a Per-Survivor-Processing receiver structure to estimate blindly the channel's second order statistics, but the receiver is asymptotically ML only. A training sequence can be exploited to reduce significantly the receiver's complexity. The second method's complexity is high.

Since this joint receiver is the MLSE receiver structure when the received pulse and noise's second order statistics are known, it has some theoretical interest. Its BER at moderate and high SNRs is essentially a lower bound on the achievable BER of any receiver in the same environment. The complexity of the joint receiver is the same as the receiver structure of [64] when the cost of estimating the second-order statistics is ignored, but it additionally performs synchronisation tasks. The joint receiver's generalisation arises as follows:

- The carrier frequency introduces a time-varying but deterministic scale factor. The ability of the receiver in [64] to deal with fast fading channels allows it to accommodate this time-varying carrier factor.

- The carrier phase offset distorts the transmitted signal by a complex factor. The predictor receiver of [64] can accommodate complex channels already, so another complex factor is irrelevant. In designing receivers for the frequency-selective channel (f-channel), the carrier phase is often modelled as a random variable, since it appears to fluctuate randomly. More accurately, the carrier phase is roughly constant, but is modified by the multipath channel. In this thesis, this second
interpretation is adopted, so the carrier phase is not a random variable but is merely unknown \emph{a priori}.

- The timing offset shifts the delay power profile along the delay axis, creating a new delay power profile. Since the receiver of [64] can accommodate arbitrary delay spread profiles already, no extension is needed to accommodate the shift also.

As will be seen, all of these effects are taken into account by sets of complex predictor tap weights.

This receiver requires beforehand only (i) a stable symbol-rate oscillator, but this is not difficult to synthesise, as well as being virtually a requirement of any receiver; (ii) an upper bound on the intervals that the delay-spread pulses occupy (i.e. the width of the delay-spread pulses plus the maximum timing error), (iii) an upper bound on the bandwidth that the Doppler spread and shifted signal occupies, and (iv) for ML performance, ideal knowledge of the received pulse and noise's second order statistics.

\section*{7.2 TRANSMITTER DESIGN}

The proposed receiver structure uses predictors. The presence of predictors influences the transmitter design through the signal constellation and mapping of bits to symbols.

Predictors use past symbols as amplitude and phase references for subsequent symbols. Therefore a phase jump (cycle slip) affects the phases of subsequent decisions, and phase lock cannot be guaranteed between transmitter and receiver over long transmission intervals in the absence of pilot information (pilot tones or pilot symbols) when the channel is purely Rayleigh. Thus the uncoded transmission is effectively a catastrophic "code," in that there are a number of valid sequences whose path metrics are the same but whose hypothesised symbol sequences are different.

If the error involves an amplitude slip, then it is ultimately corrected when a symbol is transmitted that reveals the erroneous amplitude reference, if not before. Slips involving amplitude are nearly catastrophic in large constellations, since the
anomalous reference may not be corrected for some time, resulting in long bursts of errors.

Accordingly, we conclude that the transmitter constellation and mapping should be designed with rotational-invariance and an amplitude-slip-tolerant property, in order to help reduce the number of bit errors during error events. A novel, radially symmetric constellation was proposed in section 2.1 that supports these desirable properties.

Mapping involves three sequences. There is the data source, \( \{ \alpha \} \), and the sequence of complex phasors, \( \{ \beta \} \), taken from the constellation. The mapping from \( \{ \alpha \} \) to \( \{ \beta \} \) has two stages. The \( \log_2 P \) phase bits select a sector in a rotationally invariant manner, such as differential encoding; then the \( \log_2 A \) sector (amplitude) bits select a point from that sector, in such a way that the effect of an amplitude slip is diminished. This can be achieved easily for the proposed, radially symmetric constellation, but for \( M\)-QAM constellations, an effective solution is unclear. \( M\)-PSK has one point per sector, so amplitude-slips do not arise. The third sequence, \( \{ \gamma \} \), comprises a sequence of integers, taken from \( \{ 0, ..., M-1 \} \). \( \gamma_p \) is constructed from the \( \log_2 P \) phase bits of \( \alpha_p \) and the \( \log_2 A \) coded sector bits from \( \beta_p \). \( \{ \gamma \} \) is used by the receiver, since \( \{ \beta \} \) can only be detected up to a \( P\)-ary phase ambiguity. \( \{ \alpha \} \) can be reconstructed from \( \{ \beta \} \) or \( \{ \gamma \} \).

Using \( M\)-DPSK as an example, \( \{ \alpha \} \) are taken from \( \{ 0, ..., M-1 \} \); \( \beta_p \) obeys
\[
\beta_p = \beta_{(p-1)} \exp \left( j 2 \pi \frac{\alpha_p}{M} \right), \quad \text{where } \beta_{\lfloor p/T \rfloor} = 1; \quad \text{and} \quad \gamma_p = \alpha_p, \quad \text{since there are no amplitude variations within the constellation.}
\]

7.3 RECEIVER DERIVATION

The MLSE receiver searches all allowed phasor sequences, \( \{ \beta \} \), in the transmission interval and chooses the one with maximum likelihood. The digital receiver observes received signal samples from the \( D \) diversity threads, and orders them by time and by thread, as
It is helpful to have a single index for the received samples, so define 
\[ y_m = y_{[m/D]}^{[m\text{mod}D+1]} \]
where \( y_{D+1} = y_{D+1}^d \). The vector of received signal samples up to the \( l \)th sample and the \( d \)th thread is defined as
\[
y_l^d = \begin{bmatrix} y_{[l-1]D+1}^{D+1} & \ldots & y_{[l-1]D+1}^{D+1} \end{bmatrix}^T \]
so that \( y_{D+1}^d = [\ldots, y_l^d]^T \). In the subsequent derivation, the received signal's pdf is explicitly conditioned on the synchronisation parameters, \( \eta \).

The remaining random variables are the channel, \( z(t, \xi) \), and the noise, \( n(t) \). Accordingly the received signal vector, \( y \), conditioned on the synchronisation parameters and phasor sequence is a Ricean distributed random sequence. The channel, \( z(t, \xi) \), may be correlated between threads, and now this does influence the receiver structure.

In section 4.2, the log-likelihood for a general pdf was described, using a continuous time development. A parallel derivation is outlined here, but in discrete time. The transmission interval begins at \( I_b \) sec and ends at \( I_c \) sec. These times may be finite or infinite. The MLSE receiver requires as a metric the probability of observing the sequence of received samples, conditioned on a hypothesised phasor sequence. It can be expanded by repeated application of Bayes theorem, as
\[
P_{y|\beta, \eta}(y|\beta, \eta) = \prod_{m=1}^{[I_c/I_T]} p_{y_m|\beta, \eta_{m-1}, \eta}(y_m|\beta, \eta_{m-1}, \eta) \]

The log-likelihood is defined as
The pdf of complex Gaussian random variables is given by equation (3.48), so the log-likelihood equals

\[
\ln p_y(y|\beta, \eta) = \sum_{n=1}^{[T/F]} \ln p_{m|\beta, \eta, y_{m-1}, \eta}(y_m|\beta, \eta_{m-1}, \eta)
\]  

(7.5)

Neglecting the \(ln\) term since it is independent of the hypothesised phasor sequence and the scale factor \(-1\), the log-likelihood can be used as a recursive path metric, in the form

\[
\Lambda_r(y|\beta, \eta) = \Lambda_{(r-1)}(y|\beta, \eta) + \lambda_r(y|\beta, \eta)
\]

\[
= \sum_{m=1}^{[T/F]} \left( \frac{|y_m - E(y_m|\beta, \eta_{m-1}, \eta)|^2}{\sigma^2(y_m|\beta, \eta_{m-1}, \eta)} + \ln \sigma^2(y_m|\beta, \eta_{m-1}, \eta) \right)
\]

(7.6)

where \(E(y_m|\beta, \eta_{m-1}, \eta)\) is the expected value of \(y_m\), given a hypothesised phasor sequence, the past received samples, and the synchronisation parameters; and \(\sigma^2(y_m|\beta, \eta_{m-1}, \eta)\) is the variance of this prediction. By neglecting the \(-1\) factor, the maximum likelihood phasor sequence has the minimum metric. From Appendix A, \(E(y_m|\beta, \eta_{m-1}, \eta)\) is the MMSE prediction of \(y_m\), and

\[
\left( y_m - E(y_m|\beta, \eta_{m-1}, \eta) \right) / \sqrt{\sigma^2(y_m|\beta, \eta_{m-1}, \eta)}
\]

is the Innovations process [64]. The expectation is computed by a predictor, which takes a linear combination of past samples, as

\[
E(y_m|\beta, \eta_{m-1}, \eta) = \sum_{k=1}^{[T/F]} b_{m,k}(\beta, \eta)y_{m-k}
\]

(7.8)
where \( y_m^{[n]} \) is replaced by unity according to Appendix A; and \( b^u_{n,k} (\widehat{\beta}, \eta) \) is the \( k \)th tap for the ML predictor of \( y_m \), given the hypothesised phasor sequence and the synchronisation parameters.

However, the predictor tap weights depend on the complete history of transmitted symbols, so the receiver must perform a tree search whose complexity increases exponentially in time. A finite complexity receiver truncates the lengthening predictors to some fixed order. Define \( B \) to be the number of predictor taps per thread; so there are \( DB+1 \) predictor taps in all. \( B \) is chosen to be large enough that there is a minor BER penalty only. If the received signal conditioned on the data sequence and synchronisation parameters is autoregressive, then there is no loss of performance in truncating the predictors when \( DB \) is at least the order of the process.

The BER curves in the following chapter for Rayleigh fading show that the autoregressive model can only be an approximation: at a sufficiently high SNR, additional implicit diversity orders arise. A receiver can exploit \( DB \) of these diversity orders, but floors when the \((DB+1)\)th appears. Thus \( DB \) should tend to infinity as the SNR tends to infinity. Notwithstanding, the maximum SNR is limited in a physical receiver, so a small value of \( DB \) can be selected in practice for the maximum expected SNR, and the penalty is insignificant.

The truncated predictor tap weights are chosen to satisfy the normal equations

\[
E \left( y_m | \beta, \eta \right) = \sum_{k=1}^{DB+1} b_{n,k} (\beta, \eta) E \left( y_{m-k} | \beta, \eta \right) \tag{7.9}
\]

\[
\frac{1}{2} E \left( \overline{y}_{m-p} y_{m-p} | \beta, \eta \right) = \sum_{k=1}^{DB+1} b_{n,k} (\beta, \eta) \frac{1}{2} E \left( y_{m-k} \overline{y}_{m-p} | \beta, \eta \right) \quad p = 1, ..., DB \tag{7.10}
\]

where \( y_{m-DB-1} \) is replaced by unity; and \( b_{n,k} (\beta, \eta) \) is the \( k \)th tap for the \((DB+1)\)-tap MMSE predictor of \( y_m \) given the hypothesised phasor sequence and the synchronisation parameters. \( b_{n,DB+1} = 0 \) for Rayleigh channels. The truncated predictor computes

\[
E \left( y_m | \widehat{\beta}, \eta_{m-1} \right) = \sum_{k=1}^{DB+1} b_{n,k} (\widehat{\beta}, \eta) y_{m-k} \tag{7.11}
\]

The MMSE equals
7 DESIGN OF THE MLSE PREDICTOR RECEIVER

\[
\sigma^2\left( y_n \mid \beta, y_{n-1}, \eta \right) = E \left( \left| y_n \right|^2 \mid \beta, y_{n-1}, \eta \right) - \sum_{k=1}^{\lfloor B/p \rfloor} b_{w,k} \left( \beta, \eta \right) E \left( y_{n+1} \mid y_{n-k}, \beta, \eta \right) \tag{7.12}
\]

The path metric, \( \Lambda_{(i-1)r} \left( y \mid \hat{\beta}, \eta \right) \), is a function of the hypothesised phasors, \( \left( \hat{\beta}_{(i-1)r} \right) \). The branch metric is a function of the hypothesised phasors, \( \left( \hat{\beta}_{(i-1)r} \right) \). A state, \( \sigma_{(i-1)r} = \left( \hat{\beta}_{(i-1)r} \right) \) is associated with each surviving path metric, \( \Lambda_{(i-1)r} \left( y \mid \hat{\beta}, \eta \right) \). When the path metric is extended to the \( i \)th symbol period, the branch metric, \( \lambda_{ir} \left( y \mid \hat{\beta}, \eta \right) \), is computed. It depends only on the path metric’s state and the new phasor, \( \hat{\beta}_{ir} \), up to the \( P \)-ary phase ambiguity in \( \{ \beta \} \). Thus the path metric evolution is Markovian, since it depends only on the previous state. By restricting the predictor length and ISI length to \( W \) symbol periods, the number of distinct branch metrics is finite, at \( M^{W-1} \). Thus the Viterbi algorithm can be used to find the sequence with maximum metric. It processes a partially-connected trellis, with \( M^{W-1}/P \) states, and \( M \) branches per state. Every \( W \)-phasor hypothesis vector, \( \sigma_{0 \cdots W} \left( \hat{\beta}_{ir} \right) \), or equivalently, \( \sigma_{0 \cdots W} \left( \hat{\gamma}_{ir} \right) \), labels a branch. The notation, \( \text{sector}(.) \), indicates that only the \( \log_2 A \) sector (amplitude) bits are of interest. Arranging the states on \( \gamma \) builds the rotational invariance directly into the trellis. Truncating the predictors has transformed the tree search into a trellis search. The resulting computational flow in this receiver structure is presented in figure 7.1.

In practice, a viable reduced-complexity receiver can retain fewer survivor sequences, one for each combination of the latest \( S \leq W \) hypothesised symbols. The choice for \( S \) is a trade-off between complexity and performance [61].
7.4 INDEPENDENT AND NEARLY INDEPENDENT THREADS

When the threads are ideally independent, the predictors can be substantially simplified. Only samples from the $d$th thread are useful in predicting another sample from the $d$th thread, and so the predictor can be shortened to $B+1$ taps without a reduction in performance. The same predictor tap weights are used for all threads, and they can be solved from the following system of equations,

$$E(y_i^d|\beta, \eta) = \sum_{k=1}^{B+1} b_{k,1} (\beta, \eta) E(y_{i+k}^d|\beta, \eta)$$

(7.14)

$$\frac{1}{2} E(y_i^d \bar{y}_{i+m}^d|\beta, \eta) = \sum_{k=1}^{B+1} b_{k,1} (\beta, \eta) \frac{1}{2} E(y_{i+k}^d \bar{y}_{i+m}^d|\beta, \eta) \quad m = 1, \ldots, B$$

(7.15)

where $y_i^{d_{B+1}}$ is placed by unity; and $b_{k,1} (\beta, \eta)$ is the $k$th tap for the $(B+1)$-tap MMSE predictor of $y_i^d$, $d = 1, \ldots, D$, given the hypothesised phasor sequence and the synchronisation parameters. The truncated predictor now computes
7. DESIGN OF THE MLSE PREDICTOR RECEIVER

\[ E\left(y_m \mid \beta, q_{m-1}, \eta\right) = E\left(y_{m \mod (D+1)} \mid \beta, q_{m-1}, \eta\right) = \sum_{k=1}^{B+1} b_{m \mod D \cdot k} (\beta, \eta) y_{m \mod D \cdot k} \]  \tag{7.16}

and its MMSE equals

\[ \sigma^2\left(y_m \mid \beta, q_{m-1}, \eta\right) = \sigma^2\left(y_{m \mod (D+1)} \mid \beta, q_{m-1}, \eta\right) = E\left(y_{m \mod (D+1)}^2 \mid \beta, q_{m-1}, \eta\right) - \sum_{k=1}^{B+1} b_{m \mod D \cdot k} (\beta, \eta) E\left(y_{m \mod D \cdot k} y_{m \mod (D+1) \cdot k} \mid \beta, \eta\right) \]  \tag{7.17}

The branch metric can be rewritten to show the sum over the independent diversity threads more clearly, as

\[ \lambda(y, \bar{\beta}, \eta) = \left( \sum_{i,\nu} \sum_{d=1}^{(i+D)-1} \left| y_i^d - E\left(y_i^d \mid \bar{\beta}, q_{i \cdot (D+1)-1} \cdot \eta\right) \right|^2 / \sigma^2\left(y_i^d \mid \bar{\beta}, q_{i \cdot (D+1)-1} \cdot \eta\right) + \ln \sigma^2\left(y_i^d \mid \bar{\beta}, q_{i \cdot (D+1)-1} \cdot \eta\right) \right) \]  \tag{7.18}

Threads are designed to be independent or approximately independent (proper antenna spacing for space diversity; widely spaced carrier frequencies for frequency diversity; sufficient delay for time diversity). Therefore, when the predictor length (receiver complexity) is fixed, it is best to use only samples from the \(d\)th thread in predicting the \(d\)th thread’s next sample. Unlike samples from other threads, these samples are highly correlated with the predicted sample and so they reduce the mean square prediction error more. If the received signal in one thread can be modelled as autoregressive, then there is no loss of performance. Furthermore, this arrangement simplifies the estimation of the signal’s second order statistics, since the inter-antenna autocorrelation is not required.

7.5 THE RAYLEIGH FADING CHANNEL

The Rayleigh fading channel is an important special case. This channel has a zero mean, so equation (7.9) requires that the \((DB+1)\)th tap equals zero. Accordingly, only a \(DB\)-tap predictor is required, and \(y_{m-DB-1}^d\) is not replaced by unity. When the diversity threads are ideally independent in the Rayleigh fading channel, the predictor has only \(B\) taps since the \((B+1)\)th tap equals zero. \(y_{m-B-1}^d\) is not replaced by unity.
For Rayleigh fading channels, the receiver's predictor structure can be derived by two further methods: the Innovation's Process or the Cholesky Decomposition [64]. The Cholesky decomposition approach is described here. From (3.48), the conditional probability of the zero mean, received signal vector, \( y \), is given by

\[
P_{y|\beta, \eta}(y|\beta, \eta) = \frac{1}{(2\pi)^\frac{D_r}{2}} \exp\left(-\frac{1}{2} y^H R_{yy|^\beta, \eta}^{-1} y\right)
\]

(7.19)

Assuming non-zero white noise, the \( (I_E/T) - (I_b/T) \)Dr \( \times \)
\[
(I_E/T) - (I_b/T) \)Dr received signal autocorrelation matrix,
\[
R_{yy|^\beta, \eta} = \frac{1}{2} E(yy^H|\beta, \eta)
\]

(7.20)
is positive definite and its inverse can be Cholesky decomposed as

\[
R_{yy|^\beta, \eta}^{-1} = L^H E L
\]

(7.21)
where \( L \) is a square lower triangular matrix with unit main diagonal entries, and \( E \) is a square diagonal matrix with positive real entries. Then the natural logarithm of equation (7.19) can be written as

\[
\ln p_{y|\beta, \eta}(y|\beta, \eta) = -\frac{1}{2} (Ly)^H E (Ly) - \ln\left(\frac{2^{(I_E/T)\]Dr}}{\det[E]}\right) - \ln\pi (I_E/T)\]Dr
\]

(7.22)

\[
= -\sum_{m=\frac{I_E/T}{Dr}}^{m=\frac{I_b/T}{Dr}} - \left| y_m - \sum_{k=1}^{m} i_{m-n-k} y_{m-k} \right|^2 - \ln(2/e_m) - \ln\pi
\]

where \( e_m \) is the \( m \)th diagonal entry of \( E \) and \( l_{km} \) is the \((k,m)\)th entry of \( L \), but the indices begin at \( \frac{I_b}{T} \)Dr, not the normal unity.

Neglecting the \( \ln\pi \) term since it is independent of the hypothesised phasor sequence and the scale factor -1, the log-likelihood can be used as a recursive path metric, in the form...
\[ \Lambda_r(y|\beta, \eta) = \Lambda_{(r-1)r}(y|\beta, \eta) + \lambda_r(y|\beta, \eta) \]
\[
= \left( \sum_{m=1}^{m-[I_r/T]} \sum_{k=1}^{m-[I_r/T]T_{dr}} \left( \frac{y_m - \sum_{k=1}^{I_m, m-k} y_{m-k}^2}{2/e_m} + \ln 2/e_m \right)^2 + \ln 2/e_m \right) + \sum_{m=I_r}^{(r+1)T_r - 1} \left( \frac{y_m - \sum_{k=1}^{I_m, m-k} y_{m-k}^2}{2/e_m} + \ln 2/e_m \right)^2
\]

(7.23)

This sequence metric is identified as equations (7.7) and (7.8), where \( I_{m,m-k} = h_{m,k}(\hat{\beta}, \eta) \) and \( 2/e_m = \sigma^2 \left( y_m, \beta, \eta - 1, \eta \right) \). Thus the Cholesky decomposition is equivalent to the previous analysis. As before, it is necessary to truncate the predictors for the receiver structure to have finite complexity; this assumes that \( \mathbf{L} \) has non-zero entries only on the main diagonal and the \( B \) diagonals below it.

### 7.6 THE YU AND PASUPATHY RECEIVER STRUCTURE

Yu and Pasupathy [64] describe a receiver structure for the case of Rayleigh fading, independent diversity threads, and ideal synchronisation. When the MLSE predictor receiver described here is specialised for the same conditions, its structure and metric are identical to the structure and metric of [64]. Their branch metric is written as

\[ B_k(a_{k-1}, \ldots, a_k) = \sum_{i=1}^{B} \sum_{j=1}^{B} \left| x_{k+1}^{(i)}(a_{k-1}, \ldots, a_k) \right|^2 + \ln d_{k,i}^{(i)}(a_{k-1}, \ldots, a_k) \]

(7.24)

where in this thesis' notation: \( B = \lambda; k = ir; (a_{k-1}, \ldots, a_k) = (\beta_{i+k-1}, \ldots, \beta_{k}); i = d; j = 1; \beta = r; y \) is the Innovations process, \( \left( y_m - E(y_m|\beta, \eta - 1, \eta) \right)/\sqrt{\sigma^2(y_m|\beta, \eta - 1, \eta)} \); and \( d = \sigma^2 \).
7 DESIGN OF THE MLSE PREDICTOR RECEIVER

7.7 THE VITETTA, DAM, HOLDSWORTH, AND TAYLOR RECEIVER STRUCTURE

In a series of theses and papers, Vitetta, Dam, Holdsworth and Taylor [16, 15, 60, 61, 24] describe and simulate a receiver structure for the frequency-flat, Rayleigh fading channel with $M$-PSK, one thread, ideal synchronisation, and white noise.

The research assumes a rectangular transmitter pulse, equation (2.3), and white noise. This signal model has desirable properties in that there is no ISI and a particularly simple receiver structure results. However, the receiver's anti-aliasing filter complicates the signal model. Since the faded $\text{rect}(.)$ pulse is not bandlimited, it is inevitably distorted by the anti-aliasing filter, leading to ISI.

In fact, the receiver only requires that the transmitted pulse shape be rectangular at the sampling instances, according to (2.4), so a bandlimited pulse can be designed with $r$ suitable sampling instants. If the multi-Nyquist pulse is constructed from $\text{sinc}(.)$ pulses, then the received signal has a one-sided bandwidth $1/2T_r + f_d$. An anti-aliasing filter with bandwidth, $1/2T_r$, introduces ISI and is therefore undesirable. However, if its bandwidth is doubled, to $1/T_r$, the received pulses are undistorted and the noise is white. There is a 3dB SNR loss due to the doubled noise power; but this has not been taken into account by Dam or Vitetta [16, 15, 60, 61], although it is a necessary part of their signal model.

Assuming that the received signal is undistorted by the anti-aliasing filter, the received sample vector over the whole transmission interval, $y$, can be written as

$$ y = \beta z + n $$  \hspace{1cm} (7.25)

where $\beta$ is a diagonal matrix of the transmitted phasors, $\text{diag}\left(\beta_{[1/\tau]}^{[T]}, ..., \beta_{[1/\tau]}^{[T]}\right)$, $z$ is a column vector of samples from the $t$-channel's one tap; and $n$ is a column vector of the additive noise. The signal autocorrelation matrix equals

$$ \frac{1}{2} E(\mathbf{yy}^H | \beta) = \beta \left( \frac{1}{2} E(\mathbf{zz}^H) + \frac{1}{2} E(\mathbf{nn}^H) \right) \beta^H $$  \hspace{1cm} (7.26)

Since the noise is white and the phasors are taken from the $M$-PSK signal constellation, the phasors and the channel/noise autocorrelation, $R_{xx}$, are decoupled.
The predictor tap weights have a particularly simple form, since the Cholesky decomposition of the received signal matrix reduces to the Cholesky decomposition of the channel/noise autocorrelation matrix scaled by the phasor matrices, as

$$ R_{yy}^{-1} = \left( \beta R_{xx} \beta^H \right)^{-1} = \beta L \beta^H $$

(7.27)

Assuming that the channel/noise process is autoregressive, with an order less than $B$, the only non-zero diagonals of $L \beta$ are its main diagonal and some of the $B$ diagonals below it. Then the mean square prediction error is the same for all predictors with at least $B$ taps, and all but the first $B$ entries along the main diagonal of $E_{\beta \beta}$ are the same. Furthermore, $E_{\beta \beta}$ comes from the Cholesky decomposition of the channel/noise autocorrelation matrix, so its entries are independent of the phasor sequence. Accordingly, the entries in $E_{\beta \beta}$ are the same for all samples and sequences, except for the unimportant start-up period. Since the branch metric is used for comparisons only, the presence of a constant $\sigma^2 \left(y_n, \beta, y_{n-1}, n\right)$ has no effect on the receiver's decisions, and can be discarded. Therefore the branch metric can be simplified to

$$ \lambda_{\beta}(y, \beta, n) = \left( \sum_{i=[r]}^{(r+1)T-1} y_i - \beta \sum_{k=1}^{B} b_{x,k} \bar{\beta}_{(i-1)r+j_{i-k}} \right) \left( y_i - \beta \sum_{k=1}^{B} b_{x,k} \bar{\beta}_{(i-1)r+j_{i-k}} \right) $$

(7.28)

where $b_{x,k}$ is the $k$th tap in the $B$-tap, MMSE predictor for the channel/noise process.

Referencing the indices with respect to $\left[ I_n / T \right]$, instead of unity, $b_{x,k}$ is the $(B,B-k)$th entry in $L \beta$. When $B \leq r$ and the transmitted phasors are differentially encoded as $\beta = \beta_{(r+1)T} \exp\left( j2\pi \frac{\alpha_{(r+1)T}}{M} \right)$, the branch metric is a function only of the $i$th hypothesised symbol, $\tilde{\alpha}_{(r+1)T}$, as

$$ \lambda_{\beta}(y, \beta, n) = \sum_{i=[r]}^{(r+1)T-1} y_i - \sum_{k=1}^{M_B} b_{x,k} y_{i-k} - \exp\left( j2\pi \frac{\tilde{\alpha}_{(r+1)T}}{M} \right) \sum_{k=1}^{B} b_{x,k} y_{i-k} $$

(7.29)

The receiver makes hard decisions.
7.8 RECEIVER OPERATION

In the previous sections, the receiver mathematics have been described. Although they are fundamental to the description of the receiver, they shed little light on its operation. Accordingly this section attempts to provide an intuitive appreciation of the receiver's operation.

As a first step, the predictor receiver of [24] is described, since its interpretation is particularly straightforward. QPSK is transmitted and the predictor has $B = r = 4$ taps. In figure 7.2, plot 1 shows the received signal samples over two symbol periods. The fading process changes slowly but steadily during the first symbol period (the light circles). There is a discontinuity at the end of the symbol period due to the rectangular pulse shapes. $\alpha_{ir} = 1$, so there is a 90° anticlockwise rotation. The samples over the second symbol period (the dark circles) also reveal the correlated evolution of the fading process.

The branch metric computes $|y_i - \exp(j2\pi \frac{\hat{\alpha}_{ir}}{M}) \sum_{k\mod r+1} b_{x,k} y_{l-k}|^2$ for the first sample. The light circles are used to predict the first sample in the second symbol period, then this prediction, $\sum_{k\mod r+1} b_{x,k} y_{l-k}$ (denoted by $\otimes$ in plot 2), is rotated by the hypothesised data, $\exp(j2\pi \frac{\hat{\alpha}_{ir}}{M})$ for all allowed $\alpha_{ir}$. A Euclidean distance is calculated between the received signal sample, $y_i$ (the dark circle in plots 3 - 6), and the hypothesised sample, $\exp(j2\pi \frac{\hat{\alpha}_{ir}}{M}) \sum_{k\mod r+1} b_{x,k} y_{l-k}$ (denoted by $\otimes$ in plots 3 - 6).

Clearly the distance is smallest in plot 3, so the hypothesis $\alpha_{ir} = 1$ is promising. The whole branch metric predicts all $r = 4$ of the received samples and sums the Euclidean distances between the received sample and the rotated prediction. For the correct hypothesis, $\alpha_{ir} = 1$, the predictor contents are the smoothly evolving fading process. For the incorrect hypotheses, for the remaining three predictions, the predictor contents contain a discontinuity. Accordingly, all $B = 4$ of their predictions are more inaccurate on average than the predictions for the correct hypothesis. This leads ultimately to fourth order diversity.
inaccurate on average than the predictions for the correct hypothesis. This leads ultimately to fourth order diversity.

In general, this receiver is merely predicting one tap of the stationary $t$-channel. In general, the received signal is the convolution of overlapping transmitted pulse
agglomeration is still a correlated complex Gaussian random process, so it can be predicted. Since the transmitted symbols change, the received signal is non-stationary. The receiver accommodates the problem by using different predictor tap weights for each sub-sequence of transmitted symbols.

The influence of the synchronisation parameters on the predictor tap weights merits further explanation. In the Rayleigh fading channel, the carrier phase has no influence. Predictors use past samples of the received signal as an amplitude and phase reference; and so the branch metric considers changes in amplitude and phase. The carrier phase is an absolute phase reference, and is not required when the data is differentially encoded. This can be seen in figure 7.2. A different carrier phase rotates all the points in the plots by the same amount. However, the Euclidean distance between the received samples and the hypothesised received samples is invariant to this rotation.

The effect of the carrier frequency offset is described most easily when the same phasor is always transmitted. Samples of the real and imaginary parts of the baseband received signal are then i.i.d; so the received signal evolves clockwise or anticlockwise with equal probability. The received signal’s autocorrelation matrix is real, and so too are the predictor tap weights. The real part of the signal is predicted independently from the imaginary part of the signal. When there is a residual carrier frequency offset in the received signal, there is a consistent tendency for it to evolve in a clockwise or anticlockwise direction (depending on the sign of the carrier offset). A three-dimensional plot of the received signal, with time, real amplitude and imaginary amplitude as axes shows a helical rotation in the received signal around the time axis, with a random but consistent mean rotation rate. The received signal’s autocorrelation matrix is complex, and so too are the predictors. In fact they accommodate the carrier offset by helically rotating in the opposite direction to the received signal, at its mean rotation rate.

The sequence metric is the sum of Euclidean distances between the received samples and their predicted values, conditioned on a hypothesised sequence. Therefore the sequence metric does not have particular symbol boundaries. In fact, the only symbol-rate quantity in the predictor receiver is the branch metric, which computes \( r \) Euclidean distances. The \( r \) predictions require predictor tap weights,
which in turn are computed from a hypothesis vector. Whatever \( r \) samples are selected as a symbol period, the oldest symbol can be discarded from the hypothesis vector after the \( r \)th Euclidean distance has been processed. Every symbol period, another decision is made. Therefore the sampling phase and symbol boundaries are irrelevant: since the MLSE predictor receiver makes one decision every symbol period. The need for symbol timing is merely the need to know when the receiver begins making meaningful decisions. This information is deduced by noting the substantial increase in received power at the start of transmission.

7.9 PILOT SYMBOLS

The EMLSE-\( tf \) diversity receiver needs pilot symbols to be time isolated from the data symbols, so that channel state information is available before detection. The MLSE predictor receiver does not suffer this deficiency.

Pilot symbols were used originally in the \( t \)-channel to remove the \( P \)-ary phase ambiguity. They are regularly inserted into the symbol stream, and are recovered at the receiver by a demultiplexor. An interpolating filter provides CSI between the pilot symbols. Clearly this technique is suboptimal since a pilot symbol only conveys as much channel state information as other symbols. A multi-pass approach can be employed, where first only the pilot symbols are used for CSI; second the data symbols are detected; third both the pilot and detected symbols are used for CSI; and fourth the data symbols are detected again. Although this technique is effective, the optimal approach jointly estimates the CSI and symbols.

The MLSE predictor receiver structure achieves this identically. The pilot symbols influence the receiver structure only through the branch metric: they are not demultiplexed and used for CSI. When a pilot symbol appears in the state vector, (7.13), the branch metrics are infinite for the state vectors that hypothesise a value for the pilot symbol other than its known value. The receiver uses the MAP criterion rather than MLSE, since it is exploiting the unequal symbol probabilities. Equivalently, the trellis is time-varying, with fewer states in the vicinity of the pilot symbol.
When the pilot symbols are spaced sufficiently frequently, coherent detection is possible. There is no need for rotationally invariant or amplitude-slip tolerant constellation mappings. It is not immediately clear how frequently "sufficiently frequently" is. However, the MLSE receiver should at least outperform any sub-optimal implementation, such as the traditional demultiplexor/interpolation filter structure. Therefore pilot symbols should be spaced at least as often as the Nyquist rate for the channel tap weight processes. There may be an additional requirement that the predictors should be long enough that the state vector encompasses at least one pilot symbol.

7.10 ESTIMATING THE SIGNAL AUTOCORRELATION

The previous derivation assumes that the received signal's second order statistics are available, whereas in practice they must be estimated. It is assumed that only samples from the \(d\)th thread are used to predict the \(d\)th thread's next sample.

Two methods for estimating the signal's second order statistics are proposed for Rayleigh fading channels. These are described most easily through vectors, so more notation is defined initially. A sample segment of the received signal from \(i-1\) to \((i+1)r-1\) from the \(d\)th antenna can be written as

\[
y_{d} = \beta_{d} e_{d} + n_{d}
\]

where \(y_{d} = [y_{d,0}, ..., y_{d,r-1}]^{T}\) is a \(B+r\) column vector of received samples,

\(n_{d} = [n_{d,0}, ..., n_{d,r-1}]^{T}\) is a \(B+r\) column vector of noise samples;

\[
\beta_{d} = \begin{bmatrix}
\beta_{d-1(\text{r-0})}, ..., \beta_{d-1(r-1)} & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\beta_{d-1(\text{r-1})}, ..., \beta_{d-1(\text{r-0})}
\end{bmatrix}
\]

is a \((B+r) \times (B+r)\) matrix of the data phasors and
DESIGN OF THE MLSE PREDICTOR RECEIVER 136

de_\nu^d = \begin{bmatrix}
\mathbf{c}_{\nu - B - (L + 1) r}^{d, H} c_{\nu + (L + 1) r}^{d, H} \\
\vdots \\
\mathbf{c}_{\nu - 1}^{d, H} c_{\nu + (L + 1) r}^{d, H} \\
\mathbf{c}_{\nu}^{d, H} c_{\nu + (L + 1) r}^{d, H} \\
\mathbf{c}_{\nu + 1}^{d, H} c_{\nu + (L + 1) r}^{d, H} \end{bmatrix}

is a \((B+r)L\) column vector of the channel.

Then the signal’s second order statistics can be represented as \(M^H/P\) matrices,

\[
\frac{1}{L} E\left(y_{\nu}^{d, H} y_{\nu}^{d, H} | \beta, \eta \right) = \beta_{\nu} \frac{1}{L} E\left(\mathbf{c}_{\nu}^{d, H} \mathbf{c}_{\nu}^{d, H} | \eta \right) \beta_{\nu}^{H} + \frac{1}{L} E\left(\mathbf{n}_{\nu}^{d, H} \mathbf{n}_{\nu}^{d, H} \right) 
\]

one for each distinguishable data sequence. The channel and noise’s second order
statistics are identical for all antennae and so they need not be individually indexed,
and hence

\[
\frac{1}{L} E\left(y_{\nu}^{d, H} y_{\nu}^{d, H} | \beta, \eta \right) = \beta_{\nu} \frac{1}{L} E\left(\mathbf{c}_{\nu}^{d, H} \mathbf{c}_{\nu}^{d, H} | \eta \right) \beta_{\nu}^{H} + \frac{1}{L} E\left(\mathbf{n}_{\nu}^{d, H} \mathbf{n}_{\nu}^{d, H} \right) 
\]

Equation (7.32) uses the principal \((B+1) \times (B+1)\) submatrices of
\(\frac{1}{L} E\left(y_{\nu}^{d, H} y_{\nu}^{d, H} | \beta, \eta \right)\) to compute \(B\)-tap predictors for the \(Dr\) samples, \(y_{\nu}^{1}, \ldots, y_{\nu}^{D} \).

The first method to estimate \(\frac{1}{L} E\left(y_{\nu}^{d, H} y_{\nu}^{d, H} | \beta, \eta \right)\) uses time-averaging for each of
the \(M^H/P\) sub-sequences, through the recursion,

\[
\frac{1}{L} E\left(y_{\nu}^{d, H} y_{\nu}^{d, H} | \beta, \eta \right) \sim \hat{R}_{\nu}^{d, d} = \begin{cases}
\mu_{\nu} \hat{R}_{\nu}^{d, d} | \beta, \nu (n-1) \rangle + \sum_{\nu \neq \nu} y_{\nu}^{d, H} y_{\nu}^{d, H} & \beta \text{ transmitted} \\
\hat{R}_{\nu}^{d, d} | \beta, \nu (n-1) \rangle & \beta \text{ not transmitted}
\end{cases} 
\]

where \(0 < \mu_{\nu} \leq 1\) is a possibly time-varying forget factor to accommodate channels
with non-stationary second order statistics.

Before the receiver can begin detection, at least \(\lceil B/D \rceil\) repetitions of each
sub-sequence must be received, so that the system of simultaneous equations,
equation (7.15), has a unique solution. Therefore a long training sequence is needed.
Adaptation in non-stationary environments is slow, since each signal autocorrelation
matrix and predictor tap weights can only be updated when their sub-sequence is
transmitted - i.e. only once every \(M^H/P\) symbols on average. Furthermore, decision-
aided adaptation is potentially unstable, since it is the predictor tap weights themselves that control which predictor tap weights are updated. There is little to stop a signal autocorrelation matrix for one sequence being updated only when another sequence is transmitted, since each time the error happens, the error is reinforced.

For the special case of $L = 1$ and $M$-PSK, this technique can modified to update the signal autocorrelation matrices every symbol period [15]. The received pulse autocorrelation can be updated whenever $\beta_\nu$ is transmitted, as

$$\frac{1}{2} E(e_\nu e_\nu^H | \eta) + \frac{1}{2} E(n_\nu n_\nu^H) - \hat{R}_{zz | n_\nu} = \mu_\nu \hat{R}_{zz | n_{\nu-1}} + \sum_{d=1}^{D} \beta_\nu^{-1} y_{d | \nu}^d y_{d | \nu}^{d,H} \beta_\nu^{-H} \tag{7.34}$$

allowing every signal autocorrelation matrix to be computed by

$$\hat{R}_{yy | n_\nu} = \beta_\nu \hat{R}_{zz | n_\nu} B_\nu^H \tag{7.35}$$

The second method extends the case of $L = 1$ to the case of arbitrary $L$, while retaining its advantages: namely reduced training requirements and faster adaptation rates, since the received pulse autocorrelation matrix is updated every symbol; and increased stability, since the signal autocorrelation matrices do not evolve independently. Given the received pulse and noise autocorrelation matrices, $\frac{1}{2} E(e_\nu e_\nu^H | \eta)$ and $\frac{1}{2} E(n_\nu n_\nu^H) = \frac{N_0}{\nu} \mathbb{I}$, the signal autocorrelation matrices can be computed from equation (4.32).

However, $\frac{1}{2} E(e_\nu e_\nu^H | \eta)$ is now $(B+r)L \times (B+r)L$, whereas the receiver computes noisy $(B+r) \times (B+r)$ projections of it: the $M^H / P$ signal autocorrelations, $\hat{R}_{yy | \nu, n_\nu}$ of equation (7.33). Furthermore, the received pulse autocorrelation matrix must be positive definite, to ensure that the signal autocorrelation matrices are positive definite. If this constraint is not made, the MMSE's computed in equation (7.17) can be negative, and the computed branch metric is meaningless.

These requirements are solved by the following *ad hoc* scheme. By Cholesky (square root) decomposing $\frac{1}{2} E(e_\nu e_\nu^H | \eta)$ as $S_{ee | \nu} S_{ee | \nu}^H$, and then solving for $S_{ee | \nu}$ under the constraint that its main diagonal entries are positive real, the received pulse autocorrelation matrix is always positive definite. $S_{ee | \nu}$ and the noise power are calculated from an overdetermined system of non-linear simultaneous equations, since
each entry of each signal autocorrelation matrix must equal a non-linear function of the entries of $S_{cc,ir}$, as

$$
\beta_{ir}S_{cc,ir}^H S_{cc,ir}^H \beta_{ir} + \frac{N_0}{2} I - \frac{1}{2} E\left(y_{ir} y_{ir}^H | \beta, \eta \right) = 0
$$

(7.36)

for all $(B+r) \times (B+r)$ entries of $\frac{1}{2} E\left(y_{ir} y_{ir}^H | \beta, \eta \right)$, for all $M^W/P$ signal autocorrelation matrices. In practice, $\frac{1}{2} E\left(y_{ir} y_{ir}^H | \beta, \eta \right)$ is unavailable, so $\hat{R}_{yy,\beta,\eta,ir}$ is substituted, and allowance is made for scaling due to the forgetting factor, $\mu_{ir}$. Since $\hat{R}_{yy,\beta,\eta,ir}$ is Hermitian symmetric, there are $(M^W/P)(B+r)(B+r+1)/2$ distinct equations in all. There are $(B+r)^2+1$ unknowns: $S_{cc,ir}$ and $N_0$. The off-diagonal entries of $S_{cc,ir}$ introduce two unknowns, real and imaginary but the diagonal entries only introduce one unknown.

The best scheme for solving this system of equations is unclear. We adopt here an iterative scheme that combines the Newton algorithm for solving systems of non-linear simultaneous equations with a generalisation of least squares. The $(B+r)^2+1$ unknowns from $S_{cc,ir}$ and $N_0$ are written as a column vector, $S_{cc,ir}^k$, where the superscript denotes the iteration number. The simultaneous equations expressed by equation (7.36) are written as a $(M^W/P)(B+r)(B+r+1)/2$ column vector, $y(S_{cc,ir}^k)$. The partial derivatives of the simultaneous equations with respect to the $(B+r)^2+1$ unknowns are written as a $(M^W/P)(B+r)(B+r+1)/2 \times (B+r)^2+1$ Jacobian matrix, $J(S_{cc,ir}^k)$. Then the systems of equations is solved by iteratively improving $S_{cc,ir}^k$ as

$$
S_{cc,ir}^{k+1} = S_{cc,ir}^k - \left(J^H(S_{cc,ir}^k)J(S_{cc,ir}^k)\right)^{-1} J^H(S_{cc,ir}^k)y(S_{cc,ir}^k)
$$

(7.37)

In a PSP receiver structure, each state recomputes $S_{cc,ir}$ and $N_0$ every symbol period, computes $M$ signal autocorrelation matrices by equation (7.36) for each branch, and then computes $Mr$ sets of predictor tap weights and MMSE’s, according to equations (7.15) and (7.17). Each branch of each state also updates one signal autocorrelation matrix. The computational flow in this receiver is shown in figure 7.3.

However, solving the system of non-linear equations is very complex, except for small values of $B$, $r$, and $M$, whereas updating the signal autocorrelation matrices is relatively inexpensive. Thus complexity can be reduced by
• only solving for \( S_{\ell,\beta} \) at the end of the training sequence.
• solving for \( S_{\ell,\beta} \) intermittently during data transmission.
• keeping only one copy of the signal autocorrelation matrices, and using the survivor with the best path metric to update it every symbol. PSP was initially developed for \( f \)-channel receivers used in the \( t \)-channel, so reducing the delay to tentative decisions dramatically improves their performance. While the instantaneous value of the fading is known to be non-stationary, the fading process' second order statistics are generally much more stable [13]. Accordingly, the decision delay is not as important a problem.

In principle the receiver can blindly detect the signal by replacing the length \( L_n \) training sequence with approximately \( M^{ln}/P \) states. By conditioning on the otherwise unknown data, each state can compute its own set of signal autocorrelation matrices. All but one are wrong, but that one's predictors should have a smaller mean square error than predictors calculated from miscalculated signal autocorrelation matrices, so that one should survive. However, this receiver's complexity is substantial, and a full PSP structure with a set of autocorrelation matrices per state is essential.

7.11 THE BLIND RECEIVER

Having considered schemes for estimating the received pulse's autocorrelation, it is instructive to consider the optimal arrangement, where the need for estimating the received pulse's autocorrelation is recognised from the start. There are two distinguishing characteristics. First, the metric acknowledges that the estimated received pulse's autocorrelation is inaccurate initially. Second, the number of states is not finite, at \( M^{ln} \); instead each hypothesised sequence leads to a different received pulse autocorrelation, distinct predictors, and thus a state that depends on the entire phasor sequence history. The number of states increases exponentially with time.
7 DESIGN OF THE MLSE PREDICTOR RECEIVER

Figure 7.3: Computational flow for the MLSE receiver structure for unknown channels when the unknown signal’s second order statistics are estimated in a PSP arrangement. The inter-thread correlation is not exploited.

7.12 SUMMARY

An MLSE diversity receiver has been presented that jointly performs MLSE for the case of an unknown carrier, unknown symbol timing, and an unknown fast Ricean/Rayleigh fading, frequency-selective channel, correlated between threads, and corrupted by white noise. The receiver is interesting since it performs joint synchronisation, equalisation and detection in a maximum-likelihood manner when it has ideal knowledge of the (relatively constant) received pulse and noise’s second order statistics. When less complicated synchronisation, transmitter, and channel models are considered, the receiver structure simplifies to that of [64], [60], or [15].

Two methods are developed to adaptively estimate the received pulse and noise’s second order statistics. With the second method, there are no theoretical impediments to implementing the receiver; in that the receiver does not require information that is not available. However, there is the practical problem of high complexity.
8 PERFORMANCE EVALUATION OF THE MLSE PREDICTOR RECEIVER STRUCTURE

8.1 INTRODUCTION

In this chapter, the performance of the MLSE predictor receiver structure is characterised. Two approaches are taken. First, the BER of the receiver structure in a Rayleigh fading channel is bounded analytically. Second, the system of transmitter, channel and receiver are simulated. In the remainder of the chapter, the receiver’s performance is studied for a series of transmitter, channel and receiver parameters.

8.2 RECEIVER ANALYSIS

The MLSE predictor receiver’s BER is a lower bound on the achievable BER rate in the absence of channel sounding. In this section, a fast Rayleigh fading, frequency-selective channel in white noise is considered, using the bounding technique of section 4.3. The techniques can be extended to coded transmission and Ricean fading channels. Even without channel coding, the channel’s Doppler and delay spread lead to implicit diversity and an improved BER. The rotationally invariant code is assumed to be differential encoding, and the signal’s second order statistics are ideally known. Only samples from the $d$th thread are used to predict the $d$th thread’s next sample. Without loss of generality, the beginning of the error event is aligned at $i = 0$, so equation (4.10) can be written as

$$ f: \{ \beta_2^{r}, \beta_3^{r}, \beta_0^{r} + e_0^{n,v,w}, \ldots, \beta_0^{n,v,w} e_{(u-1)r} + e_{(u-1)r} \} \exp \left( j \theta_{n,v,w} \right), ... \} \tag{8.1} $$

The pairwise probability of error depends on the hypothesised and transmitted phasors in the vicinity of the error event, since they determine which predictor tap weights are used. Clearly the phasors $\{ \beta_0^{n,v,w}, \ldots, \beta_0^{n,v,w} e_{(u-1)r} \}$ and $\{ \beta_0^{n,v,w}, \ldots, \beta_0^{n,v,w} e_{(u-1)r} \}$ affect the pairwise error probability. Define $\Phi = (L + u - 1)r - 1$. The signal samples, $y_0^{d}, \ldots, y_{\Phi}^{d}$, $d = 1, \ldots, D$, involve pulse tails from the erroneous symbols. When these
samples are predicted, or are used in a prediction, the predictor tap weights are wrong. Define \( \omega = -B \) and \( \Omega = (L + u - 1)r - 1 + B \). The signal samples \( y^d_0, \ldots, y^d_\Omega \), \( d = 1, \ldots, D \), are used with \( y^d_0, \ldots, y^d_\Omega \) in predictions, so in fact any symbols with pulse tails appearing within \( \omega, \ldots, \Omega \) are relevant. Define \( \psi = -W + 1 \) and \( \Psi = W + u - 2 \). Then \( \{ \beta^d_{\omega r}, \ldots, \beta^d_{\Omega r} \} \) and \( \{ \beta^d_{\omega,\omega,\omega}, \ldots, \beta^d_{\Omega,\Omega,\Omega} \} \) affect the pairwise probability of error. Since the ISI from these symbols is different in each case, pairwise error probabilities must be tediously computed for each ISI combination, up to the \( P \)-ary rotational ambiguity.

An upper bound on the receiver’s BER can be deduced from a union bound of error events,

\[
\text{BER} \leq \sum_{\omega=0}^{\infty} \sum_{r=1}^{M^{(r-v-1)/r}} \frac{P(\beta^d_{\omega r}) P(\beta^d_{\omega r} \rightarrow \beta^d_{\omega,\omega,\omega}) e(\beta^d_{\omega r} \rightarrow \beta^d_{\omega,\omega,\omega})}{\log_2 M} \tag{8.2}
\]

which is the union over all error events between all transmitted sequences and all possible error sequences, weighted by their transmission probability,

\[
P(\beta^d_{\omega r}) = \frac{P}{M^{r-v-1}}, \tag{8.3}
\]

the pairwise error probability, and the number of bit errors they introduce. Error events can begin in any symbol interval, during which \( \log_2 M \) bits are sent.

In practice, only the dominant short error events are considered, where up to \( E \) symbols may be in error. The range of the first summation in equation (8.2) is then \( u = 0, \ldots, E \). If the fading is fast and the SNR high, then the dominant error events are short, so the truncated bound can validly neglect long error events. However, with slow fading, most errors occur in the deep fades, when the instantaneous SNR is very low. If the fading is very slow then the deep fade lasts hundreds of symbols at reasonable SNRs, and so do the error events. Thus the upper bound is tight and easily calculated only for fast fading and high SNR. The union of \( E = 1 \) error events due to the nearest neighbours of the errored symbol is asymptotically correct at high SNRs.

To simplify subsequent notation, normalised predictor tap weights and a bias term are defined as
PERFORMANCE EVALUATION OF THE MLSE PREDICTOR RECEIVER STRUCTURE

\[ b_{l,k}^{\nu,v} = \begin{cases} \frac{1}{\sigma(y_{il}^{BD})}b_{l,k}^{\nu,v}y_{il}^{BD-1} - b_{l,k}^{\nu,v}y_{il}^{BD} & k = 0 \\ -b_{l,k}^{\nu,v}y_{il}^{BD} & k = 1, B \end{cases} \]

\[ b_{l,k}^{\nu,v,w} = \begin{cases} \frac{1}{\sigma(y_{il}^{BD})}b_{l,k}^{\nu,v,w}y_{il}^{BD-1} - b_{l,k}^{\nu,v,w}y_{il}^{BD} & k = 0 \\ -b_{l,k}^{\nu,v,w}y_{il}^{BD} & k = 1, B \end{cases} \] (8.4)

\[ \kappa_{mm} = D \sum_{j=0}^{\phi} \ln \left( \frac{\sigma^2(y_{il}^{BD})b_{l,k}^{\nu,v,y_{il}^{BD-1},\eta}}{\sigma^2(y_{il}^{BD})b_{l,k}^{\nu,v,y_{il}^{BD-1},\eta}} \right) \] (8.5)

where \( \sigma(y_{il}^{BD})b_{l,k}^{\nu,v,y_{il}^{BD-1},\eta} = \sigma(y_{il}^{BD})b_{l,k}^{\nu,v,y_{il}^{BD-1},\eta} \), \( d = 1, \ldots, D \), since the MMSE is common for all threads.

Then the pairwise probability of error is the probability that an erroneous sequence has a smaller path metric than the transmitted sequence,

\[ P(\beta^{u,v} \rightarrow \beta^{u,v,w}) = P\left( \Lambda^{u,v}_{l,k,T} > \Lambda^{u,v,w}_{l,k,T} \right) \]

\[ = P\left( \sum_{d=1}^{D} \sum_{l=0}^{D} \sum_{k=0}^{B} b_{l,k}^{u,v}b_{l,k}^{u,v} - b_{l,k}^{u,v,w}b_{l,k}^{u,v,w} \right) > \kappa_{mm} \] (8.6)

from equation (7.7).

Define \( \kappa^{u,v,w} \) as the left-hand-side of the inequality and the column vector, \( y = \left[ y_{1,1}^{u,v}, \ldots, y_{2,1}^{u,v}, \ldots, y_{D,1}^{u,v}, \ldots, y_{1,1}^{u,v} \right]^T \). \( \kappa^{u,v,w} \) is a Gaussian quadratic form in the complex Gaussian random variables, \( y \), and can be written as,

\[ \kappa^{u,v,w} = y^{H} Y^{u,v,w} y \] (8.7)

where the kernel, \( Y^{u,v,w} \), is a Hermitian symmetric matrix, defined by equation (8.6). The covariance of \( y \) is given by equation (4.32). The characteristic function of a Gaussian quadratic form, \( \kappa^{u,v,w} \), is given by

\[ \phi_{\kappa^{u,v,w}}(\zeta) = \frac{1}{\text{det} \left[ 1 - j2\zeta R_{kk} G^{u,v,w} \right]} \] (8.8)

and the pairwise probability of error can be calculated from Appendix B as
where $p_i$ is the $i$th pole of equation (8.8). This equation assumes that the poles of equation (8.8) are simple; this is satisfied unless the diversity branches are ideally independent. This more complex case is also considered in Appendix B.

### 8.3 A SIMPLE ANALYTIC EXAMPLE

The previous analysis is general and powerful, but it is not obvious. To make the analysis more accessible, the mathematics of one simple case is presented in this section.

The transmitter parameters are BPSK ($\{\beta\} \in \{\pm 1\}$), and a rectangular pulse of $H = 1$ symbol periods. A frequency-flat channel without carrier or timing offset is used, so that there is only $P = 1$ path, $\tau / T = 0$, and the received pulse is $L = 1$ symbol period long. The Bessel function autocorrelation model of equation (3.40) is used, with $f_0 T$ open. The receiver takes $r = 1$ sample per symbol and its predictors use $B = 1$ tap. $W = 2$. The receiver has $M^{W-1} / P = 1$ state, so it makes hard decisions. An error always introduces a cycle slip, as

$$\{\ldots \beta_{-2}^{u,v}, \beta_{-1}^{u,v}, -\beta_{0}^{u,v}, -\beta_{1}^{u,v}, \ldots \}$$

so $u = 0$. $\omega = -1$ and $\Omega = 0$. $\psi = -1$ and $\Psi = 0$. The BER expression of equation (8.2) simplifies to

$$BER = \sum_{v=1}^{2} \sum_{\omega_{u,\text{error}}} P(\beta_{v}^{u,v} \rightarrow \beta_{v}^{u,v}) = P(\beta_{0}^{u,v} \rightarrow \beta_{0}^{u,v})$$

since $P(\beta_{0}^{u,v}) = \frac{1}{2}$; $e(\beta_{0}^{u,v} \rightarrow \beta_{0}^{u,v}) = 1$; $\log_2 M = 1$; and the pairwise probability of $0^\circ$ being mistaken for $180^\circ$ matches the pairwise probability of $180^\circ$ being mistaken for $0^\circ$. 

From equation (4.32), the signal autocorrelation equals

\[
\frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\} = \frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\} = E_b + \frac{N_0}{T} \\
\frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\} = \frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\} = E_b J_0(2\pi f_D T)
\]

so

\[
\mathbf{R}_s = \begin{bmatrix}
E_b + \frac{N_0}{T} & E_b J_0(2\pi f_D T) \\
E_b J_0(2\pi f_D T) & E_b + \frac{N_0}{T}
\end{bmatrix}
\]

From equations (7.9) and (7.10), the predictor tap weight is

\[
b_{0,1}(\beta, \eta) = \frac{\frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\}}{\frac{1}{4} \mathbb{E}\left\{ y_n y_{n-1} \| \beta, \eta \right\}} = \frac{E_b J_0(2\pi f_D T)}{E_b + \frac{N_0}{T}}
\]

assuming \( \{ \beta \} = \{ ..., 1, 1, ... \} \).

Since the received pulses do not overlap, the MMSE is common for all branch metrics and can be neglected. Therefore \( \kappa_{mn} = 0 \). Accordingly, the pairwise probability that \( \beta^{u,v} \sim \{ ..., 1, 1, ... \} \) has a worse metric than \( \beta^{u,v,w} \sim \{ ..., 1, -1, ... \} \) is given by

\[
P(\beta^{u,v} \to \beta^{u,v,w}) = P\left( \sum_{\alpha=0}^{1} \sum_{\beta=0}^{1} \left( b_{0,\alpha,\beta}^{u,v} - b_{0,\alpha,\beta}^{u,v,w} \right) y_{n-\alpha} y_{-\beta} > 0 \right)
\]

where

\[
b_{0,0}^{u,v} = 1, \quad b_{0,1}^{u,v} = \frac{E_b J_0(2\pi f_D T)}{E_b + \frac{N_0}{T}}
\]

\[
b_{0,0}^{u,v,w} = 1, \quad b_{0,1}^{u,v,w} = \frac{E_b J_0(2\pi f_D T)}{E_b + \frac{N_0}{T}}
\]

and

\[
\mathbf{G}^{u,v,w} = \begin{bmatrix}
0 & -\frac{2E_b J_0(2\pi f_D T)}{E_b + \frac{N_0}{T}} \\
\frac{2E_b J_0(2\pi f_D T)}{E_b + \frac{N_0}{T}} & 0
\end{bmatrix}
\]

The poles of equation (8.8) obey
\[
\begin{bmatrix}
-2E_b^2J_0^2(2\pi f_D T) \\
E_b + \frac{N_0}{T} \\
-2E_bJ_0(2\pi f_D T) \\
E_b + \frac{N_0}{T} \\
\end{bmatrix}
\]
so that

\[
p_1 = \frac{j\left(E_b + \frac{N_0}{T}\right)}{4E_bJ_0(2\pi f_D T)\left(E_b + \frac{N_0}{T}\right) + E_b + \frac{N_0}{T}}
\]
\[
p_2 = \frac{-j\left(E_b + \frac{N_0}{T}\right)}{4E_bJ_0(2\pi f_D T)\left(E_b + \frac{N_0}{T} - E_bJ_0(2\pi f_D T)\right)}
\]

(8.19)

Since \(J_0(2\pi f_D T)\) is restricted to -1 ... 1, the first pole is positive imaginary if \(J_0(2\pi f_D T) > 0\); otherwise it is negative imaginary. The second pole is negative imaginary if \(J_0(2\pi f_D T) > 0\); otherwise it is positive imaginary. Thus the two poles have opposite signs. Assuming the fading is not exceedingly fast \((2\pi f_D T < 2.4)\), the pairwise probability of error given by equation (8.9) equals

\[
P(\beta^* \rightarrow \beta^*\times) = 1 + \frac{1}{1 - \frac{p_1}{p_2}} = \frac{p_1}{p_1 - p_2}
\]

(8.20)

which is also the BER, by equation (8.11).

In slow fading, \(J_0(2\pi f_D T) \approx 1\), \(p_1 \approx \frac{j\left(E_b + \frac{N_0}{T}\right)}{4E_b\left(2E_b + \frac{N_0}{T}\right)}\), \(p_2 \approx \frac{-j\left(E_b + \frac{N_0}{T}\right)}{4E_b\frac{N_0}{T}}\), and

\[
\text{BER} \approx \frac{1}{2\left(1 + \frac{E_b}{N_0/T}\right)}
\]

At low SNR, the BER is approximately 0.5; at high SNR, the BER is inversely proportional to the SNR. In fast fading with no noise,

\[
p_1 \approx \frac{j}{4E_b^2J_0(2\pi f_D T)\left(J_0(2\pi f_D T) + 1\right)} \quad \text{and} \quad p_2 \approx \frac{-j}{4E_b^2J_0(2\pi f_D T)\left(1 - J_0(2\pi f_D T)\right)}
\]

\[
\text{BER} \approx \frac{1 - J_0(2\pi f_D T)}{2}
\]

There is an error floor when \(f_D > 0\); which can be as high as 0.5.
8.4 SIMULATION

The simulation operates by repeatedly transmitting random $M$-ary symbols through a simulated fading channel and the MLSE predictor receiver, until the receiver makes at least 200 bit errors.

The transmitter is simulated by randomly generating an $M$-ary number every symbol period. For $M$-PSK, it is differentially Gray encoded; for the radially symmetric constellation, its phase and amplitude are differentially Gray encoded and mapped directly to the $M$-ary signal constellation. A sequence of complex phasors arises, where the first element in a symbol period is the data phasor and the remaining elements are zero. In principle, this sequence excites the transmitter filter, whose impulse response is $h(IT_r)$. In practice, the simulation requires $P$ independently faded paths at delays that are non-integral multiples of the sample period. Therefore the phasor sequence excites $P$ separate filters, which generate uniformly delayed versions of the transmitted signal. This has been described in section 3.9.

The $\tau_f$-fading channel is simulated as $P$ equally spaced independently Rayleigh fading taps with equal mean power. The $P$ delayed transmitted signals are multiplied individually by zero mean fading processes, generated by filtering complex white noise. The filter impulse response is given by equation (3.58), windowed with a Hanning window.

For the fading rate, $f_D T = 0.1$, a 192 tap filter is employed; for $f_D T = 0.01$, 1920 taps are used. The $P$ faded paths are summed, forming a complex baseband signal. This is multiplied by $\exp(j2\pi f_0 \text{num} T_r)$ to translate it to its residual carrier frequency, $f_0 \text{num}$, then complex Gaussian noise is added. In this way, the received signal is created.

Strictly speaking, the signal should be filtered at this point. This is achieved by generating the received signal at a sampling rate greater than $1/T_r$ Hz, filtering it to $1/T_r$ Hz, then sub-sampling the filter output. However, the purpose of this section is chiefly to verify that the analysis and simulation match. When the sampling rate is sufficient, the filter has no effect. When the sampling rate is inadequate, the filter will distort the received pulses in a way not accounted for by the analysis. As discussed in
section 3.5, the analysis is unduly complicated when the receiver's IF filter is accounted for.

Therefore the approach of [15, 61, 14, 64] is adopted. The received signal is sampled every $T_r$ sec and no IF filtering is performed. The additive noise automatically has as its one-sided bandwidth, $1/2T_r$ Hz, so the simulation departs from reality only insofar as it neglects the influence of the IF filter on the received signal. Alternatively, the IF filter can be widened so there is no distortion, but the $E_b/N_0$ definition does not account for the additional noise. However, the lack of an IF filter means that the effect of a residual carrier offset cannot be meaningfully studied: the BER is invariant to the carrier offset. This arises because the high frequency components are aliased as low frequency components, not filtered out.

The receiver’s Viterbi algorithm has a fixed decision delay of 10 symbols.

### 8.5 COMPARISON OVERVIEW

There are three goals in the following sections: to verify that the analytic results match the simulated results under the same conditions; to characterise the receiver's performance for a range of transmitter, channel and receiver parameters; and to show the complexity-performance trade-offs for the receiver parameters.

As in chapter 5, the communication system has many degrees of freedom. The transmitter is influenced by the constellation size, $M$, constellation geometry ($M$-PSK, $M$-QAM, or the $A-P-p$ radially symmetric constellation), the pulse shape, $h(t)$, its length, $H$, its excess bandwidth, $\chi$, and the encoding. The channel is characterised by the fading pdf, the SNR, the fading rate, $f_0 T$, the delay spread, $\tau/T$, the Doppler spectrum, the delay spectrum, the cross-thread correlation (the antenna spacing and orientation), the symbol timing, $t_0$, and the residual carrier offset, $f_{0,\text{nom}}$. The MLSE predictor receiver structure is affected by the number of threads, $D$, the number of samples per symbol, $r$, the number of states, $M^{K-1}$, the predictor length, $B$, whether one thread is used to predict itself only, or used jointly with the other threads for predicting the other threads, and the decision delay of the Viterbi processor. Adaptation introduces many extra degrees of freedom also.
The scheme of chapter 5 is repeated. A series of BER-SNR graphs are plotted when one parameter only is modified. In this way, the parameter’s influence on the BER is clearly evident.

Similar baseline parameters are chosen, which are repeated here. The baseline transmitter uses QPSK, a square-root raised cosine pulse, \( h(t) \), truncated to \( H = 1.5 \) symbol periods, with \( \chi = 50\% \) excess bandwidth. This is realistic, except for the truncated pulse shape which is too bandwidth inefficient. However, this considerably simplifies the computational time, by reducing the number of states in the simulated receiver, and the number of distinct transmitted sub-sequences.

The baseline channel is very challenging. The channel comprises \( P = 3 \) equally spaced paths with equal mean power. A wide delay spread is used, \( \tau/T = 0.5 \). The channel is fast fading, since \( f_D T = 0.1 \). Thus the product \( f_D \tau \) equals 0.05, or one tenth of the overspread limit. The Doppler autocorrelation model is the “windowed” Bessel function model. There is no frequency offset or timing error, so the received pulse occupies \( L = 2 \) symbol periods. The BER is invariant to a timing error, but it unduly increases the received pulse duration, \( L \), and so the analysis and simulation are more computationally expensive.

The baseline receiver has one antenna only. It takes \( r = 3 \) samples per symbol, and the predictor extends over two symbol periods, having \( B = 2r = 6 \) taps. A full complexity receiver is employed, with \( S = W = 4 \). The number of states then is \( M^{W-1}/P = 16 \).

Except for figures 8.1 and 8.2, the analytic curves are the union bound of cycle slips and one-symbol, nearest neighbour error events. By examining the error events in simulations, the cycle slip is clearly the dominant error event at high SNR, for a wide range of parameters. At moderate SNRs \( (< 25\text{dB}) \), the one-symbol, nearest neighbour error event also contributes significantly to the BER. This is confirmed in figures 8.1 and 8.2. Figure 8.1 demonstrates that the dominant error event at an asymptotically high \( E_b/N_0 \) is indeed the cycle-slip in fast fading, \( f_D T = 0.1 \). Other curves are plotted which consider many more error events; but these converge quickly towards the cycle slip curve. However, this effect is less pronounced in slower fading, \( f_D T = 0.01 \), as shown by figure 8.2. The union bounds do converge on the cycle slip;
relatively higher $E_b/N_0$. In slower fading, the mean fade duration is longer, so longer error events are more likely. Thus the union bound is loose until the mean fade duration is diminished by an increase in $E_b/N_0$. 

Figure 8.1: BER-SNR curves for different union bounds. NN = nearest neighbour; CS = cycle slip. Thus the curves are the union of the following error events: nearest neighbour cycle slips; nearest neighbour cycle slips and nearest neighbour, 1 symbol error events; and all combinations of cycle slips and 1 symbol error events.

Figure 8.2: BER-SNR curves for different union bounds, with $f_0T = 0.01$. 

![Graph](image-url)
8 PERFORMANCE EVALUATION OF THE MLSE PREDICTOR RECEIVER STRUCTURE

8.6 TRANSMITTER PARAMETERS

The prime constraint in mobile communications is the finite spectrum. Accordingly we examine the effect of the constellation size in figure 8.3. The simulated points closely match the analytic curves. The small variability is due in part to the statistical nature of simulations; and in part since the analytic curves are truncated union bounds. The receiver's performance in this extreme channel is excellent. A BER floor only appears at a high $E_b/N_0$, and it is orders of magnitude lower than other receiver designs. As we shall see, this error floor can be lowered again by increasing the sampling rate and predictor length. The increase in constellation size degrades the $E_b/N_0$ performance of the receiver. BPSK does not exhibit an error floor in the range of BERs plotted. QPSK shows a small power loss relative to BPSK at a low $E_b/N_0$, which worsens as the BER floor below $10^{-6}$ appears. 8-PSK shows a significant loss, and is outperformed by the radially symmetric constellation, $P=4/A=2/\rho=0.5$. This comparison is important, since the spectral efficiency of both signal constellations is the same. Therefore the radially symmetric constellation is superior, and it merits further research. The optimal value of the shell ratio, $\rho$, is one outstanding problem.

One concern with the baseline set of parameters is the transmitter pulse shape, which is unrealistically short. In figure 8.4, the sensitivity of the BER to the pulse's length is addressed. Again the analytic curves closely match the simulated points. For the short pulse, $H=0.5$, the total pulse length is only $L=1$ symbol periods long, so there is no ISI. The implicit delay and Doppler diversity can be easily exploited, since there is no pulse "soup," where the overlapping transmitter pulses overlap and conceal one another. Therefore its performance is excellent. For the longer pulses, $H=1.5, 2.5$, there is ISI and the BER is inferior. However, both curves are very similar, so the choice of a short transmitter pulse does not pervert the conclusions from other figures.
A square-root raised cosine pulse is used as the baseline, since it is widely used in the AWGN channel, and researchers are familiar with it. Since the channel distorts the signal in time- and frequency, its square-root Nyquist properties are unimportant. 50% excess bandwidth is used in the baseline curves, as a typical value. In figure 8.5, the sensitivity of the BER to the excess bandwidth is considered. Since the pulse is truncated so severely, the excess bandwidth does not affect the transmitter's impulse response significantly. This is seen in the figure 8.5, where all the BER curves are similar. However, the pulses with the smallest excess bandwidths exhibit the worst BER floors. Although the main lobes of their power spectral density are the narrowest,
the sidelobes in their power spectral density decay more slowly due to the pulse truncation. Accordingly, these pulses are undersampled the worst, making prediction more difficult.

8.7 CHANNEL PARAMETERS

When the Doppler and delay spreads are relatively small, the BER is improved through their implicit diversity. However, as the Doppler and delay spreads increase towards the overspread limit, the BER degrades again. Pulses smear across adjacent pulses in a quickly changing fashion, so that it becomes difficult for the receiver to distinguish one sequence from another. The BER is particularly sensitive to a large Doppler spread when the sampling rate and the predictor length are inadequate.

The overspread limit is evident in figure 8.6, where the normalised delay spread, $\tau/T$, equals 0.005, 0.05, and 0.5. The usual fast fading channel, $f_B T = 0.1$, is employed. The implicit diversity available from delay spread varies substantially, yet the BER curves are similar except at high $E_b/N_0$. Thus the implicit diversity is balanced against the more difficult prediction task. At high $E_b/N_0$, the more extreme channel exhibits the worst error floor.

Figure 8.5: BER-SNR curves for a square-root raised cosine pulse, with excess bandwidths from $\chi = 10\%$ to $90\%$. The other parameters take their baseline values.
PERFORMANCE EVALUATION OF THE MLSE PREDICTOR RECEIVER STRUCTURE

Figure 8.6: BER-SNR curves for different delay spreads, $\tau/T$. The other parameters take their baseline values.

Figure 8.7: BER-SNR curves for different delay spreads, $\tau/T$. $f_d T = 0.01$, and the other parameters take their baseline values.

In figure 8.7, the fading rate is reduced to $f_d T = 0.01$, and the delay spread is varied. For $\tau/T = 0.005$ and 0.05, the channel is dominantly time-selective, with a small amount of gain slope across the signal band. The receiver's BER is very similar for both delay spreads, but at high $E_s/N_0$, the receiver performs better with the greater delay spread. This is clearly evident when the delay spread is increased to half a symbol period, where the receiver does 5dB better at a BER of $10^{-4}$. The implicit diversity due to delay spread causes the improvement, but it has negligible effect for $\tau/T = 0.005$ and 0.05. This is the inverse of the previous figure.
Figure 8.8 plots the same curves as in figures 8.6 and 8.7, but now the delay spread is fixed at $\tau/T = 0.5$, and it is the Doppler spread that is varied. The slowly fading channel is predicted more easily, and the delay spread offers significant implicit diversity.
The baseline channel model assumes $P = 3$ independently faded paths. In figure 8.9, the effect of the number of paths is observed. For $P = 1$, the channel is frequency-flat. There is no implicit delay diversity, and the BER is approximately 3 dB worse than the $P = 2$ case. No error floor is evident in either case. As $P$ is increased, the received pulse is not lengthened, but it is more packed more densely. Adjacent pulses cannot be distinguished as easily from the pulse "soup," and the BER curves floor at high $E_s/N_0$. The worst BER floor arises when more paths are considered.

The simulations employ a windowed impulse response to generate the fading process, leading to a "windowed" autocorrelation model. The analysis uses the same autocorrelation function, so that it can be validated by the simulations. However, the Bessel function autocorrelation model of equation (3.40) is standard. In figure 8.10, the BER curves computed from the different autocorrelation models are compared. They match closely, including the BER floor, so the channel's general properties are not obscured when the "windowed" autocorrelation model is used.

8.8 RECEIVER PARAMETERS
A finite complexity receiver design requires that the predictors be truncated. In figure 8.11, this requirement’s consequences are presented. For very short predictors, the receiver’s BER is poor. There is an unsuitably high BER floor. The BER floor diminishes as the predictor length is increased, and for $B = 7$, no error floor is visible in the figure. For a receiver not operating above an $E_b/N_0$ of 10dB, $B = 2$ suffices. When the maximum $E_b/N_0$ equals 30dB, $B = 5$ is sufficient. In this way, the appropriate choice for $B$ depends on the $E_b/N_0$ range. There is no “right” predictor length. This can be seen more vividly in figure 8.12, where a simplified signal model is employed. The predictor length directly affects the BER floor.

Figure 8.11: BER-SNR curves for various predictor lengths, $B$. The other parameters take their baseline values.
Successful prediction requires an adequate sampling rate and a sufficiently long predictor. In figure 8.11, increasing the predictor length always decreases the BER. However, figure 8.13 indicates that increasing the sampling rate does not necessarily decrease the BER. The error floor is certainly reduced. When $r = 1$, the error floor is high; it reduces as $r$ increases to 2 and 3. For $r > 3$, no error floor is visible. However, at moderate $E_b/N_0$, the higher sampling rates have worse BERs. Increasing $r$ shortens the interval that the predictor can predict from, increasing the probability that the signal is faded close to (or below) the noise floor.
The analytic curve is not a bound on the BER for $r = 1$. The analytic curve only considers cycle slips and one symbol, nearest neighbour error events. However, simulations reveal that these are not the only dominant error events for $r = 1$. It is inappropriate to truncate the union bound so severely in this instance.

The sampling rate can be safely increased when the number of predictor taps, $B$, is increased proportionally. This is seen in figure 8.14, where an increase in the sampling rate consistently improves the BER.

The baseline curves use first order diversity only, and require an $E_b/N_0$ of approximately $30\text{dB} - 40\text{dB}$ to achieve a BER of $10^{-4}$. That is, the signal power is on average, one thousand to ten thousand times stronger than the noise power. Clearly this is inefficient, particularly compared to the AWGN channel, which requires the signal power to be less than ten times as strong as the mean noise power for the same BER. Furthermore, in a CCI-limited cellular system, the receiver may not be able to achieve an $E_b/N_0$ of $30\text{dB}$. The only method to reduce the required $E_b/N_0$ is by explicit diversity (including coding). The potential benefits of diversity are plotted in figure 8.15. All threads are independent, so the figure applies to all sources of diversity and it shows the maximum gain from diversity. There is a considerable reduction in the $E_b/N_0$ required to achieve a BER of $10^{-4}$ when one additional thread is available. The benefit of additional threads diminishes with increasing $D$, as can be seen when the number of threads is increased from $D = 2$ to $D = 3$.

Figure 8.14: BER-SNR curves for various $B$, where $B/r$ is kept constant. The other parameters take their baseline values.
8.9 ADAPTATION

In all the previous figures, ideal knowledge of the received signal's second order statistics is assumed. Accordingly, they present "steady-state" results, showing the receiver's asymptotic BER as the transmission duration gets very long and the receiver can estimate the second order statistics very accurately. However, it is the practical case when the receiver begins detection after a short training sequence that is more relevant. In the previous chapter, two methods for estimating the received signal's second order statistics were described. In this section, their performance is evaluated briefly through simulation.

The receiver's BER varies as a function of the training sequence duration. Adaptation of the predictor tap weights is stopped at the end of the training sequence and the receiver is simulated until the receiver makes 200 bit errors. By repeating this process 20 times for each training sequence length, the BER as a function of the training sequence length is estimated, averaged over the channel during training. The simulation results also approximately equal the BER as a function of the transmission duration, if the receiver continues to adapt the predictor tap weights after the training sequence in a decision-directed manner. The decisions must be predominantly correct.
In figure 8.16, the performance of the first adaptation method is examined. The usual baseline parameters are selected, except that the constellation is reduced to BPSK, and the predictor length and sampling rate are curtailed to \( B - r = 2 \). In this way, the receiver needs to compute four received signal autocorrelation matrices only, and each autocorrelation matrix is updated only once every four symbols on average. The forget factor, \( \mu_r \), equals unity.

The solid line is a simulated curve assuming perfect knowledge of the received signal's second order statistics. The receiver's BER using the first method can closely approximate this ideal curve for training lengths of 200, 2000 and 20000 symbols.
corresponding to 50, 500, and 5000 updates to each autocorrelation matrix on average. However, when the training sequence is 20 symbols only, each $4 \times 4$ autocorrelation matrix is updated five times only on average. The time-averaging is insufficient to show the second order statistics clearly, and the receiver's BER suffers. If the overhead of a sufficiently long training sequence is acceptable, then the first method can accurately estimate the received signal's second order statistics.

In figure 8.17, the performance of the second adaptation method is considered. The same parameters are used as for the first adaptation method. This second method performs notably worse than the first method for intermediate training lengths. For the short 20 symbol training sequence, the second method does improve the BER by pooling the sparse information from all the received signal autocorrelation matrices. However, the second method assumes that each matrix has the same underlying received pulse autocorrelation matrix, $\frac{1}{T} E(e_n e_n^H | \eta)$, and accordingly treats the entries in its Cholesky decomposition as the same unknowns. However, this is only true asymptotically, with longer time-averaging. From figure 8.17, this assumption does not hold for training lengths of 200 - 2000. The second method is successful with 20000 training symbols.

8.10 SUMMARY

The performance of the MLSE predictor receiver structure has been analysed, and the analysis shows good agreement with simulation. The BER performance of the MLSE predictor receiver has been studied for a range of transmitter, channel and receiver parameters. The receiver can tolerate large signal constellations, fast fading and a long delay spread, given a sufficiently long predictor and a fast enough sampling rate. Diversity can substantially improve the receiver's power efficiency. Given a sufficiently long training sequence, the receiver can estimate the received signal's second order statistics.
CONCLUSIONS

9.1 ACHIEVEMENTS

Receiver design is an important research topic in communications. There are many areas to study, spanning a range of channel models and receiver complexities. This thesis advances the state-of-the-art by designing optimal or near-optimal receiver structures for very general channel models.

Three novel receiver structures have been designed, simulated and analysed. The first, the EMLSE-\( tf \) diversity receiver, is the MLSE receiver structure when the receiver has perfect knowledge of the channel. The second, the MLSE predictor receiver, is the MLSE receiver structure when the receiver has perfect knowledge of the Ricean fading channel's first and second order statistics. The third, the sequence-averaged predictor receiver, is an effective, but sub-optimal receiver structure. The receivers are designed for time- and frequency-selective channels, with correlated sources of diversity.

The EMLSE-\( tf \) diversity receiver is the optimal receiver design for a fixed linear transmitter over linear channels corrupted by Gaussian noise. No other design has a longer mean time between error events, and at moderate and high SNRs, its BER is essentially a lower bound on all receiver structures. Accordingly, its performance is a benchmark for other receiver structures. When the channel model is simplified, the receiver structure also simplifies in a straightforward way to already known receiver structures. Thus the single EMLSE-\( tf \) diversity receiver derivation can replace a number of derivations for the individual receiver structures. The simplification to the frequency-flat, fast fading channel is a new and interesting result. Using pilot tones and symbols to estimate the channel before detection is described, making the receiver design more relevant. The receiver’s BER is bounded for the important fast Rayleigh fading, frequency-selective channel model.

The MLSE predictor receiver is the optimal receiver design for a fixed linear transmitter over linear Ricean channels corrupted by Gaussian noise, in the absence of channel sounding or \textit{a priori} knowledge about the channel. As with the EMLSE-\( tf \)
diversity receiver, its performance is an appropriate benchmark for other receiver structures. The receiver jointly performs synchronisation and detection, eliminating PLLs and other structures that are inappropriate for fast Rayleigh fading channels. Methods for adaptively estimating the received pulse's second order statistics are described that make the receiver design implementable.

The MLSE predictor receiver's BER is bounded for the important fast Rayleigh fading, frequency-selective channel model. This analysis shows a good agreement with simulation. A range of BER results have been calculated, revealing that the receivers can exploit the implicit Doppler and delay diversity of fast fading, frequency-selective channels. Many existing receivers either do not exploit the implicit diversity, or exhibit a BER floor.

That the BER floor usually associated with communication over fast fading, frequency-selective channels can be reduced or removed by proper receiver design is the thesis' most significant underlying contribution. New systems and services can be confidently designed for these channels, since sufficient performance can be achieved if adequate computing power is allocated to the receiver structure.

9.2 FUTURE RESEARCH

This thesis can be extended in many directions. The theoretical side can be extended by addressing other than linear modulations, such as CPM. Coding is an important area that has not been considered, yet can significantly enhance the receivers' BERs. Timely information about the channel at the transmitter was assumed unavailable, but this is not the case in slow fading channels. Accordingly, time-varying transmitters and transmitter space diversity can be examined. The latter is an attractive diversity source since in a cellular system it can localise the system complexity at the base station. Bounds on the receivers' BERs have been computed for Rayleigh fading channels only. The pdf of non-zero mean GQFs is also well-known, so the analysis can be extended to Ricean channels. Furthermore, improved algorithms are needed to estimate the time-varying first and second order statistics of a Ricean channel. The channel capacity for fast fading frequency-selective channels
needs to be calculated. The overspread limit is a hard boundary for coherent communications, but what happens when the differentially detected MLSE predictor receiver is employed? The presence of dominant co-channel interferers can be dealt with by jointly detecting both the signal of interest and the interfering signals [12]. A CCI-tolerant receiver is valuable, since the capacity of cellular systems is limited by CCI.

The benefits of directional antennae need to be probed further. When the antennae assumed in this thesis (whose radiation pattern is uniform in the two horizontal directions) are replaced with multiple directional antennae (whose radiation patterns are predominantly localised to a sector), there are two significant advantages. First, the number of paths per antenna reduces as the antenna resolution gets finer: in the limit, each antenna receives no rays or one ray only. In the latter case, there is no fading or delay spread! Second, the influence of CCI is minimised whenever the CCI arrives from a different direction than the received signal. The signal-to-interference ratio at some antennae is lower; but commensurately, at other antennae the signal-to-interference ratio is higher. In the limit as the antenna resolution gets finer, the probability tends to zero that the CCI and the received signal arrive from the same angle. Thus the problem of CCI disappears also. In addition, it is beneficial if both the base station and mobile terminal transmit only through the antenna that detects the strongest received signal. Then signals are radiated over a much smaller region, reducing the CCI to other users. This is essentially a superior form of space division multiplexing, called "directional division multiplexing." Accordingly, the potential benefits of directional antennae are considerable, albeit at the expense of additional complexity.

9.3 FUTURE SYSTEMS

In the following paragraphs, the research described in this thesis is applied as guidelines for developing successful and efficient systems. As can be seen from present trends in mobile communications, future systems must support a dense population of mobile users, each requiring a high data rate. Spectral efficiency is therefore crucial, and high carrier frequencies are ultimately inevitable.
First, there is no reason to avoid fast fading channels by increasing the symbol rate: fast fading does not cause a BER floor, and both Doppler and delay spread offer implicit diversity. In fact the only fundamental constraint on coherent communications is the overspread limit: the delay spread and Doppler spread are merely important descriptive parameters. The high data rate required by future services can be achieved either through one channel, with one carrier frequency, or through many subchannels, with many subcarrier frequencies. Both schemes are amenable to TDMA. The former scheme leads to slow fading and a long delay spread. For very high data rates, the delay spread can be tens or hundreds of symbols long; thus an MLSE receiver is extraordinarily complicated. Furthermore the interleaving required by codes to circumvent the slow fading is substantial. The latter scheme avoids these problems, but its large peak-to-average power is a considerable drawback, as the power amplification is very inefficient (due to the power amplifier’s nonlinear characteristics). Therefore neither scheme is entirely satisfactory. One reasonable approach is to choose the highest practical symbol rate (up to $1/\tau$ or $2/\tau$) as the sub-channel symbol rate, and transmit as many sub-channels as necessary to achieve the required data rate.

Second, the most successful receiver is the EMLSE-\textit{tf} diversity receiver. It can easily accommodate large constellations (e.g. 64-QAM). Its complexity is low when channel state information is available. Providing CSI through pilot tones or symbols is a straightforward method of providing CSI: it is considerably less complicated than estimating at the receiver the received signal’s second order statistics, and it avoids the need for a training sequence. The EMLSE-\textit{tf} diversity receiver’s BER performance is superior: its detection is coherent and it can fully exploit the implicit diversity. Therefore future standards for fast fading channels should include some form of channel sounding.

Third, additional diversity threads considerably improve a receiver’s power efficiency.

Fourth, in order to reduce the rate of hand-offs between cells, some system designs propose to segregate mobile terminals according to their speed. A multi-layered cell structure is employed, with complete sets of macrocells and microcells covering the same area. The fast mobile terminals use the macrocells. However, the
overspread limit is reached more quickly in this arrangement, since it is the fastest terminals that experience the longest delay spreads. Therefore the hand-off problem should be solved by other means when very high carrier frequencies are used.
A LINEAR PREDICTION

A.1 GENERAL PROPERTIES

Linear prediction underpins the MLSE receiver structure derived in Chapter 7. Accordingly it is useful to describe and to investigate the properties of linear predictors in greater detail.

The following theorem is shown. The conditional expectation, \( E(z_i|z_{i-1}, \ldots, z_{i-n}) \), is the output of a linear predictor, where \( z_i \) are samples of a random process that is complex Gaussian, possibly non-stationary, and possibly has a non-zero mean. Thus the conditional expectation can be written as

\[
E(z_i|z_{i-1}, \ldots, z_{i-n}) = \hat{z}_i = \sum_{k=1}^{n} b_k z_{i-k} + b_{n+1} \tag{A.1}
\]

or more simply as

\[
E(z_i|z_{i-1}, \ldots, z_{i-n}) = \hat{z}_i = \sum_{k=1}^{n} b_k z_{i-k} \tag{A.2}
\]
when unity is used in place of \( z_{i-k-1} \)'s actual value. This theorem is a straightforward extension of a result described in [44] for zero-mean complex Gaussian random processes. However, the predictor is only linear when \( z_{i-k-1} \) is regarded as an input (which always equals unity).

The theorem can be shown as follows. The predictor tap weights, \( b_k \), for predicting \( z_i \) are chosen according to two equations. First, the prediction error is constrained to be a zero mean process,

\[
E\left(z_i - \sum_{k=1}^{n} b_k z_{i-k}\right) = E(z_i) - \sum_{k=1}^{n} b_k E(z_{i-k}) = 0 \tag{A.3}
\]

and second, the prediction error is chosen to satisfy the Orthogonality Principle,

\[
\frac{1}{2} E\left(\left(z_i - \sum_{k=1}^{n} b_k z_{i-k}\right) \cdot \bar{z}_{i-m}\right) = \frac{1}{2} E\left(z_i \bar{z}_{i-m}\right) - \sum_{k=1}^{n} b_k \frac{1}{2} E\left(z_{i-k} \bar{z}_{i-m}\right) = 0 \quad m = 1, \ldots, B \tag{A.4}
\]
so that the prediction error, \( z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \), is orthogonal to past samples of the random process, \( z_{l-m} \).

Equations (A.3) and (A.4) provide \( B+1 \) equations for the \( B+1 \) unknowns, \( b_1, \ldots, b_{B+1} \). Equivalently, there are \( B+1 \) complex Gaussian random variables of interest: \( \left( z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \right), z_{l-1}, \ldots, z_{l-B} \). By equation (A.4), the first random variable is uncorrelated with each remaining random variable. Since they are Gaussian random variables, the first random variable is independent of each remaining random variable. Moreover, the first random variable is independent of all the random variables taken together. Therefore

\[
E \left( z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \left| z_{l-1}, \ldots, z_{l-B} \right. \right) = E \left( z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \right) = 0
\]  

(A.5)

where the expectation equals zero from equation (A.3). From the linearity of the expectation and subtraction operators,

\[
E \left( z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \left| z_{l-1}, \ldots, z_{l-B} \right. \right) = E \left( z_i \left| z_{l-1}, \ldots, z_{l-B} \right. \right) - E \left( \sum_{k=1}^{B+1} b_k z_{i-k} \left| z_{l-1}, \ldots, z_{l-B} \right. \right)
\]

(A.6)

but from equation (A.5), this equals zero also. Thus the theorem is proven,

\[
E \left( z_i \left| z_{l-1}, \ldots, z_{l-B} \right. \right) = \sum_{k=1}^{B+1} b_k z_{i-k}
\]  

(A.7)

When the random process is zero mean, equation (A.3) is only satisfied when \( b_{k+1} = 0 \), and the conditional expectation simplifies to the more familiar

\[
E \left( z_i \left| z_{l-1}, \ldots, z_{l-B} \right. \right) = \sum_{k=1}^{B} b_k z_{i-k}
\]  

(A.8)

When the random process is stationary, its mean and autocorrelation depend only on time differences,

\[
E \left( z_i \right) = E \left( z_0 \right)
\]  

(A.9)

\[
\frac{1}{N} E \left( z_i z_{i-m} \right) = R_{z_i z_{i-m}} = R_{z_0 z_{0-m}}
\]  

(A.10)
Thus the expectations in equations (A.3) and (A.4) are independent of the sample, \( l \), being predicted, and

\[
E(z_i) - \sum_{k=1}^{B+1} b_k E(z_{i-k}) = E(z_0) \left( 1 - \sum_{k=1}^{B+1} b_k \right)
\]

(A.11)

\[
\frac{1}{2} E(z_i z_{i-m}) - \sum_{k=1}^{B+1} b_k \frac{1}{2} E(z_{i-k} z_{i-m}) = R_{zz,m} - \sum_{k=1}^{B+1} b_k R_{z,m-k}
\]

(A.12)

Therefore the same predictor tap weights are used to predict all samples of the stationary random process.

The predictor tap weights that satisfy equations (A.3) and (A.4) estimate the random process with minimum mean square prediction error. The minimum mean square prediction error occurs when

\[
\frac{\partial}{\partial b_m} \left( \frac{1}{2} E \left( \left| z_i - \sum_{k=1}^{B+1} b_k z_{i-k} \right|^2 \right) \right) = 0 \quad m = 1, \ldots, B + 1
\]

(A.13)

or equivalently when

\[
\frac{1}{2} E(z_i z_{i-m}) \frac{\partial \overline{b}_m}{\partial b_m} - \frac{1}{2} E(z_{i-m} z_i) + \sum_{k=1}^{B+1} b_k \frac{1}{2} E(z_{i-k} z_{i-m}) \frac{\partial \overline{b}_m}{\partial b_m} + \sum_{k=1}^{B+1} b_k \frac{1}{2} E(z_{i-m} z_{i-k}) = 0
\]

\[
\text{m} = 1, \ldots, B + 1
\]

(A.14)

which has the solution

\[
\frac{1}{2} E(z_i z_{i-m}) = \sum_{k=1}^{B+1} b_k \frac{1}{2} E(z_{i-k} z_{i-m}) \quad m = 1, \ldots, B + 1
\]

(A.15)

Identifying \( z_{i,B+1} \) as unity, this condition mimics the two conditions of equations (A.3) and (A.4),

\[
E(z_i) - \sum_{k=1}^{B+1} b_k E(z_{i-k}) = 0
\]

(A.16)

\[
\frac{1}{2} E(z_i z_{i-m}) - \sum_{k=1}^{B+1} b_k \frac{1}{2} E(z_{i-k} z_{i-m}) = 0 \quad m = 1, \ldots, B
\]

(A.17)

The MMSE equals
A LINEAR PREDICTION

Given that $B$ extends back to the start of transmission, the prediction error, $z_l - \sum_{k=1}^{B+1} b_k z_{l-k}$, is uncorrelated with all previous process samples. When scaled by the root mean square prediction error,

$$\sqrt{\frac{1}{B+1} E\left(\left| z_l - \sum_{k=1}^{B+1} b_k z_{l-k}\right|^2 \right)}$$

the resulting sequence is known as the Innovations Process. It is a white noise process, since it is zero mean, uncorrelated with past process samples, and its power is
constant. The adjective "innovations" is used since the sequence comprises the unpredictable part of the random process.

The predictor's structure is shown in figure A.1 for both non-zero mean and zero mean random processes. New samples of the random process are shifted into the tapped delay line. The prediction is a weighted sum of these past samples.

### A.2 PERFECT PREDICTION

With the appropriate predictor tap weights, a predictor with $B$ taps can perfectly predict a $B$-1 order polynomial. More generally, if a random process obeys a $B$-1 order polynomial model over $t = [t_b; t_E]$ and the predictor's taps are spaced every $T_r$ sec, then the random process can be perfectly predicted over $t = [t_b + BT_r; t_E]$. The predictor computes

$$\hat{z}_i = \sum_{k=1}^{B} b_k z_{i-k}$$

(A.20)

where the predictor tap weights are given by

$$b_k = (-1)^{k-1} C_k^B = (-1)^{k-1} \frac{B!}{(B-k)!k!}$$

(A.21)

Defining the $B$-1 order polynomial around a reference time, $t_R = (l - B/2)T_r$, as

$$z_i = \sum_{n=0}^{B-1} x_n (IT_r - t_R)^n$$

(A.22)

where $x_n$ are the polynomial coefficients, the predictor calculates

$$\hat{z}_i = \sum_{k=1}^{B} (-1)^{k-1} C_k^B \sum_{n=0}^{B-1} x_n (IT_r - t_R - kT_r)^n$$

$$= \sum_{n=0}^{B-1} x_n \sum_{k=1}^{B} (-1)^{k-1} C_k^B \sum_{p=0}^{n} C_p^n (IT_r - t_R)^p (-kT_r)^{n-p}$$

(A.23)

$$= \sum_{n=0}^{B-1} x_n C_p^n \sum_{p=0}^{n} (IT_r - t_R)^p T_r^{n-p} \sum_{k=1}^{B} (-1)^{k-1} C_k^B (-k)^{n-p}$$

However it is known that
\[
\sum_{k=1}^{B} (-1)^{k-1} C_k^B (-k)^{m-p} = \begin{cases} 
0 & m - p \in \{1, \ldots, B-1\} \\
1 & m = p 
\end{cases} \quad (A.24)
\]

so that the prediction is indeed perfect,

\[
\hat{z}_j = \sum_{m=0}^{B-1} x_m (l_{T_e} - t_R)^m = z_j \quad (A.25)
\]

This result is interesting when viewed with the work in [2] on power series models. The same random process is expanded in a Taylor series around the same reference point,

\[
z_j = \sum_{k=0}^{B-1} \frac{z^{(k)}(T_e)}{k!} (l_{T_e} - t_R)^k + r_R \quad (A.26)
\]

where \( r_R \) is the error due to truncating the Taylor series at the \( (B-1) \)th term.

The \( t \)-power series model of \( T(f,t) \) can be used to show that the Taylor series rapidly converges over an interval around \( t_R \). Considering the special case, \( T(f,t) = T(0,l_{T_e}) = z_h \), [2] reports that the error term rapidly decays to zero as \( B \) gets large, when

\[
2\pi(2f_d)BT_e << 1 \quad (A.27)
\]

where \( 2f_d \) is the random process’s two sided bandwidth and \( BT_e \) is the duration where convergence is required.

Therefore a bandlimited random process can be expressed as a polynomial expansion, with an error that decreases to zero as the polynomial’s order increases. Since a length \( B \) predictor can perfectly predict a polynomial of order \( B-1 \), a bandlimited random process can be predicted with an error that decreases to zero as the predictor length, \( B \), increases to infinity.

That is, bandlimited random processes can be predicted with an arbitrarily low error, so long as the predictor is long enough and the sampling rate is high enough. Alternatively, if the predictor is too short or the sampling rate inadequate, then there is a prediction error floor.

This problem appears in the literature. It is seen most obviously in [17]. The FIR filter does not generate a bandlimited random process, so it cannot be perfectly predicted. As expected, the prediction error floor reduces as the sampling rate increases. In the fading channel, a prediction error floor leads to a BER floor. Since
the BER curves [15] all have the same predictor-length to samples-per-symbol ratio, it is unclear immediately whether the BER floor is due to an inadequate sampling rate or insufficient predictor taps. Further simulations reveal that the BER floor is substantially reduced by increasing the predictor length.

Inadequate prediction also plagues the receiver performance in [14, 64]. The receiver principle is as follows. The received signal is sampled at its Nyquist rate. Each received signal sample expresses a snapshot of the channel. Treating each snapshot as a state, the channel can be regarded as following a series of states. A state space model of the channel arises, and the optimum receiver structure uses Kalman filters to estimate the channel and its correlation matrices. However, the receiver structure appears to exhibit a BER floor in simulations [14, 64].

Here a first order AR model of a zero-mean channel is assumed. The tap autocorrelation decays exponentially,

$$R_n(\Delta r) = \exp(-f_D|\Delta r|)$$  \hspace{1cm} (A.28)

where $f_D$ is some measure of the Doppler spread. The Doppler spectrum is not bandlimited,

$$S_x(\Delta f) = \frac{2f_D}{f_D^2 + (2\pi f)^2}$$  \hspace{1cm} (A.29)

and so the received signal is not bandlimited either. Any finite sampling rate is insufficient, leading to a prediction error floor and hence a BER floor. It is this problem that plagues [14, 64] and causes the BER floor; their channel model violates their sampling model.

In the t-channel with constant envelope signalling [37, 17, 61], the receiver predicts the channel directly, albeit corrupted by additive noise. The channel’s second order statistics are needed to compute the MMSE predictor tap weights. A sub-optimum design can use the predictor tap weights of equation (A.21) instead. Since they are calculated for predicting the noiseless fading process, they are asymptotically accurate, and there is no BER floor if the sampling rate and predictor length are chosen large enough. However, there is some noise enhancement at low SNR. For instance the MMSE predictor tap weights for a slow fading channel are approximately $1/B$, so the noise power in the prediction is $1/B$ the noise power in any one sample.
For the predictor tap weights of equation (A.21), the noise enhancement is given by
\[ \sum_{i=1}^{B} (C_{i}^{a})^2. \] For \( B = 2 \), there is a 10dB penalty; for \( B = 3 \), the penalty is 17.6dB.
Gaussian quadratic forms (GQF) occur extensively in the analysis of complex Gaussian fading channels and noise, since the metric usually involves a Euclidean (squared) distance. In this appendix, complex methods for inverting the characteristic function are described in more detail.

The BER can be bounded by considering pairwise error probabilities: namely the probability that one sequence's metric exceeds the transmitted sequence's metric. In complex Gaussian fading channels, both metrics are GQFs, and so too is their difference. When the channel is unknown, a deterministic bias term is also present. The pairwise error probability is calculated by integrating the pdf of the metric difference over the positive axis. Therefore the pdf of a GQF is required.

A real GQF, \( \kappa \), is given by

\[
\kappa = g^H G g
\]  

(B.1)

where \( g \) is a vector of zero mean, i.i.d complex Gaussian random variables with autocorrelation

\[
R_{gg} = \frac{1}{4} E(g g^H)
\]  

(B.2)

and \( G \) is a Hermitian symmetric matrix that specifies which quadratic terms are present in the GQF. Calculating the GQFs pdf is achieved by exploiting the pdf's characteristic function. Given a pdf, \( p_\kappa(\kappa) \), its characteristic function, \( P_\xi(\xi) \), is given by the integral transform,

\[
P_\xi(\xi) = \int p_\kappa(\kappa) \exp\{j\xi \kappa\} d\kappa
\]  

(B.3)

which is closely allied to the Fourier Transform. The pdf is computed from the inverse transform of the characteristic function,

\[
p_\kappa(\kappa) = \frac{1}{2\pi} \int P_\xi(\xi) \exp\{-j\xi \kappa\} d\xi
\]  

(B.4)

A general result for the characteristic function of a Gaussian quadratic form is well known [55, 50]
\[ P_\zeta (\zeta) = \frac{1}{\text{det}[1 - j2\zeta \mathbf{R}_{ss} G]} \]  
(B.5)

For an \( I \times I \) \( \mathbf{R}_{ss} G \) matrix, the determinant is an \( I \)th order polynomial in \( \zeta \). The characteristic function can be written as the product,

\[ P_\zeta (\zeta) = \prod_{i=1}^{I} \frac{1}{1 - \zeta / p_i} \]  
(B.6)

where the poles, \( p_i \), are zeros of the determinant.

The pairwise probability of error is the probability that \( \kappa > \kappa_{\text{min}} \),

\[
\int_{\kappa_{\text{min}}}^{\infty} p_\kappa (\kappa) d\kappa = \int_{\kappa_{\text{min}}}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} P_\zeta (\zeta) \exp\{-j\zeta \kappa\} d\zeta d\kappa
\]

\[ = \int_{\kappa_{\text{min}}}^{\infty} \frac{1}{2\pi} P_\zeta (\zeta) \int_{-\infty}^{\infty} \exp\{-j\zeta \kappa\} d\kappa d\zeta \]

\[ = \int_{-\infty}^{\infty} P_\zeta (\zeta) \left( \pi \delta (\zeta) + \frac{\exp(-j\zeta \kappa_{\text{min}})}{j\zeta} \right) d\zeta \]

\[ = \frac{1}{2} + \int_{-\infty}^{\infty} \frac{P_\zeta (\zeta) \exp(-j\zeta \kappa_{\text{min}})}{j2\pi \zeta} d\zeta \]  
(B.7)

which is a line integral along the real axis.

Figure B.1(a): Contour of integration for \( \kappa_{\text{min}} \leq 0 \)

Figure B.1(b): Contour of integration for \( \kappa_{\text{min}} \geq 0 \)
It is useful to define a contour that encompasses this interval, such as figure B.1(a). It extends from -R to positive R, with a semicircle of radius δ around the origin to avoid the pole there, then around the positive imaginary plane in a semicircle of radius R (figure B.1(a)). R then tends to infinity and δ tends to zero. The Residue Theorem states that the integral around the contour in an anti-clockwise direction is $2\pi j$ multiplied by the sum of the residues, $r_i$, at the poles, $p_i$, enclosed by the contour.

An $n_i$th order residue equals

$$\text{Res}_i \{ X(\zeta) \} = r_i = \lim_{\zeta \to p_i} \frac{d^{(n_i-1)}}{d\zeta^{(n_i-1)}} \left( (\zeta - p_i)^{n_i} X(\zeta) \right)$$

The contribution from the large semi-circle tends to zero if there exists a $k > 1$ and an $M$ such that the modulus of the integrand is bounded above by $MR^k$. The characteristic function is an inverse polynomial with an order that at least exceeds quadratic, so this is satisfied as long as $\exp(-j\kappa_{\min})$ is bounded. For $\kappa_{\min} \leq 0$, the exponential is of the form

$$\left| \exp(-j\kappa_{\min}) \right| = \left| \exp(-j Re \kappa_{\min}) \right|, \quad 0 \leq 0 \leq \pi$$

$$= \left| \exp(\kappa_{\min} R \sin \theta) \right|, \quad 0 \leq 0 \leq \pi$$

$$\leq 1$$

which is indeed bounded. The contribution from the small semicircle is given by

$$\oint_{\delta \to 0} \frac{P_\kappa(\delta \exp(j\theta)) \exp(-j\delta \exp(j\theta)\kappa_{\min})}{\sqrt{2\pi \delta \exp(j\theta)}} d\theta = \int_{0}^{\pi} \frac{1}{\sqrt{2\pi}} d\theta = -\frac{1}{2}$$

and so from equation (B.7), the pairwise probability of error equals

$$\int P_\kappa(\kappa) d\kappa = 1 + \sum_{\kappa_{\min}} \text{Res}_i \left\{ \frac{P_\kappa(\zeta) \exp(-j\zeta\kappa_{\min})}{\zeta} \right\}$$

However, the proposed contour is inadequate for $\kappa_{\min} > 0$. A suitable contour is shown in figure B.1(b). This contour is traversed in clockwise direction, so the integral equals $-2\pi j$ multiplied by the sum of the residues enclosed by the contour. Again we must show that $\exp(-j\zeta\kappa_{\min})$ is bounded. For $\kappa_{\min} \geq 0$, the exponential is of the form
\( |\exp(-j\zeta_{\kappa_{\min}})| = |\exp(-j\Re\Theta_{\kappa_{\min}})|, \quad \pi \leq 0 \leq 2\pi \)
\( = |\exp(\kappa_{\min}R\sin\theta)|, \quad \pi \leq 0 \leq 2\pi \)
\( \leq 1 \) \hspace{1cm} (B.12)

which is indeed bounded. The contribution from the small semicircle is given by

\[
\gamma_0 \int_{-\pi}^{\pi} \frac{P_z(\delta \exp(j\theta)) \exp(-j\delta \exp(j\theta)\kappa_{\min})}{j2\pi \delta \exp(j\theta)} j\delta \exp(j\theta) d\theta = \frac{1}{2\pi} d\theta = \frac{1}{2} 
\]

and so from equation (B.7), the pairwise probability of error equals

\[
\int_{\kappa_{\min}}^{\infty} p_z(\kappa) d\kappa = -\sum_{i=1}^{i} \Re s \left\{ P_z(\zeta) \exp(-j\zeta_{\kappa_{\min}}) / \zeta \right\} \hspace{1cm} (B.14)
\]

In many instances the poles must be located numerically. This is related to the eigenvalue problem, and most numerical packages have library routines for this task. In particular, the eigenvalue problem is

\[
\left| R_{xx} G - \lambda I \right| = 0 \hspace{1cm} (B.15)
\]

so the poles of the characteristic function are given by \( p_i = -j/2\lambda_i \). However, the eigenvalue calculation can be numerically inaccurate, particularly for large \( I \). Proceeding regardless, erroneous and even negative pairwise probabilities arise. Instead, only the significant (i.e. greater that \( \lambda \times \text{machine epsilon} \times \lambda_{\max} \)) eigenvalues should be retained. Thus fewer than \( I \) poles and residues are used to calculate the BER.

Multiple poles arise whenever there are \( D \) ideally independent diversity threads. Then the random variables from each thread are independent, and the optimal metric is the sum of each thread’s metric. Thus both \( R_{xx} \) and \( G \) are only non-zero on \( D \) identical diagonal blocks. The determinant in equation (B.5) equals the product of the block determinants, so every pole is repeated \( D \) times.

However, from equation (4.35), ideal independence only arises when there is no Doppler spread and the antennae are located at the zeros of the spatial autocorrelation function, or when the antennae are infinitely spaced. Neither of these cases are important, since the union bound technique is not tight when the Doppler spread gets small, and infinitely spaced antennae cannot be used in practice. Notwithstanding, the analysis is undertaken for the case of \( D = 2, 3 \), as well as for the
case of simple poles. When the threads are not ideally independent, the simple pole analysis applies.

When all poles in equations (B.11) and (B.14) are simple, the residues are calculated as

\[ r_i = \lim_{\zeta \to p_i} \left( \zeta - p_i \right) \frac{d}{d\zeta} \left( \zeta - p_i \right) \exp\left( -j\zeta \kappa_{\text{min}} \right) / \zeta = -\exp\left( -j p_i \kappa_{\text{min}} \right) \prod_{\substack{k \in I \setminus \{i\}}} \frac{1}{\left( 1 - p_i / p_k \right)} \]

(B.16)

When all poles in equations (B.11) and (B.14) are double, the residues are calculated as

\[
\begin{align*}
r_i &= \lim_{\zeta \to p_i} \frac{d^2}{d\zeta^2} \left( \zeta - p_i \right)^2 \frac{d}{d\zeta} \exp\left( -j\zeta \kappa_{\text{min}} \right) / \zeta \\
&= -\left( 1 + j p_i \kappa_{\text{min}} \right) \exp\left( -j p_i \kappa_{\text{min}} \right) \prod_{\substack{k \in I \setminus \{i\}}} \frac{1}{\left( 1 - p_i / p_k \right)^2} \\
&\quad + p_i \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} \frac{1}{p_i \left( 1 - p_i / p_l \right)^3} \prod_{\substack{k \in I \setminus \{i,l\}}} \frac{1}{\left( 1 - p_i / p_k \right)^2} \\
&= \left( 1 + j p_i \kappa_{\text{min}} - \frac{1}{2} p_i^2 \kappa_{\text{min}}^2 \right) \exp\left( -j p_i \kappa_{\text{min}} \right) \prod_{\substack{k \in I \setminus \{i\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad + p_i \left( 1 + j p_i \kappa_{\text{min}} \right) \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 3 \frac{1}{p_i \left( 1 - p_i / p_l \right)^4} \prod_{\substack{k \in I \setminus \{i,l\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad - \frac{1}{2} p_i^2 \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 12 \frac{1}{p_i^5 \left( 1 - p_i / p_l \right)^5} \prod_{\substack{k \in I \setminus \{i,l\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad - \frac{1}{2} p_i^2 \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 3 \frac{1}{p_i^4 \left( 1 - p_i / p_l \right)^4} \sum_{\substack{m \in I \setminus \{i,l\}}} 3 \frac{1}{p_m^3 \left( 1 - p_i / p_m \right)^4} \prod_{\substack{k \in I \setminus \{i,l,m\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3}
\end{align*}
\]

(B.17)

When all poles in equations (B.11) and (B.14) are triple, the residues are calculated as

\[
\begin{align*}
r_i &= \lim_{\zeta \to p_i} \frac{d^3}{d\zeta^3} \left( \zeta - p_i \right)^3 \frac{d}{d\zeta} \exp\left( -j\zeta \kappa_{\text{min}} \right) / \zeta \\
&= -\left( 1 + j p_i \kappa_{\text{min}} - \frac{3}{2} p_i^2 \kappa_{\text{min}}^2 + \frac{1}{6} p_i^3 \kappa_{\text{min}}^3 \right) \exp\left( -j p_i \kappa_{\text{min}} \right) \prod_{\substack{k \in I \setminus \{i\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad + p_i \left( 1 + j p_i \kappa_{\text{min}} \right) \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 3 \frac{1}{p_i \left( 1 - p_i / p_l \right)^4} \prod_{\substack{k \in I \setminus \{i,l\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad - \frac{1}{2} p_i^2 \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 12 \frac{1}{p_i^5 \left( 1 - p_i / p_l \right)^5} \prod_{\substack{k \in I \setminus \{i,l\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad - \frac{1}{2} p_i^2 \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 18 \frac{1}{p_i^4 \left( 1 - p_i / p_l \right)^4} \sum_{\substack{m \in I \setminus \{i,l\}}} 3 \frac{1}{p_m^3 \left( 1 - p_i / p_m \right)^4} \prod_{\substack{k \in I \setminus \{i,l,m\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3} \\
&\quad - \frac{1}{2} p_i^2 \exp\left( -j p_i \kappa_{\text{min}} \right) \sum_{\substack{l \in I \setminus \{i\}}} 3 \frac{1}{p_i^3 \left( 1 - p_i / p_l \right)^3} \sum_{\substack{m \in I \setminus \{i,l\}}} 3 \frac{1}{p_m^2 \left( 1 - p_i / p_m \right)^3} \sum_{\substack{n \in I \setminus \{i,l,m\}}} 3 \frac{1}{p_n \left( 1 - p_i / p_n \right)^3} \prod_{\substack{k \in I \setminus \{i,l,m,n\}}} \frac{1}{\left( 1 - p_i / p_k \right)^3}
\end{align*}
\]

(B.18)
C. THE SEQUENCE-AVERAGED PREDICTOR RECEIVER

C.1 INTRODUCTION

In early research, this author could not devise a viable method for estimating the channel's second order statistics every symbol period, whichever sequence was transmitted. Accordingly, a "sequence-averaged" predictor receiver structure was conceived, where "sequence-averaged second order statistics" are computed each symbol period. For simplicity, $M$-PSK, a Rayleigh fading channel, one thread, ideal carrier recovery and symbol timing are assumed: $f_0_{nom} = 0; \phi = 0; t_0 = 0$. The receiver structure is described here, but its importance is questionable. It is not an MLSE receiver structure, so its theoretical value is slight. Its complexity is high, so it is not suited to a practical implementation either. Finally, its BER performance is considerably worse than the MLSE predictor receiver.

The MLSE receiver structures for purely frequency-selective channels [10], and purely time selective channels using $M$-PSK and rectangular pulses were known [61]. The sequence-averaged receiver structure arises from unifying these two divergent receiver structures. Accordingly the receiver synthesis begins by identifying their commonality. In white noise, both use implicitly or explicitly a Euclidean branch metric,

$$\lambda_n(y|\hat{\beta}) = \sum_{n=p}^{(n+1)-1} \frac{\left| y_n - E\left(y_n|\hat{\beta}, y_{n-1}\right) \right|^2}{\sigma^2(y_n|\hat{\beta}, y_{n-1})} + \ln\sigma^2\left(y_n|\hat{\beta}, y_{n-1}\right)$$  \hspace{1cm} (C.1)

where

$$y_{n-1} = \begin{bmatrix} \cdots & y_{n-2} & y_{n-1} \end{bmatrix}^T$$  \hspace{1cm} (C.2)

and

$$\sigma^2\left(y_n|\hat{\beta}, y_{n-1}\right) = \sigma^2\left(y_n|y_{n-1}\right)$$  \hspace{1cm} (C.3)

for purely frequency-selective channels and purely time selective channels using $M$-PSK and rectangular pulses.
The purely time selective channel with rectangular pulses forms

\[
E(y_n|\beta, y_{n-1}) = \beta_{\{n \rightarrow r \rightarrow \beta\}} \sum_{k=1}^{p} b_k \bigg( \beta_{\{n \rightarrow r \rightarrow \beta\}} r_{\beta} y_{n-k} + n_{n-k} \bigg)
\]

\[
= \beta_{\{n \rightarrow r \rightarrow \beta\}} \sum_{k=1}^{p} b_k \left( z_{n-k} + \bar{\beta}_{\{n \rightarrow r \rightarrow \beta\}} r_{\beta} n_{n-k} \right)
\]

as the MMSE prediction when the hypothesised sequence, \{\hat{\beta}\}, matches the transmitted sequence \{\beta\}. The important properties of this formulation are threefold.

The receiver (data dependently) unrotates the received signal to get an estimate of the underlying channel process, \(z_{n-k}\). It takes the (data-independent) weighted sum of these estimates, where the only the predictor tap weights characterise the channel statistics. The final result is (data dependently) rotated by the hypothesised symbol.

The purely frequency-selective channel forms,

\[
E(y_n|\beta, y_{n-1}) = \sum_{i=-L+1}^{L} \sum_{j=1}^{n} \beta_{\{n,m \rightarrow L\rightarrow \beta\}} \hat{c}_{n-m,0} = \sum_{i=-L+1}^{L} \sum_{j=1}^{n} \beta_{\{n,m \rightarrow L\rightarrow \beta\}} \hat{c}_{n-m,0}
\]

where the received pulse shape obeys

\[
y_{\{n,r\}} = \begin{bmatrix}
\beta_{\{n,L+1\rightarrow \beta\}} & \cdots & \beta_{0} \\
\vdots & & \vdots \\
\beta_{\{n,L+1\rightarrow \beta\}} & \cdots & \beta_{0} \\
\end{bmatrix} \begin{bmatrix}
c_{\{n,L+1\rightarrow \beta\},0} \\
\vdots \\
c_{\{n,L+1\rightarrow \beta\},0} \\
\end{bmatrix} + \begin{bmatrix}
n_{0} \\
\vdots \\
n_{n_L} \\
\end{bmatrix}
\]

or

\[
y_{\{n,r\}} = \beta \cdot c + n
\]

The MMSE estimate of the \(Lr\) unknowns, \(c\), comes from least squares theory, as

\[
\hat{c} = \left( \beta \beta^H \right)^{-1} \beta^H y_{\{n,r\}}
\]
In this instance, a particular form of the receiver is sought. Observing that \( \beta_f \) can be written as
\[
\beta_f = \beta \begin{bmatrix} I_{L_r, L_r} \\
\vdots \\
I_{L_r, L_r} \end{bmatrix} = \beta F
\]
where
\[
\beta = \text{diag} \left( \beta_{(-L+1)r}, \ldots, \beta_{0}, \ldots, \beta_{Lr}, \ldots, \beta_{(L+1)n_r} \right)
\]
then the MMSE estimate of \( c_f \) equals
\[
\hat{c}_f = \left( F^H \beta^H \beta F \right)^{-1} F^H \beta^H y
\]

With this representation, estimating \( E(y_n | \beta, y_{n-1}) \) has the same threefold structure as for the time-selective case. The received signal is (data dependently) unrotated by the hypothesised signal, according to \( \beta^H y \),
\[
\overline{\beta}_{kr} y_n = c_{n-kr, 0} + \sum_{i=-[n/r]-1}^{[n/r]} \beta_{kr} c_{n-i r, 0} + \overline{\beta}_{kr} n_n \quad k = -L + 1 + \lfloor n/r \rfloor, \ldots, \lfloor n/r \rfloor
\]
The first term is one of the \( L_r \) received pulse unknowns, corrupted by additive terms only: received pulse tails from adjacent symbols and noise. Thus \( \beta^H y \) generates a clean estimate for every received pulse tail present in each received sample.

![Figure C.1: Operation of sequence-averaged prediction. Predicting \( y_n \) requires that \( L \) pulse tails be predicted (the solid circles). The dashed enclosure denotes that each pulse tail is predicted using all \( L \) pulse tails in the past \( B \) samples (the empty circles). Assuming independent phasors and many samples, \( (F^H \beta^H \beta F)^{-1} F^H \) becomes data independent. However, the metric is conditioned on the phasors, so the assumption of...](image-url)
assumption of independent phasors cannot be justified. Proceeding regardless, the \((\mathbf{F}^H\mathbf{\beta}\mathbf{F})^{-1}\mathbf{F}^H\) factor plays the same role as the predictor for the time-selective case: it (data independently) predicts the received pulses. The channel estimates are then combined in a (data dependent) way, to form \(E(y_n|\hat{\beta}, y_{n-1})\).

### C.2 THE SEQUENCE-AVERAGED PREDICTOR RECEIVER

The sequence-averaged predictor receiver’s metric adopts this threefold structure also:

- Past samples are unrotated by the hypothesised phasor sequence \(L\) times, to get clean estimates of the \(L\) received pulse tails in the sample.

- The clean estimates are used to predict the received pulse tails in the current sample, in a least squares, data-independent manner.

- The predicted received pulse tails are combined according to the hypothesised symbol sequence, forming \(\hat{E}(y_n|\hat{\beta}, y_{n-1}) \approx E(y_n|\hat{\beta}, y_{n-1})\).

The second step is the source of sub-optimality, but it centralises the receiver’s knowledge of the channel statistics, so the predictor tap weights can be updated on a symbol-by-symbol basis. The prediction arrangement is represented in figure C.1.

The branch metric of equation (C.1) also requires the sequence-averaged minimum mean square error (SAMMSE), \(\hat{\sigma}^2(y_{[n/r]}|y_{[(n-1)/r]} \approx \sigma^2(y_n|\hat{\beta}, y_{n-1})\). Having averaged across phasor sequences, the SAMMSE is independent of the phasor sequence, \(\{\beta\}\), so the \(\ln \hat{\sigma}^2(y_{[n/r]}|y_{[(n-1)/r]} \) term in the branch metric can be neglected. However, the SAMMSE varies with each sample within a symbol interval, so they are required to weight each sample’s Euclidean distance, as

\[
\lambda_r(y|\hat{\beta}) = \sum_{n=ir}^{(i+1)r-1} \frac{\hat{E}(y_n|\hat{\beta}, y_{n-1})^2}{\hat{\sigma}^2(y_{[n/r]}|y_{[(n-1)/r]})} \tag{C.13}
\]
The SAMMSE predictor taps weights minimise the sequence-averaged mean square error (SAMSE),

\[
\text{SAMSE} = \frac{1}{2} E \left[ y_n - \sum_{k=0}^{\lceil n/r \rceil} \beta_{kn} \sum_{l=1}^{B} \sum_{m=-L+1}^{L} b_{klmn} \beta_{mv} y_{n-l} \right]^2
\]  
\text{(C.14)}

where the expectation is taken over both the received pulse shape and the transmitted phasors. The predictor tap weights \( b_{klmn} \) are indexed as follows:

- To form \( \hat{E}(y_n|\beta, y_{n-1}) \), \( L \) received pulse tails are required. Therefore there are \( L \) distinct predictors, indexed by \( k \).
- The predictor uses \( B \) past samples of the received signal. \( l \) indexes the sample under consideration, \( y_{n-l} \).
- Each received sample is unrotated \( L \) different times by the transmitted symbols. This unrotation is indexed by \( m \).
- The received pulse shapes repeat every symbol, so only the \( r \) samples within a symbol period require different predictors. These are indexed by \( \lfloor n/r \rfloor \).

In summary, there are \( Lr \) predictors, each with \( BL \) taps.

To minimise the SAMSE, each \( b \) term is differentiated in turn with respect to \( b_{stuv} \), where \( stuv \) encompass every \( b \) term, i.e. \( s = -L+1+\lfloor v/r \rfloor, \ldots, \lfloor v/r \rfloor; t = 1, \ldots, B; u = -L+1+\lfloor (v-t)/r \rfloor, \ldots, \lfloor v/r \rfloor; v = 0, \ldots, r-1. \)

\[
\frac{d(\text{SAMSE})}{db_{stuv}} = \\
+ \sum_{k=0}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \sum_{m=-L+1}^{L} \frac{dB_{klnm}}{dB_{stuv}} \left[ \sum_{k=0}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \sum_{m=-L+1}^{L} b_{klmn} \frac{1}{2} E \left( \beta_{kn} \bar{\beta}_{mv} \beta_{mv} \beta_{mv} y_{n-l} y_{n-l} \right) \right] \\
- \frac{1}{2} E \left( \beta_{mv} \bar{\beta}_{mv} \beta_{mv} y_{n-r} \right) \\
+ \sum_{k=0}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \sum_{m=-L+1}^{L} \frac{DB_{klnm}}{DB_{stuv}} \left[ \sum_{k=0}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \sum_{m=-L+1}^{L} b_{klmn} \frac{1}{2} E \left( \beta_{kn} \bar{\beta}_{mv} \beta_{mv} y_{n-l} y_{n-l} \right) \right] \\
- \frac{1}{2} E \left( \beta_{mv} \bar{\beta}_{mv} \beta_{mv} y_{n-r} \right)
\]  
\text{(C.15)}
which equals zero for a minimum. This occurs when

\[
\sum_{k=1}^{[v/r]} \sum_{l=1}^{[v-t)/r]} \sum_{n=1}^{B} \sum_{m=1}^{B} \tilde{b}_{klmn} \frac{1}{2} E(\beta_{kr} \tilde{\beta}_{mr} \beta_{vr} y_{v-t} \bar{y}_{v-t}) = \frac{1}{2} E(\tilde{\beta}_{kr} \beta_{mr} y_r \bar{y}_{v-t})
\]  
(C.16)

for \( s = -L+1 + \lfloor v/r \rfloor, \ldots, \lfloor v/r \rfloor \); \( t = 1, \ldots, B \); \( u = -L+1 + \lfloor (v-t)/r \rfloor, \ldots, \lfloor (v-t)/r \rfloor \); \( v = 0, \ldots, r-1 \). Equation (C.16) is a system of linear equations, the normal equations, where there are \( BL^2r \) equations and \( BL^2r \) unknowns. However the system can be underdetermined, so the singular value decomposition (SVD) [41] should be used to calculate \( b_{klmn} \). Since the SVD is an iteratively computed \( O(B^3r^3) \) decomposition, this system is very expensive to solve.

The SAMMSE equals

\[
\hat{\sigma}^2(Y_{\lfloor v/r \rfloor}, Y_{\lfloor (v-t)/r \rfloor}) = \frac{1}{2} E\left( y_r - \sum_{k=1}^{[v/r]} \sum_{l=1}^{[v-t)/r]} \sum_{n=1}^{B} \sum_{m=1}^{B} \tilde{b}_{klmn} \frac{1}{2} E(\beta_{kr} \beta_{mr} y_{v-t} \bar{y}_{v-t}) \right)^2 + \frac{1}{2} E\left( \sum_{k=1}^{[v/r]} \sum_{l=1}^{[v-t)/r]} \sum_{n=1}^{B} \sum_{m=1}^{B} \tilde{b}_{klmn} \frac{1}{2} E(\beta_{kr} \beta_{mr} y_{v-t} \bar{y}_{v-t}) \right) + \frac{1}{2} E\left( \sum_{k=1}^{[v/r]} \sum_{l=1}^{[v-t)/r]} \sum_{n=1}^{B} \sum_{m=1}^{B} \tilde{b}_{klmn} \frac{1}{2} E(\beta_{kr} \beta_{mr} y_{v-t} \bar{y}_{v-t}) \right) 
\]  
(C.17)

which can simplified by using equation (C.16), as

\[
\hat{\sigma}^2(Y_{\lfloor v/r \rfloor}, Y_{\lfloor (v-t)/r \rfloor}) = \frac{1}{2} E\left( y_r^2 \right) - 2\Re\left( \sum_{k=1}^{[v/r]} \sum_{l=1}^{[v-t)/r]} \sum_{n=1}^{B} \sum_{m=1}^{B} \tilde{b}_{klmn} \frac{1}{2} E(\beta_{kr} \beta_{mr} y_{v-t} \bar{y}_{v-t}) \right) 
\]  
(C.18)

but this simplified expression can be inaccurate, since equation (C.16) is solved imprecisely through a pseudoinverse. Negative SAMMSEs can arise. Accordingly it is more reliable to use equation (C.17) to guarantee the validity of the SAMMSE.

The quantities \( \frac{1}{2} E(y_r^2) \), \( \frac{1}{2} E(\beta_{kr} \beta_{mr} y_r \bar{y}_{v-t}) \), and \( \frac{1}{2} E(\beta_{kr} \beta_{mr} y_r \bar{y}_{v-t}) \) are the "sequence-averaged" second order statistics. They are given by
\[
\frac{1}{2} E\left\{ |y_1|^2 \right\} = \sum_{\ell = k - L}^{k} \sum_{r = \ell}^{\ell + L} E(\beta_{\ell \rho} \overline{\beta}_{\ell \rho}) \frac{1}{2} E\left( c_{\ell - \ell \rho}, \overline{c}_{\ell - \ell \rho} \right) + \frac{1}{2} E(n, \overline{n}) \quad (C.19)
\]

\[
\frac{1}{2} E(\overline{\beta}_{\ell \rho} \beta_{\ell \rho}, \overline{y}_{\ell - \ell \rho}) = E(\overline{\beta}_{\ell \rho} \beta_{\ell \rho}) \frac{1}{2} E(n, \overline{n}) \quad (C.20)
\]

\[
\frac{1}{2} E(\beta_{\ell \rho} \overline{\beta}_{\ell \rho}, \overline{y}_{\ell - \ell \rho}) = E(\beta_{\ell \rho} \overline{\beta}_{\ell \rho}) \frac{1}{2} E(n, \overline{n}) \quad (C.21)
\]

where the phasor expectations equal

\[
E(\overline{\beta}_{\ell \rho} \beta_{\ell \rho}) = \delta_{\ell \rho, \ell \rho} \quad (C.22)
\]

\[
E(\beta_{\ell \rho} \overline{\beta}_{\ell \rho}, \beta_{\ell \rho}, \overline{\beta}_{\ell \rho}) = \delta_{\ell \rho, \ell \rho} \delta_{\ell \rho, \ell \rho} + \delta_{\ell \rho, \ell \rho} \delta_{\ell \rho, \ell \rho} - \delta_{\ell \rho, \ell \rho, \ell \rho} \quad (C.23)
\]

\[
E(\beta_{\ell \rho} \overline{\beta}_{\ell \rho}, \beta_{\ell \rho}, \overline{\beta}_{\ell \rho}) = \left\{ \begin{array}{c}
\delta_{t, \rho} \delta_{u, \rho} \delta_{r, \rho} \delta_{r, k} \\
+ \delta_{k, t} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{k, u} \delta_{t, r} \delta_{r, k} \\
+ \delta_{k, r} \delta_{t, u} \delta_{r, k} \\
+ \delta_{k, t} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{k, u} \delta_{t, r} \delta_{r, k} \\
+ \delta_{k, r} \delta_{t, u} \delta_{r, k} \\
+ \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{t, u} \delta_{r, r} \delta_{r, k} \\
+ \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{t, u} \delta_{r, r} \delta_{r, k} \\
+ \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{t, u} \delta_{r, r} \delta_{r, k} \\
+ \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
+ \delta_{t, u} \delta_{r, r} \delta_{r, k} \\
+ \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
+ 4 \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \\
\end{array} \right\} + \delta_{t, r} \delta_{u, \rho} \delta_{r, k} \delta_{r, k} \\
\]

Equations (C.22) - (C.24) can be directly substituted into equation (C.16), and the Kronecker deltas applied to the limits of summation. However, each term in equations (C.22) - (C.24) leads to a term in the final expression, so the final result is unwieldy and obscure.

An implementation of the sequence-averaged predictor receiver can compute the sequence-averaged second order statistics through time-averaging,

\[
E\left\{ |y_1|^2 \right\} \sim \sum_{i} \mu_{i} \left| y_{i - \ell \rho} \right|^2 \quad (C.25)
\]

\[
\frac{1}{2} E(\overline{\beta}_{\ell \rho} \beta_{\ell \rho}, \overline{y}_{\ell - \ell \rho}) \sim \sum_{i} \mu_{i} \overline{\beta}_{\ell \rho}(\ell \rho) \beta_{\ell \rho}(\ell \rho) y_{i - \ell \rho} \overline{y}_{i - \ell \rho} \quad (C.26)
\]
where $0 < \mu_{tr} \leq 1$ is a possibly time-varying forget factor to accommodate channels with non-stationary second order statistics. However, this assumes that the hypothesised phasors are independent, which cannot be justified since the metric is conditioned on the hypothesised phasor sequence.

This completes the mathematical description of the receiver structure. In the following paragraphs, its structure is outlined. It resembles the MLSE predictor receiver structure of Chapter 4, but with the sub-optimal branch metric of equation (C.13). The path metric is then

\[
\frac{1}{2} E \left( \beta y \bar{\beta}_n y_{v-1} y_{v-1} \right) = \sum_i \mu_{tr} \beta_{(i-1),r} \bar{\beta}_{(i-1),r} \beta y_{v-1} y_{v-1} y_{r-1}
\]

(C.27)

Figure C.2: Computational flow in the sequence-averaged predictor receiver.
Exhaustive comparison is needed to find the maximum metric. Since the evolution of the path metric is Markovian with a finite number of states, the task can be undertaken efficiently by the Viterbi algorithm. The path metric can be written in recursive form, as an old path metric updated by a branch metric,

\[
\Lambda_v(y|\hat{\beta}) = \Lambda_{(i-1)r}(y|\hat{\beta}) + \Lambda_v(y|\hat{\beta})
\]

\[
= \sum_{k=-\infty}^{(i+1)r-1} \sum_{n=kr}^{(i+1)r-1} \left[ y_n - \sum_{k}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \hat{\beta}_{kl} \sum_{m=1}^{B} b_{kln} \hat{\beta}_{ln} y_{n-l} - L+1+L+1 \left[ \sum_{l=1}^{B} b_{kln} \hat{\beta}_{ln} y_{n-l} \right] \right]^2 + \sum_{n=kr}^{(i+1)r-1} \left[ y_n - \sum_{k}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \hat{\beta}_{kl} \sum_{m=1}^{B} b_{kln} \hat{\beta}_{ln} y_{n-l} - L+1+L+1 \left[ \sum_{l=1}^{B} b_{kln} \hat{\beta}_{ln} y_{n-l} \right] \right]^2
\]

\[
= \sum_{k=-\infty}^{(i+1)r-1} \sum_{n=kr}^{(i+1)r-1} \frac{\left[ y_n - \sum_{k}^{\lfloor n/r \rfloor} \sum_{l=1}^{B} \hat{\beta}_{kl} \sum_{m=1}^{B} b_{kln} \hat{\beta}_{ln} y_{n-l} \right]^2}{\sigma^2 \left(y_{\lfloor n/r \rfloor} y_{\lfloor (n-1)/r \rfloor} \right)}
\]
The path metric, $\Lambda_{(i-1)r}(\hat{y}I\hat{b})$, is a function of the hypothesised phasors, $(\hat{b}_{-\infty} \ldots \hat{b}_{(i-1)r})$. The branch metric is a function of the $W = L + \lceil B/r \rceil$ hypothesised phasors, $(\hat{b}_{(i-W+1)r} \ldots \hat{b}_{ir})$. A state,

$$\sigma_{(i-1)r} = (\hat{b}_{(i-W+1)r} \ldots \hat{b}_{(i-1)r})$$

(C.30)
is associated with each surviving path metric, $\Lambda_{(i-1)r}(\hat{y}I\hat{b})$. When the path metric is extended to the $i$th symbol period, the branch metric, $\lambda_{ir}(\hat{y}I\hat{b})$, is computed. It depends only on the path metric's state and the new phasor, $\hat{b}_{ir}$. Thus the path metric evolution is Markovian, since it depends only on the previous state. By restricting the predictor length and ISI length to $W$ symbol periods, the number of states is finite, at $M^{W-1}$.

Thus the Viterbi algorithm can be used to find the sequence with maximum metric. It processes a partially-connected trellis, with $M^{W-1}$ states, and $M$ branches per state. Every $W$-phasor hypothesis vector, $(\hat{b}_{(i-W+1)r} \ldots \hat{b}_{(i-1)r} \hat{b}_{ir})$, labels a branch.

Figure C.2 is a diagram of the receiver's operation.

The sequence-averaged receiver's performance is considerably inferior that the MLSE receiver's performance. Figure C.3 shows that as the $E_b/N_0$ increases, the BER diminishes at a much slower rate for the sequence-averaged receiver.
REFERENCES


REFERENCES


[62] *Private communication with G. M. Vitetta and D. P. Taylor*


