# A Likelihood Ratio Analysis of Digital Phase Modulation

J.A. Schoonees

Thesis Presented for the Degree of
Doctor of Philosophy
in the Department of Electrical Engineering
University of Cape Town
10 March 1998

**Abstract**

# A Likelihood Ratio Analysis of Digital Phase Modulation

Although the likelihood ratio forms the theoretical basis for maximum likelihood (ML) detection in coherent digital communication systems, it has not been applied directly to the problem of designing good trellis-coded modulation (TCM) schemes. The remarkably simple optimal receiver of minimum shift keying (MSK) has been shown to result from the mathematical simplification of its likelihood ratio into a single term. The log-likelihood ratio then becomes a linear sum of metrics which can be implemented as a so-called simplified receiver, comprising only a few adders and delay elements. This thesis project investigated the possible existence of coded modulation schemes with similarly simplifying likelihood ratios, which would have almost trivially simple receivers compared to the Viterbi decoders which are typically required for maximum likelihood sequence estimation (MLSE). A useful notation, called the likelihood transform, was presented to aid the analysis of likelihood ratios. The work concentrated initially on computer-aided searches, first for trellis codes which may give rise to simplifying likelihood ratios for continuous phase modulation (CPM), and then for mathematical identities which may aid in the simplification of generic likelihood ratios for equal-energy modulation. The first search yielded no simplified receivers, and all the identities produced by the second search had structures similar to the likelihood ratio of MSK. These observations prompted a formal proof of the non-existence of simplified receivers which use information from more than two symbols in their observation period. This result strictly bounds the error performance that is possible with a simplified receiver. It was also proved that simplified receivers are only optimal for modulation schemes which use no more than two pairs of antipodal signals, and that only binary modulation schemes can have simplified receivers which use information from all the symbols in their observation period.

Johann August Schoonees
Keam Holdem Associates
27 Gillies Avenue
Newmarket
Auckland
New Zealand
10 March 1998

# Contents


**Glossary** xi
- Symbols . . . xi
- Abbreviations . . . xii

**Acknowledgements** xiii

**Summary** xiv

**1 Introduction** 1
- 1.1 Organization . . . 6
- 1.2 The Communication System . . . 6
- 1.3 Traditional Description of the Continuous Phase Modulated Signal 8
- 1.4 Rimoldi's Decomposition Approach to CPM . . . 10
  - 1.4.1 The memoryless modulator . . . 12
  - 1.4.2 The continuous-phase encoder . . . 13
  - 1.4.3 The advantages of the decomposition approach . . . 15
  - 1.4.4 Signal notation for tilted CPFSK representation . . . 18
- 1.5 Minimum Shift Keying . . . 19
  - 1.5.1 The MSK modulator . . . 19
  - 1.5.2 Massey's MSK detector . . . 22
- 1.6 Classical Detection Theory . . . 26
- 1.7 Likelihood Parameter Receivers . . . 29
- 1.8 The Gelfond-Schneider Theorem . . . 33
- 1.9 Euclidean Distances . . . 34

**2 Problem Statement, Objectives, and Literature Review** 39
- 2.1 Problem Statement . . . 39
- 2.2 Objectives . . . 40
- 2.3 Literature Review . . . 43

**3 Likelihood Ratio Analysis of Continuous-Phase Frequency Shift Keying** 49
- 3.1 Deriving a Likelihood Ratio for MSK . . . 50
- 3.2 The Likelihood Transform . . . 52
  - 3.2.1 Definition of the likelihood transform . . . 52
  - 3.2.2 The significance of the likelihood transform . . . 53
- 3.3 The Assumptions in the MSK Likelihood Ratio . . . 55
  - 3.3.1 Arbitrary starting phase state . . . 55

3.3.2 Limited observation period . . . . . . . . . . . . . . . . . . 57
3.4 Generalizing the MSK Analysis . . . . . . . . . . . . . . . . . . . . 59
3.4.1 Arbitrary starting phase states for full-response CPFSK . 59
3.4.2 The significance of starting phase states in CPM with coding . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.4.3 Extending to multi-level signalling . . . . . . . . . . . . . 64
3.4.4 The choice of observation period . . . . . . . . . . . . . . 65
3.5 Catastrophic Trellises . . . . . . . . . . . . . . . . . . . . . . . . 66

**4 Searches for Simplified Receivers** **69**
4.1 Constructing Receivers From Simplified Likelihood Ratios . . . . 70
4.2 Search for Simplified Receivers for Specific Modulations . . . . . 75
4.2.1 Four-level CPFSK with $h = 1/4$ . . . . . . . . . . . . . . 75
4.2.2 Quadrature Minimum Shift Keying . . . . . . . . . . . . . 77
4.2.3 Likelihood ratio analysis of QMSK . . . . . . . . . . . . . 84
4.2.4 The search for a simplified receiver for QMSK . . . . . . . 87
4.2.5 Number of trellises tested . . . . . . . . . . . . . . . . . . 94
4.3 Search for Identities to Aid Simplification . . . . . . . . . . . . . 96
4.3.1 Search for MSK-like identities . . . . . . . . . . . . . . . . 97
4.3.2 A more general search for identities . . . . . . . . . . . . 100
4.3.3 Search for identities derived from trellises . . . . . . . . . 104

**5 Interpretation of the Search Results** **113**
5.1 Interpreting the Results of the QMSK Trellis Search . . . . . . . 114
5.1.1 A Re-examination of Catastrophic Schemes . . . . . . . . 115
5.2 Interpreting the Identities Found . . . . . . . . . . . . . . . . . . 119
5.2.1 The MSK identity . . . . . . . . . . . . . . . . . . . . . . 119
5.2.2 The family of MSK identities . . . . . . . . . . . . . . . . 120
5.2.3 The identities over three symbols . . . . . . . . . . . . . . 122
5.2.4 Extending identities for longer observation periods . . . . 127

**6 A Formal Analysis of Simplified Receivers** **131**
6.1 Generalizing the Found Identities . . . . . . . . . . . . . . . . . . 132
6.1.1 Construction procedure for arbitrary simplifying likelihood ratios . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
6.1.2 Properties of the construction procedure . . . . . . . . . . 137
6.2 Bounds on the Performance of Simplified Receivers . . . . . . . . 138
6.2.1 Simplification of likelihood ratios . . . . . . . . . . . . . . 138
6.2.2 Performance bounds . . . . . . . . . . . . . . . . . . . . . 143
6.2.3 Implications for the performance of simplified receivers . . 151

**7 Conclusion** **152**

**A The Number of QMSK Trellis Codes** **159**

**B The Search Algorithm for Identities** **166**
B.1 Simplification . . . . . . . . . . . . . . . . . . . . . . . . . . . 166
B.2 Mathematically Distinct Expressions . . . . . . . . . . . . . . . . . 167
B.3 Generating Multi-Sets of Variables . . . . . . . . . . . . . . . . . . 169
B.4 Generating All Combinations of Reciprocals . . . . . . . . . . . . 170
B.5 Partitioning the Multi-Set . . . . . . . . . . . . . . . . . . . . . . . 170
B.6 Generating All Ordered Partitions of a Set . . . . . . . . . . . . . 171
B.7 Generating Likelihood Ratio Expressions . . . . . . . . . . . . . . 171
B.8 Simplification of the expression . . . . . . . . . . . . . . . . . . . 172

**C Results of the Search for Likelihood Ratio Identities** **174**

**Bibliography** **180**

# List of Tables


1.1 Differential MSK (Type I) signal set . . . . . . . . . . . . . . . . 21

4.1 Hodgart's QMSK encoder A . . . . . . . . . . . . . . . . . . . . . . 81
4.2 Hodgart's QMSK encoder B . . . . . . . . . . . . . . . . . . . . . . 82
4.3 Number of possible trellises . . . . . . . . . . . . . . . . . . . . . 88
4.4 Incremental normalized correlation coefficients . . . . . . . . . . . 93
4.5 Example calculation of NSED . . . . . . . . . . . . . . . . . . . . . 93
4.6 Results of relaxing qualifying tests . . . . . . . . . . . . . . . . . 94
4.7 Total numbers of trellises tested . . . . . . . . . . . . . . . . . . . 96
4.8 Powers and numbers of terms searched . . . . . . . . . . . . . . . . 99
4.9 Number of possible expressions . . . . . . . . . . . . . . . . . . . . 103
4.10 Number of ordered partitions of a set of branch labels . . . . . . 110

5.1 Signal set of differential offset QPSK . . . . . . . . . . . . . . . . 120

A.1 Number of possible trellises . . . . . . . . . . . . . . . . . . . . . 161
A.2 Number of combinations of used states . . . . . . . . . . . . . . . . 164

# List of Figures


1.1 Block diagram of a coded modulation transmitter . . . . . . . . . 7
1.2 CPFSK phase trees . . . . . . . . . . . . . . . . . . . . . . . . 10
1.3 Memoryless modulator . . . . . . . . . . . . . . . . . . . . . . 14
1.4 Continuous-phase encoder . . . . . . . . . . . . . . . . . . . . . 14
1.5 CPE trellis diagram . . . . . . . . . . . . . . . . . . . . . . . . 15
1.6 Differential CPM . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.7 MSK (FFSK) modulator block diagram . . . . . . . . . . . . . . 20
1.8 Differential MSK (Type I) modulator block diagram . . . . . . . 21
1.9 The Massey MSK detector . . . . . . . . . . . . . . . . . . . . . 22
1.10 The MSK trellis . . . . . . . . . . . . . . . . . . . . . . . . . . 23
1.11 The Massey-Hodgart MSK receiver . . . . . . . . . . . . . . . . 25
1.12 Likelihood parameter receiver . . . . . . . . . . . . . . . . . . . 33

3.1 MSK trellis diagram labelled with LTVs . . . . . . . . . . . . . . 56
3.2 MSK trellis with extended observation period . . . . . . . . . . . 58
3.3 Trellis invariant over starting phase state . . . . . . . . . . . . . 62
3.4 Example of a catastrophic trellis . . . . . . . . . . . . . . . . . . 66

4.1 Derived maximum likelihood receiver . . . . . . . . . . . . . . . . 72
4.2 Trellis derived from likelihood functions . . . . . . . . . . . . . . 72
4.3 CPM with parallel branches . . . . . . . . . . . . . . . . . . . . 73
4.4 4-CPFSK trellis diagram . . . . . . . . . . . . . . . . . . . . . . 76
4.5 QMSK modulator waveforms . . . . . . . . . . . . . . . . . . . . 79
4.6 QMSK phase trellis . . . . . . . . . . . . . . . . . . . . . . . . . 80
4.7 Trellis for Hodgart's QMSK encoder A . . . . . . . . . . . . . . . 82
4.8 Trellis for Hodgart's QMSK encoder B . . . . . . . . . . . . . . . 82
4.9 Trellis for Hodgart's QMSK encoder C . . . . . . . . . . . . . . . 83
4.10 QMSK modulator block diagram . . . . . . . . . . . . . . . . . . 84
4.11 Combined trellis for Hodgart's scheme A . . . . . . . . . . . . . 85
4.12 Combined trellis for Hodgart's QMSK scheme C . . . . . . . . . 86
4.13 Example of a catastrophic QMSK trellis . . . . . . . . . . . . . . 91
4.14 QMSK trellis without SLR test . . . . . . . . . . . . . . . . . . . 95
4.15 QMSK trellis without setting a minimum distance . . . . . . . . 95
4.16 Trellis corresponding to general likelihood ratio . . . . . . . . . . 101
4.17 Different trellises with same likelihood ratio . . . . . . . . . . . . 106
4.18 Tree for general likelihood ratio . . . . . . . . . . . . . . . . . . 107

5.1 Catastrophic QMSK trellis of Figure 4.13 . . . . . . . . . . . . . 116

5.2 Trellis of a hypothetical limited catastrophic scheme . . . . . . . 117
5.3 Tree resulting from an MSK-like identity . . . . . . . . . . . . . . 121
5.4 Tree and trellis which drop an interval . . . . . . . . . . . . . . . 123
5.5 Trellis diagram for coded QPSK . . . . . . . . . . . . . . . . . . . 124
5.6 Tree and trellis of apparently optimal identity . . . . . . . . . . . 125
5.7 Degenerate tree after time shift . . . . . . . . . . . . . . . . . . . 126
5.8 Tree resulting from repeated MSK identity . . . . . . . . . . . . 128
5.9 Block diagram of simplified receiver . . . . . . . . . . . . . . . . 129
5.10 Regular trellis using two out of three symbols . . . . . . . . . . . 129

6.1 Tree illustrating the construction procedure . . . . . . . . . . . . 136
6.2 Three branches from a common state in a tree . . . . . . . . . . 146
6.3 A binary tree to show possible branch labelling . . . . . . . . . . 148
6.4 Binary uncoded antipodal modulation from contribution of $a^2$ . . 149
6.5 Two-state trellis resulting from contributions $a/b$ and $ab$ . . . . . 150

A.1 Trellises containing mutually exclusive sub-graphs . . . . . . . . 160
A.2 Trellises with identical signal sets . . . . . . . . . . . . . . . . . . 162

# Glossary

## Symbols

| | |
|---|---|
| $\boldsymbol{A}$ | memoryless mapping matrix between CPE and MM |
| $D$ | delay operator |
| $d^2(s,s')$ | normalized squared Euclidean distance |
| $d^2_{\text{free}}$ | (squared) free Euclidean distance |
| $E$ | symbol energy |
| $E(R)$ | expectation operator, expected value of $R$ |
| $\eta$ | threshold in likelihood ratio test |
| $f_{R\|H}(R\|H)$ | conditional probability density function |
| $h$ | modulation index, frequency deviation ratio |
| $J$ | cycle length of time-varying CE |
| $L$ | phase response length, memory |
| $\lambda(S_{vu})$ | correlation metric, matched filter output |
| $\Lambda, \Lambda_{vu}, \Lambda(R)$ | likelihood ratio |
| $l_j, l_{vu}$ | likelihood parameter |
| $M$ | number of possible input symbols |
| $m$ | number of signals in signal set |
| $N$ | length of observation period |
| $n$ | symbol time index, symbol number |
| $N_0$ | single-sided noise power spectral density |
| $\omega_0$ | (average) lowest signalling frequency |
| $\omega_c$ | centre frequency, carrier frequency |
| $P$ | denominator of a rational modulation index |
| $P(H)$ | probability of $H$, *a priori* probability |
| $\varphi_0$ | arbitrary phase offset |
| $\psi(\tau, \boldsymbol{u})$ | tilted phase |
| $Q$ | total number of input sequences of length $N-1$ |
| $r$ | observation of $R$, sampled output of matched filter |
| $R$ | observation random variable, matched filter output |
| $r(t)$ | received signal |
| $s(t, \boldsymbol{u})$ | transmitted signal |
| $\sigma_n$ | CPE state |
| $S_{vu}$ | abbreviation of $s_{vu}(\tau)$ |
| $s_{vu}(\tau)$ | MM output signal in $n$-th interval |
| $T$ | symbol period |
| $t$ | time |

| | |
|---|---|
| $\tau$ | time shifted to origin, $0 \le \tau < T$ |
| $\theta(\tau)$ | data-independent component of signal phase |
| $u_n$ | CPE input, CE output data symbol |
| $\boldsymbol{u}$ | input data sequence |
| $\boldsymbol{U}$ | sequence of 'unwanted' parameters in compound hypothesis |
| $v_n$ | differential input symbol, phase state |
| $\boldsymbol{v}$ | differential input data sequence |
| $w(t)$ | stationary white Gaussian noise process |
| $Z$ | number of CE states |
| $\zeta_n$ | CE state |

# Abbreviations

| | |
|---|---|
| AMF | average matched filter |
| AWGN | additive white Gaussian noise |
| CE | channel encoder |
| CPE | continuous-phase encoder |
| CPFSK | continuous-phase frequency shift keying |
| CPM | continuous phase modulation |
| FFSK | fast frequency shift keying |
| LTV | likelihood transform variable |
| MAP | maximum *a posteriori* |
| MLSE | maximum likelihood sequence estimation |
| MM | memoryless modulator |
| MSK | minimum shift keying |
| NSED | normalized squared Euclidean distance |
| QMSK | quadrature minimum shift keying |
| TCM | trellis-coded modulation |
| TFM | tamed frequency modulation |

# Acknowledgements

Heartfelt thanks to Robin Braun, for his generous support over the years; to the members of the Digital Modulation Research Group, for their companionship; to the anonymous examiners, for their helpful suggestions; to the Foundation for Research Development, for partial funding of this work; to Stephen Hodgart, for sharing his many innovative ideas with me; to John Holdem, Rick Keam, Bruce Reeler and Jon Tapson, for reading and commenting on drafts; to Ken Hughes and Zaid Kimmie, for informative discussions mathematical; to Chris and Heather Pienaar for their hospitality; and to Kate Schoonees, for hours of late-night help in preparing the figures and equations in this dissertation.

# Summary

Trellis-coded modulation and the search for good modulation codes is presently one of the most active research fields in telecommunications. Since coded modulation is created with a transmitter which has memory, it is best received using some form of sequence estimation. When the usual assumptions about equal *a priori* information symbol probabilities and additive white Gaussian noise in the channel are made, the optimal receiver uses maximum likelihood sequence estimation (MLSE). All the known sequence estimation algorithms, maximum likelihood or not, are relatively complex to implement in hardware. The widely used MLSE Viterbi algorithm can usually only be applied to the simpler coded modulation schemes.

In contrast with the complexity of MLSE algorithms, minimum shift keying (MSK) possesses at least three remarkably simple maximum likelihood receivers. MSK is an important member of the class of continuous phase modulation (CPM), and may be regarded with the rest of CPM as a form of coded modulation. The reason for the simplicity of MSK's receivers becomes apparent when the Bayesian likelihood ratio is written down for MSK, and is seen to simplify to a single term. It appears that this simplification was first noticed by G.D. Forney, Jr. Taking the logarithm of the simplified likelihood ratio renders it linear and leads directly to an almost trivially simple linear MSK receiver, first described by J.L. Massey. Other MSK receivers have equally simple structures, but Massey's version is the only one which is a direct implementation of the simplified log-likelihood ratio. In this dissertation, a receiver which results from a simplified log-likelihood ratio is called a *simplified receiver*, and it can be shown that it always makes maximum likelihood decisions.

The hypothesis of this thesis is that if a simple optimal MSK-like (simplified) receiver exists for any other digital modulation scheme, the key to its existence would be the simplification of its likelihood ratio. The objective of the project was to examine as many likelihood ratios as possible, in the hope of finding a suitable simplification or, failing that, of learning enough about the characteristics of simplifying likelihood ratios to attempt a formal analysis.

The method followed was to start by searching the space of trellis codes and their associated likelihood ratios with the help of computers. In this respect, this work is similar to a lot of recent research into finding good modulation codes. The space of all possible codes is vast, and researchers are forced to confine themselves to specific parts of it.

In this project, the search for simplifying likelihood ratios was limited at first by considering only likelihood ratios of specific CPM schemes, and later by considering mathematical identities which may be used to simplify likelihood ratios. In both cases the search space grows combinatorially with the number of encoder states and the length of the observation period, and only the simpler forms of likelihood ratios could be evaluated exhaustively.

The two specific modulation schemes that were considered were four-level continuous-phase frequency shift keying with modulation index $h = 1/4$, and a coded version called quadrature minimum shift keying (QMSK). In the latter case a computer program was written to generate all possible code trellises for a given number of states and given observation period in the receiver. Each trellis was examined to see if its associated likelihood ratios simplified, in which case a simplified receiver would have been found. Many such trellises were found, but all of them are catastrophic in the sense that it is possible to find pairs of parallel paths through the trellis that accumulate no Euclidean distance between them in signal space.

The catastrophic trellises were then re-examined to determine if any of them could nevertheless result in practical receivers. It was found that a certain class of catastrophic codes could indeed be useful, but none of the simplified QMSK likelihood ratios fell into this class. The fact that no likelihood ratio simplification occurred during the investigation of QMSK codes encouraged the belief that none was possible within the definition of QMSK.

The search for mathematical identities was done by generating all possible mathematically distinct expressions that can occur in a likelihood ratio from a trellis of given size. Each expression was simplified to see if it could be written as a single term (comprising only products and quotients of variables), in which case an identity has been found which can be used to simplify the associated likelihood ratio. Many such identities were found, and it was observed that they all shared structural similarities with the likelihood ratio of MSK.

The mathematical identities that were found all resulted from exhaustive searches of subspaces of the search space, and it could therefore be stated with certainty that no others exist in those subspaces. These results presented an opportunity to formalize the mathematical description of simplifying likelihood ratios, which in turn led to a series of rigorous proofs of performance bounds on the corresponding receivers.

The most important result of this work was the finding that MSK, together with a related family of antipodal modulation schemes which share essential properties of its likelihood ratio, is unique in possessing a simplified receiver which is in some sense optimal. It was found that simplified receivers which use information from all the symbols in their observation period are optimal only

1. for binary modulations;

2. for antipodal bi-orthogonal signal sets with at most four signals; and

3. for observation periods of no more than two symbol intervals.

If it is not required that information from all the symbols in the observation period is used, these bounds may be relaxed at the cost of performance.

The value of the work lies in the conclusive bounds that were set on what can be achieved with simplifying likelihood ratios. In addition, a concise notation, called the likelihood transform, was developed for evaluating likelihood ratios quickly and showing their correspondence to maximum likelihood receiver structures. The likelihood transform may be generally useful outside the immediate context, specifically in the analysis of generic receiver types independently of specific choices of signal set.

# Chapter 1

# Introduction

Trellis-coded modulation (TCM) [37, 82] and continuous phase modulation (CPM) [10, 9] have become major research fields in recent years and have resulted in significant advances in telecommunications technology. Coded modulation is based on the idea that modulation and coding should be combined and optimized together to improve the performance of digital modulation systems.

Coded modulation differs essentially from uncoded modulation in that successive symbols, passed to the modulator for transmission as signals, are not statistically independent [14]. A channel encoder (CE) introduces memory into the transmitter which has the effect of causing any transmitted signal to depend not only on the current input to the transmitter, but also on the current memory state of the CE. Significant performance gains can then be achieved by using a signal set which is larger than that which is strictly needed to transmit the information.

For example, if a communication system needs to transmit two bits per signalling interval, the conventional uncoded signal set would need to contain no more than four signals. If this number were doubled to eight, then a CE which takes two bits as its input and produces three bits at its output every signalling interval (a rate-2/3 encoder) can be used to map the input symbols to the signal set. With good design, this redundancy in the signal set can result in a lowering of the error rate, or equivalently permit a poorer signal-to-noise ratio at the same error rate. The signal-to-noise margin so gained is often called the coding gain of the encoder, following traditional error-correcting coding terminology. The increased freedom of choice of signals from the signal set permits sequences of signals to be spaced further apart in the signal space. If the expanded signal set occupies the same bandwidth as the original one, then this improvement in power efficiency comes at no cost to the bandwidth efficiency of the system.

The traditional uncoded coherent modulations of amplitude shift keying (ASK), phase shift keying (PSK) and frequency shift keying (FSK) can all be optimally received by examining only one received signal at a time in the receiver, if the channel is memoryless (that is, in the absence of intersymbol interference). The receiver for a coded modulation, however, has to observe a sequence of two or more received signals to make an optimal decision. In the theoretical limit, an observation period equal in length to the entire transmitted sequence is required.

An important tool in the reception of coded modulation, is the Viterbi algorithm [84], which was initially proposed as an optimal decoder for convolutional codes. When implemented in a receiver to do maximum likelihood sequence estimation (MLSE), the algorithm is often called a Viterbi decoder. It is widely used and was shown by Forney [24] to be a maximum likelihood decoder, that is, it always selects the code word which maximizes the log-likelihood function.

The great advantage of the Viterbi decoder over other methods is that its computational complexity and storage requirements grow only linearly with the length of the observation period. Unfortunately, its complexity increases exponentially with the amount of memory present in the CE. When convolutional coding is used, the amount of memory in the encoder is usually quantified as the constraint length, which is defined as the number of output symbols of the encoder which are influenced by a single input symbol. Larger coding gains are generally achieved by using codes with longer constraint lengths. The Viterbi algorithm can only be implemented practically for relatively short constraint lengths. For longer codes, suboptimal algorithms are therefore sometimes used, which have complexities which are essentially independent of the constraint length. The Fano algorithm [23] is the best known of these, and falls in the class of so-called sequential decoding algorithms.

All these decoding algorithms, even the suboptimal ones, are complex when compared to the receivers of uncoded modulations. The complexity refers both to the number of calculations and comparisons that have to be made to make a decision, and, related to that, the circuitry required to implement the algorithm. The choice of code and decoder creates an important constraint on the maximum data rate that can be supported with the available technology. To some extent, the complexity of decoders has become more manageable in recent years with the advent of powerful digital signal processing integrated circuits. However, system cost is still strongly influenced by the complexity of its components.

Coded modulation is mostly assumed to refer to TCM, which typically uses a trellis encoder in combination with a PSK or ASK modulator (in one or more dimensions). It is less often acknowledged that CPM belongs equally well in the class of coded modulations. CPM had been invented earlier as a family of modulations with good power and bandwidth efficiency, which lends itself naturally to trellis phase coding [4, 3]. CPM signals have constant amplitude envelopes, and are distinguished by the requirement that their signal phases must be continuous for the duration of the transmission.

Simultaneously with the advent of TCM, it was realized that CPM's relatively good power efficiency may be explained by regarding it as a form of coded modulation. An early example is Forney's use [24] of CPM as an illustration of the application of the Viterbi algorithm in communication receivers. Rimoldi [65], following an earlier concept by Massey [49], proposed a useful decomposition of CPM which makes this interpretation explicit: the continuous-phase property is enforced with a so-called continuous-phase encoder (CPE) which precedes a memoryless modulator (MM). Not surprisingly, therefore, CPM receivers are as complex as those of other coded modulations. They typically also require the Viterbi algorithm to demodulate the signal optimally through MLSE.

The CPM class may conveniently be divided into full-response signalling, in which each symbol is transmitted within a single signalling period, and partial response signalling, in which each symbol is intentionally overlapped in time with its neighbours to improve bandwidth and power efficiency. Full-response CPM is nevertheless continuous-phase encoded. An important subclass of CPM is continuous-phase frequency shift keying (CPFSK) which is characterized by its linear phase response function. CPFSK may also use either full-response or partial response signalling. CPFSK is often used to represent CPM more generally because of its relative simplicity, and because it possesses most of the essential characteristics of many CPM schemes. A particularly well-known member of the CPFSK family, and the cornerstone of CPM, is minimum shift keying (MSK), which is characterized by having a modulation index $h = 1/2$. (The modulation index $h$ is defined as the smallest frequency difference between the signals in the signal set, normalized to the symbol rate.)

It may seem surprising then that MSK possesses several unusually simple receivers, the best-known ones being due to de Buda [20] and Amoroso and Kivett [2], as well as some elegant reformulations due to Massey [48], Ryu and Un [70] and Rimoldi [67].

Massey's MSK receiver [48] is based on the direct implementation of the Bayesian likelihood ratio test [80] for MSK. A general formulation of the likelihood ratio test for full-response CPFSK exists, due to Osborne and Luntz [56]. The Osborne-Luntz receiver is highly complex for most applications and impractical for actual implementation. Its main application is as a means of calculating the theoretical error performance of general CPM schemes. When the Osborne-Luntz likelihood ratio test is applied to MSK, a remarkable simplification occurs which also renders the test linear. The almost trivially simple MSK receiver is derived directly from the simplified likelihood ratio test, and is nonetheless a maximum likelihood receiver, optimal in the sense of minimizing the bit error rate for equally likely input symbols. In this dissertation, receivers which result from such simplification of the likelihood ratio will be called *simplified receivers.*

It is tempting to ask whether any other coded modulation scheme possesses a likelihood ratio which simplifies in a similar manner. If one is found, it will simultaneously have a simplified receiver. Although the Osborne-Luntz receiver was originally proposed for CPFSK, it is applicable to any CPM [3, p.238], in fact to any coded modulation with equal-energy signals. It may be used as a tool for generating maximum likelihood receivers for the large class of equal-energy coded modulations. The task is then to identify those modulation schemes that have simplifying likelihood ratios, because they will automatically have maximum likelihood receivers with very low complexity.

For all practical purposes, equal-energy signalling refers to constant amplitude envelope signalling, which is another way of denoting digital phase modulation. The emphasis in this dissertation will be on CPM as an important class of power and bandwidth efficient modulations with 'built-in' continuous-phase coding. The fact that MSK with its simplified receiver is a CPM scheme provides additional motivation for this emphasis. In later chapters, after the introduction of the likelihood transform in Chapter 3, more general (discontinuous) phase modulation is considered on an equal footing with CPM.

In attempting to identify simplifying likelihood ratios, the main problem is the enormous size of the space of possible modulation codes. This is the same problem faced by all code designers, and a lot of research has been done in the field recently. Systematic searches have to be constrained either by using particular mathematical formulations of codes such as partitions of lattices into cosets [17, 25, 26], nonlinear parity check equations [60] or algebraic group systems [12], or by applying various heuristic synthesis methods [88, 91, 89, 18] following the precedent originally set by Ungerboeck [82].

It would therefore have been most useful to have had some formal limitations of the space of all simplifying likelihood ratios in order to constrain the search or even to make a search unnecessary. At the outset of this project such information was not available. The project proceeded by exploring the code space of a specific CPM with coding, and the space of mathematical expressions that constitute general likelihood ratios. With the information gained from these searches, it was possible to return to a formal analysis of the limits of the space of simplifying likelihood ratios and the implications that these limits have on the performance of simplified receivers.

## 1.1 Organization

This dissertation is organized as follows: The present chapter gives some theoretical background, and the next chapter presents the problem statement, objectives and a literature review. Chapter 3 introduces likelihood ratio analysis and the likelihood transform, using MSK as the most relevant example. Chapter 4 describes the methods used in the search in detail, and also presents the results of the search. Chapter 5 offers a brief analysis of the search results. Chapter 6 proves a set of bounds on simplifying likelihood ratios and the performance of simplified receivers, followed by a discussion and conclusions in Chapter 7.

The rest of this chapter introduces the theoretical concepts needed for the dissertation. First, in section 1.2, some notation is defined for the communications channel. The traditional description of CPM [10, 9] is given in section 1.3, as a basis for the emphasis that is placed on CPM. A motivation for regarding CPM as a form of coded modulation is given in section 1.4 which discusses Rimoldi's [65] decomposition approach to CPM. Then MSK [22, 20, 48] is discussed in detail in section 1.5 because of its importance in having a simplifying likelihood ratio. Classical detection theory [80] is discussed next in section 1.6 as an introduction to Osborne and Luntz's [56] likelihood parameter receiver covered in section 1.7. The Gelfond-Schneider theorem [11] is mentioned in section 1.8 because it will be used in Chapter 3 to formulate the likelihood transform. Finally, the important concept of Euclidean distance between signals and sequences of signals is defined in section 1.9.

## 1.2 The Communication System

The communication channel is assumed to be an additive white Gaussian noise (AWGN) channel throughout this dissertation. The received signal is $r(t) = s(t)+w(t)$ where $w(t)$ is a stationary Gaussian-distributed random process with constant two-sided power spectral density $N_0/2$. One signal from a given signal set is transmitted every symbol interval of length $T$ seconds, and all signals are assumed to have identical energies $E$ in every symbol interval. The symbol intervals are numbered, for ease of reference, as $n = 0, 1, 2, \ldots$

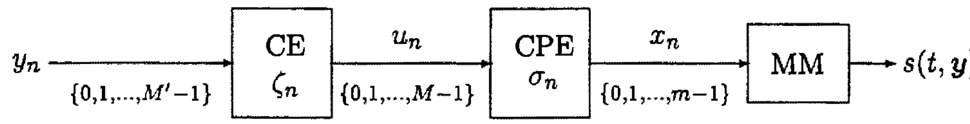



Figure 1.1: General system block diagram of a coded modulation transmitter, showing a channel encoder (CE), continuous-phase encoder (CPE), and memoryless modulator (MM).

The modulation schemes of interest here are almost all coded, in other words the transmitter possesses memory which imposes some order on the sequence of transmitted signals. Figure 1.1 shows a general transmitter block diagram. CPM is also a coded modulation scheme even when an external CE is not used. The default CPM 'encoder' is usually called a CPE to distinguish it from additional coding. There is some potential for confusion when a phrases such as 'coded CPM' and 'uncoded CPM' are used: does 'coded' refer to the CPE or to the presence of a CE? These terms are therefore avoided in this dissertation, in favour of the phrases 'CPM with coding' and 'CPM without coding' which hopefully makes clear the presence or absence, respectively, of a CE in addition to the CPE.

A related source of possible confusion is in the notation for the number of symbols in the input alphabet of coded modulation systems. The generally accepted conventions in the literature are not entirely consistent. The lower case $m$ will be used throughout to denote the number of signals in the signal set. The upper case $M$ will be used generally to denote the size of the input symbol alphabet from the information source. However, in CPM systems (with or without coding) $M$ will always be the number of possible input symbols to the CPE. If a CE is also present in addition to the CPE, it will sometimes be necessary to refer explicitly to the number of possible inputs to the CE, to distinguish that number from $M$. The word 'level', as in 'four-level CPFSK', will also be used to refer to the number of input symbols in the same way as $M$. For uncoded modulation $m = M$, but for coded modulation generally $m > M$.

In the context of this thesis, *equal-energy signalling* will be taken to mean that the signals in the signal set have the same energy in every signalling interval.

The receiver is assumed to be coherent, with perfect carrier synchronization and symbol timing available to it. It will generally comprise a bank of matched filters (usually implemented as correlators) and samplers at its input, followed by some form of soft-decision sequence estimation. By this it is meant that the sampled matched filter outputs are not quantized, but are passed on directly to the sequence estimator as analogue values. In some cases, the sequence estimator observes the $N$ most recent matched filter samples at a time; this is called the observation period or observation window. In other cases, there is no explicit limit on the number of samples observed, as for example in the case of the Viterbi decoder. The output of the sequence estimator is one symbol estimate every symbol interval, delayed by a certain time since its transmission.

A glossary of symbols is given after the table of contents.

## 1.3 Traditional Description of the Continuous Phase Modulated Signal

This section gives a general introduction to CPM and CPFSK, following the definitive approach of [10], [9] and [3].

The CPM waveform is described as

$$s(t, \boldsymbol{\alpha}) = \sqrt{\frac{2E}{T}} \cos\left[\omega_c t + \varphi(t, \boldsymbol{\alpha}) + \varphi_0\right], \quad t \geq 0 \tag{1.1}$$

where

$$\varphi(t, \boldsymbol{\alpha}) = 2\pi h \sum_{i=0}^{\infty} \alpha_i q(t - iT), \quad t \geq 0 \tag{1.2}$$

is the information-carrying phase which completely describes the CPM signal (apart from the symbol energy $E$, the carrier or centre frequency $\omega_c$ and the arbitrary phase offset $\varphi_0$, all parameters which do not affect theoretical power or bandwidth efficiency). $T$ is the symbol period.

(1.2) may be extended if necessary to include phase-modulated signals which are not necessarily of continuous phase, by allowing the phase function $\varphi(t, \boldsymbol{\alpha})$ to be discontinuous. The information sequence $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \ldots)$ is chosen from an input alphabet of $M$ symbols with

$$\alpha_n \in \begin{cases} \{\pm 1, \pm 3, \ldots, \pm(M-1)\} & M \text{ even} \\ \{0, \pm 2, \ldots, \pm(M-1)\} & M \text{ odd} \end{cases} \qquad n = 0, 1, 2, \ldots \tag{1.3}$$

The modulation index $h$ is assumed to be a rational number

$$h = \frac{K}{P} \tag{1.4}$$

where $K$ and $P$ are relatively prime positive integers. This constraint gives the phase trellis (modulo $2\pi$) a periodic structure with a finite number of phase states. $q(t)$ is the monotonically increasing phase response function, and is constrained by

$$q(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t \geq LT \end{cases} \tag{1.5}$$

where $L \geq 1$ is the phase response length. The first derivative of $q(t)$ with respect to time is the frequency pulse

$$g(t) = \frac{d\,q(t)}{dt} \tag{1.6}$$

which has an area of $1/2$.

CPM schemes may be characterized as being either full-response or partial response schemes, depending on the length of the phase response $L$. This is the number of input symbols which affect the change in phase over the current symbol interval, or alternatively, the number of symbol intervals over which one input symbol affects the signal phase. Full-response schemes have $L = 1$, and partial response schemes have $L > 1$. $L$ is sometimes called the *memory* of the scheme.

The frequency pulse $g(t)$ provides a convenient notation for denoting some important CPM schemes [3], depending on the pulse shape. LREC denotes a rectangular pulse of length $L$, and LRC denotes a raised cosine pulse of length $L$. For example, a partial response scheme which has a raised cosine frequency pulse spread over two symbol intervals, is called a 2RC scheme.

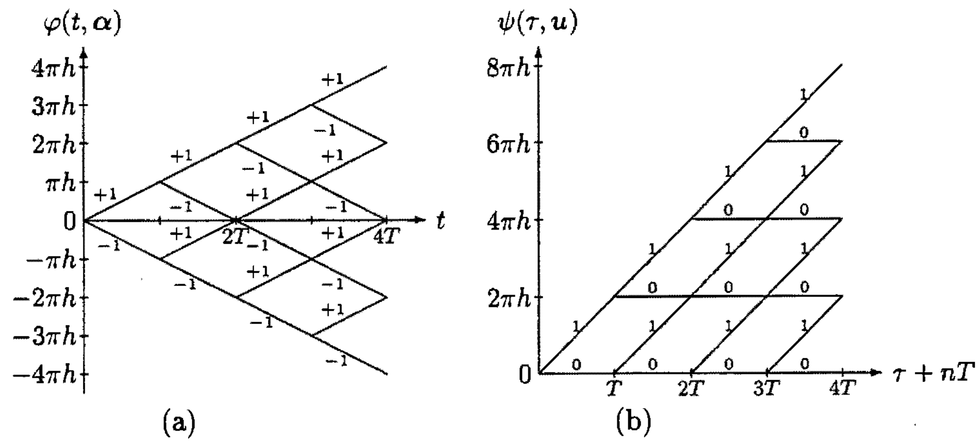



Figure 1.2: The traditional phase tree (a), and its tilted phase equivalent (b), here shown for full-response CPFSK with $M = 2$. The branches are labelled with input symbols.

An LREC scheme results if the phase response function is defined as

$$q(t) = \begin{cases} 0 & t < 0 \\ \frac{t}{2LT} & 0 \leq t < LT \\ \frac{1}{2} & t \geq LT \end{cases} \tag{1.7}$$

The phase tree is then composed entirely of straight line segments, and the modulation is therefore also known as *continuous-phase frequency shift keying* (CPFSK). LREC and CPFSK are different names for exactly the same modulation. Full-response CPFSK is also called 1REC. Its phase tree is illustrated in Figure 1.2(a).

## 1.4 Rimoldi's Decomposition Approach to CPM

Rimoldi's separation of the continuous-phase coding and memoryless modulating functions of a CPM transmitter is by now a standard description of CPM, and is especially useful when considering CPM with coding. This decomposition also plays an important role in the conceptual unification of the fields of TCM and CPM, because it makes it possible to regard the continuous-phase property of CPM as standard convolutional coding.

This section and its subsections are a summary of Rimoldi's well-known paper [65]. It is included in some detail here because, without it, it would be difficult in the rest of the dissertation to treat CPM and TCM together as general coded modulation, and also because the focus of this work is initially on CPM.

An explicit decomposition of the CPM modulator into a *continuous-phase encoder* (CPE) and a *memoryless modulator* (MM) is achieved by translating the traditional phase $\varphi(t, \boldsymbol{\alpha})$ of (1.2) into a *tilted phase.* The tilted-phase description of CPM is achieved by redefining the CPM signal during the $n$-th symbol interval as follows:

$$
\begin{aligned}
s(t, \boldsymbol{u}) &= \sqrt{\frac{2E}{T}} \cos\left[\omega_0 t + \psi(t - nT, \boldsymbol{u}) + \varphi_0\right], \quad t \geq 0 && (1.8)\\
\psi(\tau, \boldsymbol{u}) &= 2\pi h \left[ v_n + 2\sum_{j=0}^{L-1} u_{n-j} q(\tau + jT) \right] + \theta(\tau), \quad 0 \leq \tau < T && (1.9)\\
v_n &= \left[ \sum_{i=0}^{n-L} u_i \right]_{\bmod P}, \quad n = 0, 1, 2, \ldots && (1.10)\\
\theta(\tau) &= \pi h (M-1)\tau/T - 2\pi h(M-1)\sum_{j=0}^{L-1} q(\tau + jT) + \\
&\quad \frac{\pi h}{2}(M-1)(L-1), \quad 0 \leq \tau < T && (1.11)
\end{aligned}
$$

The input data sequence $\boldsymbol{u}$ now comprises $u_n \in \{0, 1, \ldots, M-1\}$, $n = 0, 1, 2, \ldots$ It is assumed that $u_n = 0$ for $n = -1, -2, -3, \ldots$ and that the frequency pulse $g(t) = d\,q(t)/dt$ is symmetrical around $t = LT/2$. (These assumptions are not strictly necessary, but they simplify the derivation of (1.8) - (1.11) from the traditional description.)

The reference frequency $\omega_0$ corresponds to the average signal frequency when the all-0 data sequence is being transmitted, that is, nominally the lowest signalling frequency. The information-carrying phase $\psi(\tau, \boldsymbol{u})$ is now called the tilted phase, because it is expressed relative to $\omega_0$ instead of the centre frequency $\omega_c$. Figure 1.2(b) shows an example for $M = 2$ CPFSK. The tilted phase has the important characteristic of being defined over a single symbol interval $0 \leq \tau < T$, relative to the start of the symbol interval, which makes it invariant over time shifts of integral multiples of the symbol period $T$.

The phase response has a finite memory of length $L$. Input symbols that were generated more than $L$ symbols before the current symbol cannot influence the shape or slope of the phase trajectory in the current interval. However, they do contribute a phase offset or accumulated phase which depends on the history of the phase trajectory since $t = 0$. This accumulated phase is represented by the phase state $v_n$ of (1.10), which contains only time-independent data-dependent terms.

Phases that differ by multiples of $2\pi$ are physically indistinguishable, and therefore phase trees have to be wrapped into *phase trellises* by reducing them modulo $2\pi$. It is easy to show that reducing $\psi(\tau, \boldsymbol{u})$ modulo $2\pi$, reduces the time-independent terms in $\psi(\tau, \boldsymbol{u})/(2\pi h)$ modulo $P$. ($P$ was defined earlier in (1.4) as the denominator of the rational modulation index $h$.) The phase state $v_n$ is therefore defined as the sum of all the input symbols up to $L$ symbols before the current one, modulo $P$. The single phase value $2\pi h v_n$ records the accumulated phase since $t = 0$ up to $t = (n - L + 1)T$, in order to provide the required phase continuity.

The data-independent component $\theta(\tau)$ of (1.11) provides a constant frequency shift as well as a constant phase offset[1] to ensure that the phase tree starts at 0 radians at time $t = 0$. This offset component is not especially important, except for Rimoldi's observation that it should have a constant first derivative with respect to time in order not to increase the bandwidth of the signal unnecessarily without increasing the information rate.

Once again, the information-carrying phase $\psi(\tau, \boldsymbol{u})$ in (1.8) may be extended arbitrarily to general (discontinuous) phase modulation, by not insisting on the continuous-phase definition of (1.9).

### 1.4.1 The memoryless modulator

Examining (1.9), it is apparent that the CPM modulator output is completely determined in any symbol interval $n$ by the vector

$$x_n = [u_n, u_{n-1}, \cdots, u_{n-L+1}, v_n] \qquad (1.12)$$

[1] The expression in (1.11) is the correct one, there is a small mistake in Rimoldi's equation (11) in [65].

comprising the current input $u_n$, the $L-1$ previous data inputs which still influence the phase trajectory, and the phase state $v_n$. If $x_n$ is the input to a modulator which maps to a set of time waveforms to produce the CPM output, then that modulator need not have any memory of previous inputs. In this sense, it is an MM.

The concept can be made explicit as follows. In (1.8) and (1.9), the shifted-time variable

$$\tau = t - nT, \quad 0 \leq \tau < T \tag{1.13}$$

was implicitly introduced as a way of showing that the tilted phase $\psi(\tau, \boldsymbol{u})$ is time-invariant. Substituting (1.13) into (1.8) and writing $s(\tau, x_n)$ instead of $s(t, \boldsymbol{u})$, and $\psi(\tau, x_n)$ instead of $\psi(\tau, \boldsymbol{u})$ (thereby restricting our attention to the current symbol interval $n$ only), we obtain

$$s(\tau, x_n) = \sqrt{\frac{2E}{T}} \cos\left[\omega_0(\tau + nT) + \psi(\tau, x_n) + \varphi_0\right], \quad 0 \leq \tau < T, \;\; n = 0, 1, 2, \ldots \tag{1.14}$$

Apart from the carrier $\omega_0(\tau + nT)$, the waveform is entirely specified in terms of the relative time $\tau$. It is not really necessary to do so, but $s(\tau, x_n)$ can be made independent of the absolute time $t = \tau + nT$ by constraining $\omega_0 T = 2\pi k$ for some integer $k$. When this is done, the signal set in every symbol interval is identical to that in any other symbol interval. The important quantity is $\psi(\tau, x_n)$, which must be invariant over time shifts of multiples of $T$ for this description to be valid.

The MM can be implemented as a table look-up of baseband waveforms (specified by $\psi(\tau, x_n)$) followed by a frequency conversion up to $\omega_0$, to place the signal in the desired passband. Figure 1.3 shows the block diagram, which can be implemented practically as an in-phase and quadrature (I-Q) modulator.

### 1.4.2 The continuous-phase encoder

Some manipulation of the phase state definition $v_n$ from (1.10) results in the recursive formulation

$$v_{n+1} = \left[v_n + u_{n-L+1}\right]_{\text{mod } P} \tag{1.15}$$

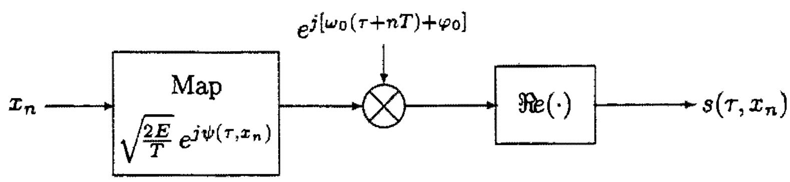



Figure 1.3: Memoryless modulator implemented as a baseband waveform mapping followed by frequency translation.

where by definition $v_0 = 0$. The vector input $x_n$ to the MM can be realized as shown in Figure 1.4.

The *state* of the CPE is defined as the last $L - 1$ input symbols together with the phase state:

$$\sigma_n = [u_{n-1}, u_{n-2}, \cdots, u_{n-L+1}, v_n] \tag{1.16}$$

A code trellis diagram can be drawn for the CPE similar to those for convolutional encoders [24]. Each state (dot) in the trellis represents a value of the phase state $\sigma_n$ in interval $n$, and each branch (line) represents a particular output signal during that symbol interval. Branches are also sometimes called *edges*. A trellis diagram for $M = 4$ CPM with $h = 1/4$ and $L = 1$ is shown in Figure 1.5. There are $M = 4$ branches from each state, and $PM^{L-1} = 4$ states.

The information contained in a trellis diagram can be represented equally well in a state transition matrix [44], or analytically as a functional transformation [1, 16], but trellis diagrams will suffice for the present purpose.

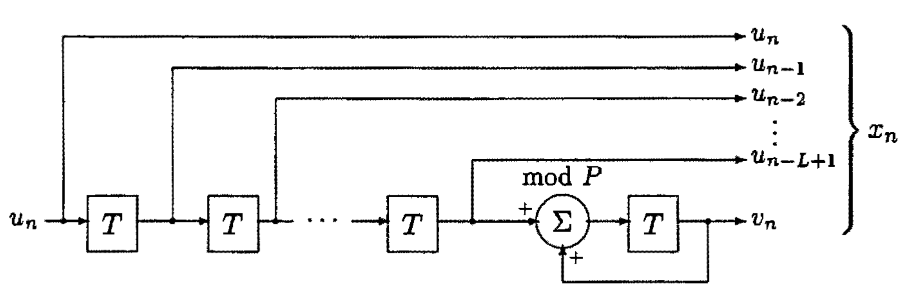



Figure 1.4: The CPE for general CPM, a linear sequential circuit with one data input $u_n$ and $L + 1$ outputs forming the vector $x_n$. $T$ represents a time delay of one symbol interval.

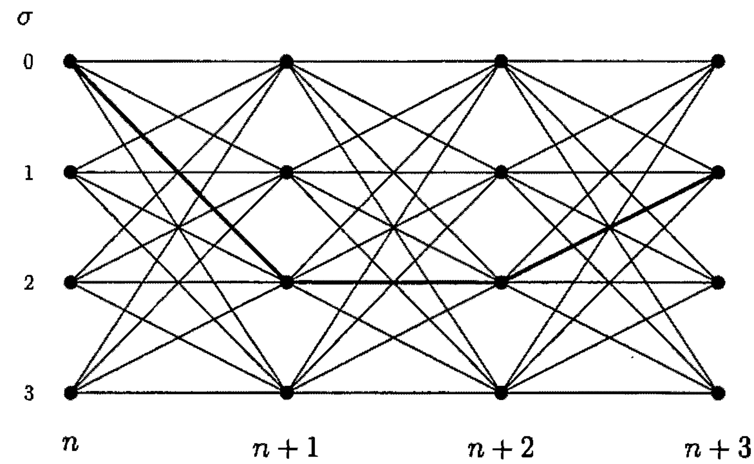



Figure 1.5: Trellis diagram for the CPE of $M = 4$ CPM with $L = 1$ and $h = 1/4$. The signal trajectory for the data input sequence $(2, 0, 3)$ is shown in a heavy line, assuming $\sigma_n = 0$ at $t = nT$.

### 1.4.3 The advantages of the decomposition approach

Although it had long been recognized that the continuous-phase property of the waveform represents a form of coding, the traditional description of CPM does not make a clear separation of the continuous-phase encoding from the rest of the modulation, which could be described as a memoryless mapping.

The decomposition approach has two important implications. The first one is that the CPE can be studied separately as a time-invariant linear sequential circuit, using the same theory as that developed for convolutional encoders. It can be cascaded or combined with an external convolutional encoder to increase the free Euclidean distance (see section 1.9 below) of the resulting coded modulation, while ensuring that the continuous-phase requirement of the transmitted waveform is met.

In contrast, in the traditional description the information-carrying phase is not cyclostationary, in the sense that the set of all possible phase trajectories in even-numbered symbol intervals is not a time translate of the set in odd-numbered intervals. The corresponding signal set therefore also varies between even and odd intervals.

The second implication is that the MM may be cascaded with the channel and a demodulator operating over one symbol interval, to form a discrete memoryless channel [90]. This channel can then be studied, with the extensive theory available for it, without the distraction of having to provide continuous phase explicitly.

The following is an example of the kind of manipulation that is possible with the decomposition approach, and introduces differential CPM:

Still following Rimoldi [65], it is assumed that the modulation index is $h = K/P$ with $K$ and $P$ mutually prime integers. I further assume that $P = M$, in order to bring the input symbols $u_n$ and the phase state $v_n$ into the same finite field modulo $P$. This is in any case a common choice for $P$ in practice [3], although Rimoldi considers the more general case where $M = P^k$ for some positive integer $k$. The choice of $k = 1$ simplifies the derivation offered below while retaining the essence of the argument.

With the above assumptions, the CPE defined in (1.15) and Figure 1.4 has the transfer function

$$G(D) = \frac{X(D)}{U(D)} = \left[1, D, D^2, \cdots, D^{L-1}, \frac{D^L}{1-D}\right] \tag{1.17}$$

where

$$U(D) = u_n + u_{n-1}D + u_{n-2}D^2 + \cdots \tag{1.18}$$

is the $D$-transform of the input sequence, $X(D)$ is the $D$-transform of the output vector $x_n$, and $D$ is the delay operator [44].

Factoring out $1/(1-D)$ from (1.17) and reorganizing the result gives

$$\begin{aligned} G(D) &= \frac{1}{1-D}\left[(1-D), D(1-D), \cdots, D^{L-1}(1-D), D^L\right] \\ &= \frac{1}{1-D}\left[1, D, \cdots, D^L\right] \boldsymbol{A} \end{aligned} \tag{1.19}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{1.20}$$

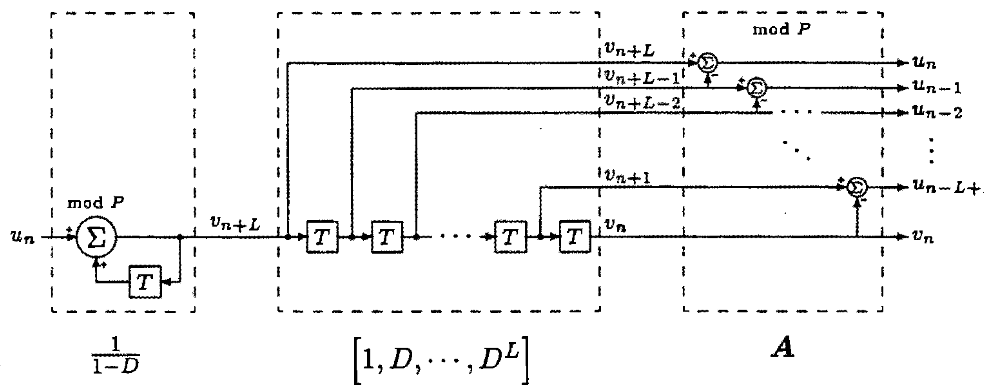



Figure 1.6: Manipulation of the basic CPE to produce a differential CPE and MM. If the first dashed box is omitted, differential CPM is produced. The last dashed box may be merged with the MM, leaving the central dashed box as the new CPE.

is a memoryless $(L+1) \times (L+1)$ mapping matrix with elements

$$a_{ij} = \begin{cases} 1 & i = j \\ -1 & i = j+1 \\ 0 & \text{otherwise} \end{cases} \tag{1.21}$$

The resulting structure is shown in Figure 1.6.

Two remarks can be made here. Firstly, factoring out and removing the $1/(1+D)$ precoding is permissible and results in an equivalent CPE because the input symbols are assumed to occur with equal *a priori* probability. The input becomes the data sequence $\boldsymbol{v}$ and the resulting modulation is *differential CPM*: CPM with the input data differentially encoded before transmission.[2]

Secondly, the $\boldsymbol{A}$-mapping may be incorporated into the MM, because it represents a memoryless mapping of the vector $x'_n = [v_{n+L}, v_{n+L-1}, \cdots, v_n]$ into $x_n$ (see Figure 1.6). If $x'_n$ is used as the input to the differential MM instead of $x_n$, the only thing that changes is its mapping from input to the particular output waveform selected.

This leaves the differential CPE with transfer function

$$G_M(D) = \left[1, D, \cdots, D^L\right] \tag{1.22}$$

[2]Rimoldi's definition of differential CPM was broader to include modulation indices with $M = P^k$. Differential CPM is in general not equivalent to his *modified CPM* which results when $k > 1$. My narrower definition will suffice for the schemes of interest here.

which can be implemented simply as a shift register. Also note that the current phase state $v_n$ is equal to the symbol $v_{n+L}$ which was the input to the CPE $L$ symbols previously. The input symbols are associated with future phase states, and the differences between consecutive symbols contribute to the current frequency. In contrast, the non-differential inputs $u_n$ directly contribute to the current signal frequency (see (1.9)), and sum together to form future phase states ((1.10) or (1.15)).

It is also interesting to note that absorbing the $\boldsymbol{A}$-mapping into the MM transforms its input-to-signal mapping from a linear code to a Gray code, at least for binary inputs.

### 1.4.4 Signal notation for tilted CPFSK representation

The tilted-phase representation of full-response CPFSK is obtained by substituting (1.7) into (1.9) with $L = 1$:

$$\psi(\tau, \boldsymbol{u}) = 2\pi h(v_n + u_n\tau/T), \quad 0 \leq \tau < T \tag{1.23}$$

where $v_n$ is defined as before in (1.10) or (1.15), and $\theta(\tau)$ reduces to 0. This is the phase tree illustrated in Figure 1.2(b) for $M = 2$.

I now define a convenient signal notation for full-response signalling in this dissertation, based on the MM concept. When $L = 1$ in (1.14), $x_n = [u_n, v_n]$. The output signal in the $n$-th interval is then called

$$s_{vu}(\tau) = s(t - nT, x_n) \tag{1.24}$$

where the $\tau$ makes explicit that the waveform does not depend on the absolute time $t$, and $v$ and $u$ refer to the inputs $v_n$ and $u_n$ respectively. This makes it easier to refer to a certain signal in a given symbol interval, without having to refer to the entire input sequence. For example, if $v_n = 2$ and $u_n = 3$, the signal is called $s_{23}(\tau)$ during the $n$-th interval.

Note that the signal $s_{vu}(\tau)$ refers to conventional non-differential CPM in which the $u$ in the subscript refers to the current signalling frequency selected by

$u_n$. This is a more general notation than choosing to have the pair of subscripts refer to the current and next state, because the signal is then not restricted to the condition that $P \geq M$, as had to be assumed for differential CPM in (1.17) and (1.19).

The $s_{vu}(\tau)$ notation is only applicable to full-response signalling, because the signal depends on the current phase state $v_n$ and the current input $u_n$ only.

## 1.5 Minimum Shift Keying

Minimum shift keying (MSK) is of particular interest in this dissertation, because it is the only CPM for which maximum likelihood coherent CPM receivers[3] exist which do not use either the Viterbi algorithm or the likelihood parameter receiver to be discussed in section 1.7. If one restricts the modulation index to $h = 1/2$, simplified suboptimal receivers may be designed for other CPM schemes than MSK, by adapting an MSK receiver [77, 78, 2, 55, 3]. Practical CPM receivers with $h \neq 1/2$ are designed using the average matched filter (AMF) concept [56, 32] at low signal-to-noise ratio, or linear pulse shape optimization [41, 54] at high signal-to-noise ratio.

### 1.5.1 The MSK modulator

MSK is full-response CPFSK with $M = 2$ and $h = 1/2$. It is the cornerstone of CPM, although it has many interpretations that enable it to span more than one category of modulation scheme [20, 30, 58, 67]. In terms of the CPM notation developed in section 1.4.4, the MSK signal can be described as

$$s(t, \boldsymbol{u}) = \sqrt{\frac{2E}{T}} \cos[\omega_0 t + \psi(t - nT, \boldsymbol{u}) + \varphi_0], \quad t \geq 0 \tag{1.25}$$

where

$$\psi(\tau, \boldsymbol{u}) = \pi(v_n + u_n \tau / T), \quad 0 \leq \tau < T \tag{1.26}$$

[3] I prefer to use the word 'receiver' for the entire practical reception system, which includes carrier synchronization and symbol timing extraction, and to use 'detector', 'demodulator' or 'decoder' for symbol detection or sequence estimation. 'Receiver' is however widely used to denote the latter concept only, with coherent synchronization usually assumed. I shall therefore use these terms interchangeably.

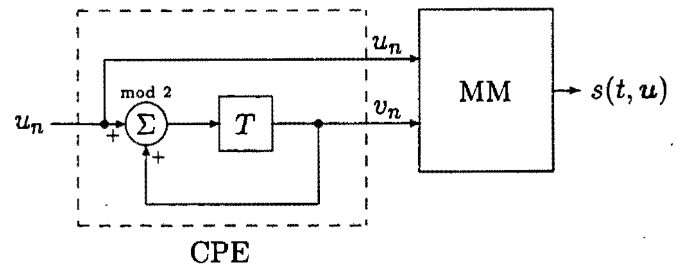



Figure 1.7: MSK (FFSK) modulator block diagram, showing decomposition into a CPE and an MM.

is the information-carrying phase and $\boldsymbol{u} = (u_0, u_1, u_2, \ldots)$ is the input data sequence, $u_n \in \{0, 1\}$, $n = 0, 1, 2, \ldots$

The phase state is

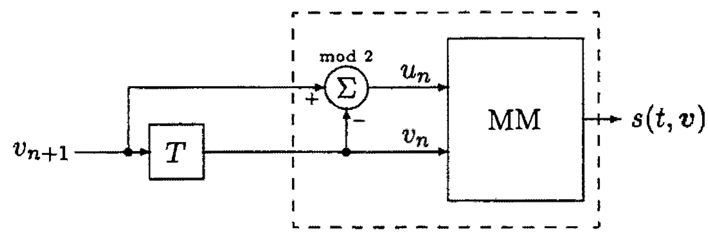
$$\sigma_n = v_n = \left[\sum_{i=0}^{n-1} u_i\right]_{\text{mod } 2}, \quad n = 1, 2, 3, \ldots \tag{1.27}$$

The modulator block diagram is shown in Figure 1.7.

There is a direct mapping of input data $u_n$ to signal frequency in the modulator of Figure 1.7: $u_n = 0$ produces the lower signalling frequency at the output and $u_n = 1$ maps to the higher frequency. 'Straight' MSK with this data-to-frequency mapping was called fast frequency shift keying (FFSK) by de Buda, who first invented a receiver for it [20].

The differential CPM description of MSK produces differential MSK (sometimes called DMSK). The differential MSK equivalent of Figure 1.6 is shown in Figure 1.8. The current input data is now the next phase state $v_{n+1}$, instead of the frequency (or phase increment) as before. The actual frequency transmitted is $\omega_0 + \pi[v_{n+1} - v_n]_{\text{mod } 2}$. In other words, if the last two input symbols are the same, transmit the lower signalling frequency, if they differ, transmit the higher frequency.

Differential MSK was in fact invented first [22], and was simply called minimum shift keying. Subsequently, it has become known as Type I MSK [58, 59]. In general, one may use one of several different data-to-frequency mappings for MSK [15], but the two mentioned here are sufficiently representative. I shall assume the differential MSK (Type I) signal for the rest of this chapter, because




Modified MM

Figure 1.8: Differential MSK (Type I) modulator block diagram. The input symbol is explicitly labelled $v_{n+1}$ here, instead of the usual $u_n$, for comparison with Figure 1.7. The $\mathcal{A}$-mapping may be combined with the MM as shown in the dashed box.

it is the more general description underlying the various MSK receivers that are possible.

If it is assumed for convenience that $\omega_0 T = 2\pi k$ for some integer $k$, the MSK signal set can be written entirely in terms of the time $\tau$ relative to the start of the symbol interval. Table 1.1 lists the output signals $s_{vu}(\tau)$ for all combinations of the input pair $(v_n, u_n)$. The signal notation $s_{vu}(\tau)$ was defined in section 1.4.4. Referring to Figure 1.8, the current transmitter input is $v_{n+1}$, and the internal frequency variable $u_n$ is derived from $u_n = [v_{n+1} - v_n]_{\text{mod } 2}$. The higher signalling frequency is denoted by $\omega_1$, which for $h = 1/2$ is $\omega_1 = \omega_0 + \pi/T$. I have also assumed $\varphi_0 = 0$.

It is convenient here to stress the formulation of MSK as a CPM. This is not the only formulation of MSK, or even the principal formulation. MSK is in fact best known as an offset quadrature modulation with sinusoidally shaped pulses [51, 30, 58]. This view arises naturally from the description of the MSK signal set as comprising two orthogonal pairs of antipodal signals,

| $v_n$ | $v_{n+1}$ | $u_n$ | $s_{vu}(\tau)$ |
|---|---|---|---|
| 0 | 0 | 0 | $s_{00}(\tau) = \sqrt{\frac{2E}{T}} \cos \omega_0 \tau$ |
| 0 | 1 | 1 | $s_{01}(\tau) = \sqrt{\frac{2E}{T}} \cos \omega_1 \tau$ |
| 1 | 0 | 1 | $s_{11}(\tau) = -s_{01}(\tau)$ |
| 1 | 1 | 0 | $s_{10}(\tau) = -s_{00}(\tau)$ |

Table 1.1: The differential MSK (Type I) signal set.

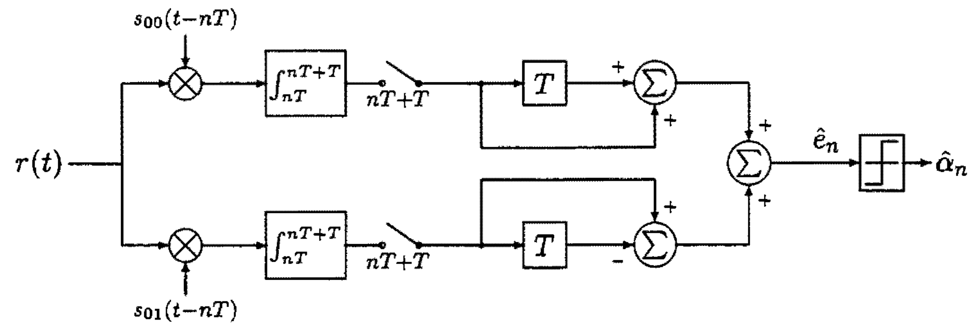



Figure 1.9: The Massey detector for differential MSK.

as described in standard communications textbooks [59, 31, 42, 63, 73]. It is this quadrature description which underlies the practical MSK modulator and receiver implementations first proposed by de Buda [20].

It is also no accident that certain types of offset quadrature phase shift keying (QPSK) and half-rate coded QPSK can be received with a receiver which is identical to Figure 1.9 except for the choice of matched filters [48]: these possibilities will be discussed in detail in the remainder of the thesis.

There is no conflict between different descriptions of MSK: they can each be translated to any other, and which one to use depends solely on what characteristic of the scheme one wants to emphasize. Rimoldi gives an interesting synthesis of five different descriptions of MSK in [67], including one which regards MSK as a re-transmission of the *same* data symbols, staggered in time, in orthogonal channels.

### 1.5.2 Massey's MSK detector

An interesting optimal coherent receiver for MSK is due to Massey [48]. The Massey detector for differential MSK is shown in Figure 1.9.

De Buda was the first to show that MSK may be optimally demodulated over two symbol intervals only [20]. Massey [48] obtained the same result by different means. I present Massey's proof here because it is related to the simplification of MSK's likelihood ratio, to be discussed in Chapter 3. The MSK trellis is shown in Figure 1.10. The branches are labelled with the signals transmitted during that interval. Note that this labelling is independent of the

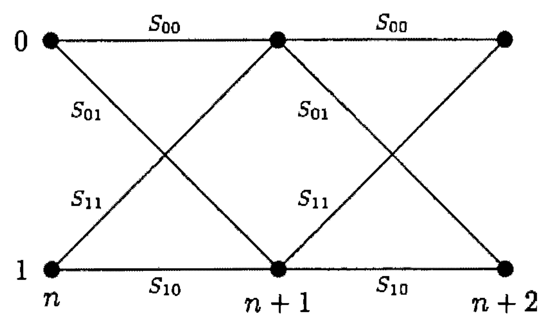



Figure 1.10: The MSK trellis, with branches labelled with the transmitted signal.

actual data-to-signal mapping, so that the diagram is valid for both FFSK and Type I MSK.

To start with, assume that all symbols are equally probable. This means that the optimal receiver need only be a maximum likelihood receiver, rather than a more general maximum *a posteriori* probability (MAP) receiver [90, p.214]. It also has to be assumed that the signals $s_{vu}(\tau)$ have equal energy $E$, to make comparisons between them meaningful.

Following the approach of Wozencraft and Jacobs [90, p.419], assume that a 'genie' tells us that the phase state at time $t = nT$ is $\sigma_n = 0$, and that $\sigma_{n+2} = 0$ as well. The MSK receiver now has to decide, after observing the received waveform $r(t)$, which of two trajectories the transmitted signal followed. Did it go via state $\sigma_{n+1} = 0$, or via $\sigma_{n+1} = 1$? The maximum likelihood decision in the presence of AWGN is to choose [90, p.238]

$$\lambda_n(S_{00}) + \lambda_{n+1}(S_{00}) \underset{\sigma_{n+1}=1}{\overset{\sigma_{n+1}=0}{\gtrless}} \lambda_n(S_{01}) + \lambda_{n+1}(S_{11}) \tag{1.28}$$

where

$$\lambda_n(S_{vu}) = \int_{nT}^{nT+T} r(t) s_{vu}(t - nT)\, dt \tag{1.29}$$

and where I have shortened $s_{vu}(\tau)$ to $S_{vu}$.

On the other hand, if the 'genie' had told us that the final state was $\sigma_{n+2} = 1$, the optimal decision would be

$$\lambda_n(S_{00}) + \lambda_{n+1}(S_{01}) \underset{\sigma_{n+1}=1}{\overset{\sigma_{n+1}=0}{\gtrless}} \lambda_n(S_{01}) + \lambda_{n+1}(S_{10}) \tag{1.30}$$

Two more similar decision rules can be formulated for the cases where it is known that the initial state is $\sigma_n = 1$:

$$\lambda_n(S_{11}) + \lambda_{n+1}(S_{00}) \underset{\sigma_{n+1}=1}{\overset{\sigma_{n+1}=0}{\gtrless}} \lambda_n(S_{10}) + \lambda_{n+1}(S_{11}) \tag{1.31}$$

$$\lambda_n(S_{11}) + \lambda_{n+1}(S_{01}) \underset{\sigma_{n+1}=1}{\overset{\sigma_{n+1}=0}{\gtrless}} \lambda_n(S_{10}) + \lambda_{n+1}(S_{10}) \tag{1.32}$$

Which decision rule to use seems to depend on the choice of initial and final states. But, from Table 1.1 it can be seen that $s_{10}(\tau) = -s_{00}(\tau)$, and $s_{11}(\tau) = -s_{01}(\tau)$. This implies that

$$\lambda_n(S_{10}) = -\lambda_n(S_{00}) \quad \text{and} \quad \lambda_n(S_{11}) = -\lambda_n(S_{01}), \qquad n = 0, 1, 2, \ldots \tag{1.33}$$

Substituting (1.33) into (1.28) and (1.30), and the other cases as well, shows that they are all the same decision rule: choose

$$\lambda_n(S_{00}) + \lambda_{n+1}(S_{00}) - \lambda_n(S_{01}) + \lambda_{n+1}(S_{01}) \underset{\sigma_{n+1}=1}{\overset{\sigma_{n+1}=0}{\gtrless}} 0 \tag{1.34}$$

Having decided what state the signal went through, what was the transmitted symbol? In differential MSK, the current phase state $\sigma_n$ is simply the previous input symbol $v_n$. The decision rule (1.34) therefore directly detects the transmitted symbol $v_n$ after a total delay of two symbol periods. If the FFSK mapping had been used in the transmitter, the Massey receiver would have to be followed by a differential decoder $(1+D)$ to obtain the input symbol [65, 15].

The output of the Massey detector $\hat{e}_n$ is a polar value, nominally $\pm 2E$ in the absence of noise. After hard limiting, $\hat{\alpha}_n \in \{\pm 1\}$, which is translated to the estimated input data through

$$\hat{v}_{n-2} = \frac{1 - \hat{\alpha}_n}{2} \tag{1.35}$$

Before hard limiting, the output $\hat{e}_n$ ideally has just two possible amplitudes in the absence of noise. (Strictly speaking, at low carrier frequencies relative to the bit rate the last statement is exactly true for $\omega_c = \pi k/(2T)$ where $k$ is an

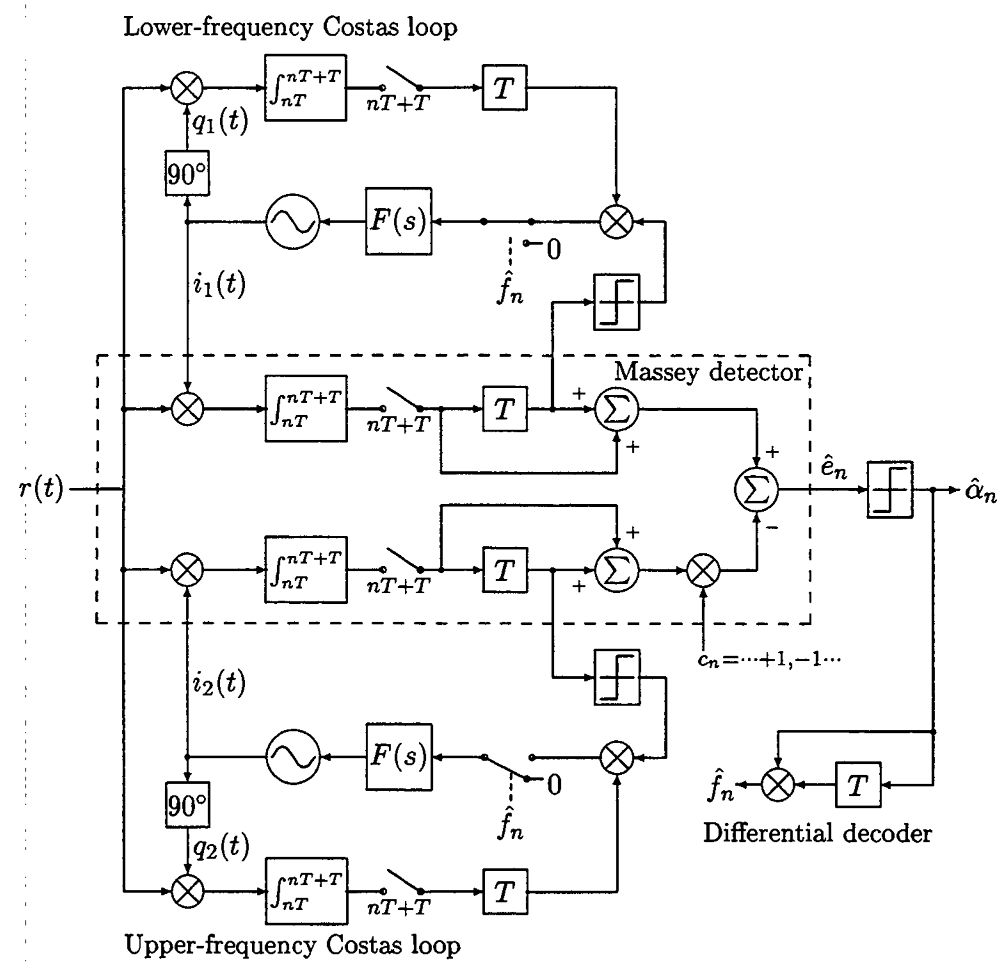



Figure 1.11: Massey-Hodgart MSK receiver block diagram. The section in the dashed box is the Massey detector. Quadrature correlators are added above and below this to complete two Costas loops for carrier recovery. Bit clock recovery is not shown here.

odd integer, $k \geq 3$ [72].) In an AWGN channel these two possible amplitudes are corrupted by the additive noise.

If one should choose to hard-limit $\hat{e}_n$, then the system has made a maximum likelihood bit decision $\hat{\alpha}_n = \text{sgn}(\hat{e}_n)$ at this point. Its theoretical performance will follow the standard coherent antipodal error rate performance curve. Alternatively one could choose *not* to limit the signal at this point and pass $\hat{e}_n$ on to a soft-decision decoder, assuming that the MSK modulation is part of a higher-level coded structure [49].

Massey's MSK receiver described above was rediscovered independently by Hodgart [35], who completed it by integrating it with carrier and symbol synchronization circuits. The Massey-Hodgart MSK receiver is interesting because of its simplicity, robustness and good performance, on a par with the 'classical' MSK receiver implementations of de Buda [20] and Amoroso and Kivett [2]. Its block diagram is shown in Figure 1.11, and more detail may be found in [35]. It is the only known practical MSK receiver implementation which explicitly uses the Massey decoder.

## 1.6 Classical Detection Theory

This section briefly summarizes the expressions for the likelihood ratio test, the MAP detector, maximum likelihood detector, and composite hypothesis test, to provide a basis for understanding the origin of likelihood parameter receivers and the optimality of Massey's MSK receiver. The material in this section is covered in more depth in, for example, Van Trees [80] or Davenport and Root [19].

The likelihood ratio test is usually described as a decision on binary hypotheses $H_0$ or $H_1$, only one of which may occur at any one time. In the communications context, the hypotheses usually refer to one of two symbols being transmitted through a channel. The criterion of optimality used for deriving the likelihood ratio test is the Bayes criterion of minimizing some average 'cost'. The likelihood ratio is the quotient of two likelihood functions, one for each hypothesis:

$$\Lambda(R) = \frac{f_{R|H_1}(R|H_1)}{f_{R|H_0}(R|H_0)} \tag{1.36}$$

It expresses the likelihood that hypothesis $H_1$ is true as opposed to hypothesis $H_0$, as a function of the observation variable $R$. $R$ may in general be a vector of random variables, but I shall only need the one-dimensional case here. Because $R$ is a random variable, $\Lambda(R)$ is one too. To make a decision on the hypotheses, an observation $R = r$ is needed and then $\Lambda(r)$ is compared against a threshold $\eta$:

$$\Lambda(r) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{1.37}$$

The $\gtrless$ notation means that hypothesis $H_1$ is chosen if $\Lambda(r) > \eta$, else $H_0$ is chosen. This is the likelihood ratio test.

The likelihood parameter $f_{R|H_k}(r|H_k)$ is the probability density of observing the outcome $R = r$ given that $H_k$ is true. It is easy to write an expression for such a density function if AWGN is assumed at the receiver: for a one-dimensional observation $r$ it is simply the familiar normal distribution, with mean equal to the expected value of the observation (given that $H_k$ is true), and variance equal to the two-sided noise spectral density $N_0/2$:

$$f_{R|H_k}(r|H_k) = \frac{e^{-\frac{1}{N_0}[r-E(R|H_k)]^2}}{\sqrt{\pi N_0}} \tag{1.38}$$

where $E(\cdot)$ is the expectation operator.

The threshold $\eta$ is chosen according to the *a priori* probabilities of each hypothesis, and the costs associated with each decision. If the common assumption is made (valid for most communication systems) that the differences in cost between an incorrect and a correct decision are equal for each hypothesis (that is, choosing $H_1$ when $H_0$ is in fact true, is as 'bad' as choosing $H_0$ when $H_1$ is true), then

$$\eta = \frac{P(H_0)}{P(H_1)} \tag{1.39}$$

where $P(H_k)$ is the *a priori* probability that $H_k$ is true. With this special cost assignment, it can be shown that the decision also minimizes the average probability of error.

Another useful formulation is the *maximum a posteriori* (MAP) detector, which is easily derived from the likelihood ratio test by using Bayes' rule. The MAP detector chooses the most probable hypothesis, given a particular observation $R = r$, in order to minimize the average probability of error:

$$P(H_1|R=r) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0|R=r) \tag{1.40}$$

If it is further assumed, for either the likelihood ratio test or the MAP detector, that the *a priori* probabilities of the hypotheses are equal, that is $P(H_0) = P(H_1) = 1/2$, then the threshold $\eta = 1$. A decision can be made by simply comparing the likelihood parameters

$$f_{R|H_1}(r|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} f_{R|H_0}(r|H_0) \tag{1.41}$$

for the observation $R = r$. [4]

This is called the *maximum likelihood* decision, and it is optimal (in the sense of minimizing the average probability of error) only for equally likely hypotheses and equally costly errors. In the more general case of $M$ hypotheses (corresponding to $M$-ary signalling), the maximum likelihood detector makes an observation $R = r$, evaluates all $M$ likelihood functions, and chooses the hypothesis (transmitted message) corresponding to the largest likelihood parameter.

In terms of data transmission systems, the maximum likelihood decision can be shown to be equivalent to choosing the symbol (hypothesis) corresponding to the transmitted signal closest to the observation $R = r$. The concept of 'closest' is usually defined in an orthonormal signal space, which is exactly analogous to Euclidean space. Euclidean distances between CPM signals are defined in section 1.9.

Sometimes, the likelihood functions depend on parameters in addition to the observation variable $R$. For example, if two consecutive observations of two consecutive symbols in a signal are made, with the aim of making a decision on the first symbol only, the information in the second observation may well contribute usefully to the decision. In MSK, for example, the requirement of continuous phase constrains the choice of signal in any given symbol interval, depending on the signal in the previous interval. Without knowing what the second symbol was, however, the second symbol becomes a parameter in the likelihood functions of the test, which now has so-called composite hypotheses. If such an 'unwanted' parameter $U$ is random and its distributions are known,

[4] There does not seem to be a consistent definition in the literature of these terms. I shall use *likelihood function* to refer to the probability distribution function of a random variable $R$ which appears in either the numerator or the denominator of the likelihood ratio $\Lambda(R)$, and *likelihood parameter* to refer to the value of a likelihood function which is calculated by the receiver on an observation $R = r$. Likelihood parameters are sometimes called *decision functions* [90, p.214].

then the composite hypothesis problem reduces to a likelihood ratio test

$$\begin{aligned}\Lambda(R) &= \frac{f_{R|H_1}(R|H_1)}{f_{R|H_0}(R|H_0)} \\ &= \frac{\int_U f_{R|U,H_1}(R|U,H_1) f_{U|H_1}(U|H_1)\, dU}{\int_U f_{R|U,H_0}(R|U,H_0) f_{U|H_0}(U|H_0)\, dU}\end{aligned} \tag{1.42}$$

where the integration is over the range of $U$. Once again, $U$ may be a vector of parameters, in which case the integration is done over each component in turn.

The compound hypothesis test is useful when the likelihood function is only known in terms of the unwanted parameter $U$ and the observation variable $R$, in which case $U$ is then integrated out. The next section on likelihood parameter receivers provides an example of the use of the composite hypothesis test.

## 1.7 Likelihood Parameter Receivers

An asymptotically optimal receiver for coherent binary CPFSK with arbitrary modulation index, was first described by Osborne and Luntz in 1974 [56]. Schonhoff [71] extended the concept in 1976 to include $M$-ary CPFSK. Their receivers for both binary and $M$-ary CPFSK are directly based on the likelihood ratio test. The receiver makes an optimal (for the given observation period) decision on the first received symbol after observing the received signal for a fixed number of symbol periods $N$. The longer the observation period $N$ becomes, the closer to a global maximum likelihood the decisions become.

This class of receivers directly evaluates likelihood functions to yield a set of likelihood parameters. I shall call them *likelihood parameter receivers.*

The significance of the CPFSK likelihood parameter receiver is due to the error performance bounds that may be calculated from it for various modulation indices $h$ and observation periods $N$. It is clear that likelihood parameter receivers are hardly practical in their 'pure' form. Apart from an exponential growth of the number of reference sequences with length of observation period, likelihood parameter receivers are nonlinear, and require *a priori* knowledge of either the signal-to-noise ratio $E/N_0$ or the noise power spectral density $N_0$.

The following description of the likelihood parameter receiver for binary and

$M$-ary CPFSK follows the papers by Osborne and Luntz [56] and Schonhoff [71]. Although they derived their receivers for full-response CPFSK, these are quite generally applicable to CPM with any phase response and modulation index [3, p.238].

The full-response CPFSK signal may be described by (1.8) and (1.23):

$$s(t, \boldsymbol{u}) = \sqrt{\frac{2E}{T}} \cos[\omega_0 t + \psi(t - nT, \boldsymbol{u}) + \varphi_0], \quad t \geq 0 \tag{1.43}$$

where

$$\psi(\tau, \boldsymbol{u}) = 2\pi h \left( \sum_{i=0}^{n-1} u_i + u_n \tau / T \right), \quad 0 \leq \tau < T \tag{1.44}$$

is the information-carrying phase during the $n$-th symbol interval, and $\boldsymbol{u} = (u_0, u_1, u_2, \ldots)$ is the input data sequence. For $M$-ary CPFSK we have $u_n \in \{0, 1, \ldots, M-1\}$, $n = 0, 1, 2, \ldots$

The received signal $r(t) = s(t, \boldsymbol{u}) + w(t)$ is corrupted with AWGN with power spectral density $N_0/2$. Carrier and symbol synchronization is assumed, and therefore the arbitrary starting phase $\varphi_0$ may be assumed to be known and set to 0 without loss of generality. The assumption of a known starting phase is fundamental to all the receiver designs considered here, and it has important implications for the interpretation of the detected data. This is discussed in more detail in section 3.4.1.

I now misuse the current notation somewhat to follow Osborne and Luntz's convention, and denote the signal during the observation period as $s(t, u_0, \boldsymbol{U})$. $u_0$ is the first symbol, the one to be decided, and $\boldsymbol{U}_i$, $i = 1, 2, \ldots, Q$ is the set of all possible data sequences each comprising the $(N-1)$-vector $[u_1, u_2, \ldots, u_{N-1}]$, and $Q = M^{N-1}$ is the total number of such vectors. Thus the sequence of $N$ symbols in the observation window is partitioned into the first symbol, and the rest.

This is a composite hypothesis problem in which the $N-1$ symbols in $\boldsymbol{U}$ are the unwanted parameters. The maximum likelihood decision is made by comparing the $M$ likelihood parameters

$$l_j(r) = \int_{\boldsymbol{U}} f_{R|\boldsymbol{U}}(r|\boldsymbol{U}, u_0 = j) f_{\boldsymbol{U}|u_0}(\boldsymbol{U}|u_0) \, d\boldsymbol{U}, \quad j = 0, 1, \ldots, M-1 \tag{1.45}$$

and choosing the symbol corresponding to the largest one.[5] The integral over $\boldsymbol{U}$ is in fact the $(N-1)$-fold integral

$$\int_{\boldsymbol{U}} \cdots \, d\boldsymbol{U} = \int_{u_{N-1}} \int_{u_{N-2}} \cdots \int_{u_1} \cdots \, du_1 \, du_2 \cdots du_{N-1}$$

The joint probability density function of $\boldsymbol{U}$ is given by

$$\begin{aligned} f_{\boldsymbol{U}|u_0}(\boldsymbol{U}|u_0 = j) &= f_{\boldsymbol{U}}(\boldsymbol{U}) \\ &= f_{u_1}(u_1) f_{u_2}(u_2) \cdots f_{u_{N-1}}(u_{N-1}) \end{aligned} \tag{1.46}$$

where

$$f_{u_k}(u_k) = \frac{1}{M} \sum_{i=0}^{M-1} \delta(u_k - i), \quad k = 1, 2, \ldots, N-1 \tag{1.47}$$

because the input symbols are assumed to be independent and identically distributed.

The maximum likelihood receiver includes a bank of $M$ matched filters [90]. If the input to the $j$-th matched filter is $r(t) = s(t, k, \boldsymbol{U}) + w(t)$, where $s(t, k, \boldsymbol{U})$ is the transmitted signal, then the sampled matched filter output is a random variable

$$R_j = S_{kj} + W_j \tag{1.48}$$

where

$$R_j = \int_0^{NT} r(t) s(t, j, \boldsymbol{U}) \, dt \tag{1.49}$$

$$S_{kj} = \int_0^{NT} s(t, k, \boldsymbol{U}) s(t, j, \boldsymbol{U}) \, dt \tag{1.50}$$

$$W_j = \int_0^{NT} w(t) s(t, j, \boldsymbol{U}) \, dt \tag{1.51}$$

$R_j$ is normally distributed with mean $E(R_j) = S_{kj}$ and variance $E[(R_j - S_{kj})^2] = NE N_0/2$. The likelihood function is then

$$f_{R_j|\boldsymbol{U}}(R_j|\boldsymbol{U}) = \frac{e^{-\frac{1}{NEN_0}(R_j - S_{kj})^2}}{\sqrt{\pi NEN_0}}$$

[5]In the interest of simple notation, I use $u_k$ to refer both to an input symbol value and its random variable. The meaning should be clear from the context.

$$= \frac{e^{-\frac{1}{NEN_0}(R_j^2+S_{kj}^2)}}{\sqrt{\pi N_0}} \cdot e^{\frac{2}{NEN_0}R_j S_{kj}} \tag{1.52}$$

If all the transmitted signals are assumed to have equal energy $\int_0^T s^2(t,k,\boldsymbol{U})\,dt = E,\ k = 0,1,\ldots,M-1$ and if the signal $s(t,k,\boldsymbol{U}) = s(t,j,\boldsymbol{U})$ was transmitted, then

$$\begin{aligned} f_{R_j|\boldsymbol{U}}(R_j|\boldsymbol{U},u_0=j) &= \frac{e^{-\frac{1}{NEN_0}\left[R_j^2+(NE)^2\right]}}{\sqrt{\pi NEN_0}} \cdot e^{\frac{2}{NEN_0}\int_0^{NT} r(t)s(t,j,\boldsymbol{U})\,dt \cdot NE} \\ &= F \cdot e^{\frac{2}{N_0}\int_0^{NT} r(t)s(t,j,\boldsymbol{U})\,dt} \end{aligned} \tag{1.53}$$

If equal-energy signals are assumed, the first exponential factor in (1.53) is independent of $u_0 = j$ and may be omitted from the comparison of the likelihood functions as a constant factor $F$.

Removing such constant factors from (1.53), and substituting it together with (1.46) and (1.47) in (1.45), gives $M$ likelihood parameters:

$$\begin{aligned} l_j(r) &= e^{\frac{2}{N_0}\int_0^{NT} r(t)s(t,j,\boldsymbol{U}_1)\,dt} + e^{\frac{2}{N_0}\int_0^{NT} r(t)s(t,j,\boldsymbol{U}_2)\,dt} + \cdots \\ &\cdots + e^{\frac{2}{N_0}\int_0^{NT} r(t)s(t,j,\boldsymbol{U}_Q)\,dt}, \quad j = 0,1,\ldots,M-1 \end{aligned} \tag{1.54}$$

These are compared and the symbol corresponding to the largest one selected as the maximum likelihood estimate. Figure 1.12 shows a block diagram of the likelihood parameter receiver for $M$-ary CPFSK.

The likelihood parameter receiver of (1.54) leads directly to thinking of ways of approximating $e^x$, particularly through truncated power series. For example, the Maclaurin series approximation to $M$-th order is

$$e^x \approx 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^M}{M!} \tag{1.55}$$

A first-order linear approximation $e^x = 1+x$ leads to the *average matched filter* mentioned before as a simplified suboptimal receiver [56, 32][3, p.329].

Higher order approximations may be used, and Taylor series expansions may be taken around different centres in an attempt at optimizing the receiver for the expected $E/N_0$. This approach will not be followed further here.

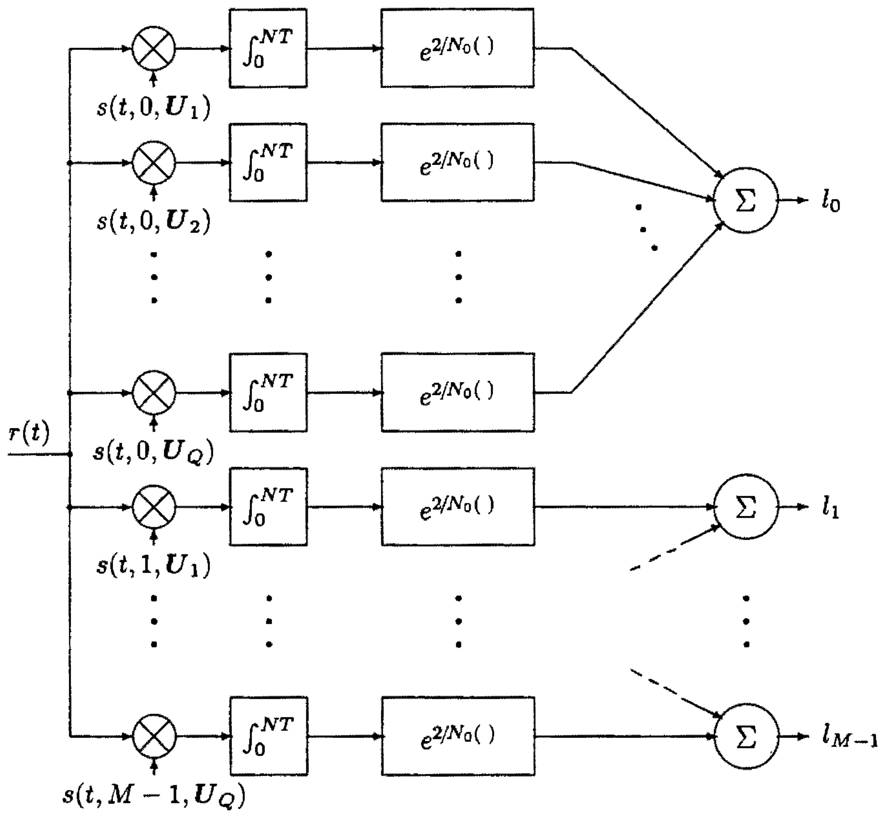



Figure 1.12: The likelihood parameter receiver for $M$-ary CPFSK.

It may not have been recognized before that the likelihood parameter receiver is optimal (in the maximum likelihood sense) for any coded modulation with equal-energy signals, not only for CPM. It is regarded as the basic MLSE receiver for any coded modulation in the context of this dissertation, and is equally applicable to TCM with equal-energy signals.

## 1.8 The Gelfond-Schneider Theorem

In Paris in 1900, at the International Congress of Mathematicians, the great mathematician David Hilbert presented a short list of 23 unsolved problems. "Hilbert's problems", which include Fermat's Last Theorem and the Riemann Hypothesis, subsequently became famous and had a considerable influence on the development of modern mathematics [11, p.9][39, p.268].

Hilbert's seventh problem raised the question of whether an irrational logarithm of an algebraic number to an algebraic base is transcendental. (An irrational number is one which cannot be written as a quotient of integers. An algebraic number is one which can be written as a root of a polynomial with rational coefficients. A transcendental number is one which is not algebraic.) It was solved in 1934 independently by Gelfond and Schneider [11]. The Gelfond-Schneider theorem states that for any non-zero algebraic numbers $x_1$, $x_2$, $y_1$ and $y_2$,

$$y_1 \log x_1 \neq y_2 \log x_2 \tag{1.56}$$

when $\log x_1$ and $\log x_2$ are linearly independent over the rational numbers [11, p.10].

The Gelfond-Schneider theorem has an important implication for the simplification of expressions that contain exponentials, such as (1.54) above. It means that exponentials with irrational bases and irrational exponents, such as $e^x$, $x$ irrational, are transcendental, that is, if $f()$ is an algebraic function, then $f(e^{x_1}, e^{x_2}, \cdots) = 0$ has no solutions for irrational $x_1, x_2, \ldots$ The number $x$ is restricted to the irrationals in order to exclude trivial identities.

Any expression containing exponentials to the base $e$ can only be simplified by treating the exponentials as indivisible units and doing the usual algebraic simplification on these units. One should not expect to find identities of exponentials to aid simplification, because exponentials are transcendental and identities therefore do not exist. This property is the basis of the likelihood transform, which is defined in Chapter 3. [6]

## 1.9 Euclidean Distances

This section defines the normalized squared Euclidean distance (NSED) and free Euclidean distance, and points out their importance to the error rate of a modulation scheme. This information may be found in more detail in many texts, for example [90], [3] or [14].

The maximum likelihood receiver evaluates the 'distance' between the received signal and a reference signal to decide how closely they match. For a modulation scheme to have good error performance, its signals must therefore

[6] I am indebted to Dr. Ken Hughes of the Department of Mathematics at the University of Cape Town, for bringing the Gelfond-Schneider theorem to my attention.

be as widely separated in signal space as possible, subject to the total energy available. The closest pair of signals (or closest pair of paths in a trellis of signals) overwhelmingly determines the error performance at moderately high signal-to-noise ratio.

It is convenient to define a distance in signal space, called a Euclidean distance because of its close analogy with Euclidean space, with which to characterize the error performance of CPM schemes. Assume that two equal-energy transmissions $s(t, \boldsymbol{u})$ and $s(t, \boldsymbol{u}')$ differ over an observation period $0 \leq t < NT$, and are identical otherwise. The maximum likelihood receiver would calculate the squared Euclidean distance between them as

$$\begin{aligned} \|s - s'\|^2 &= \int_0^{NT} [s(t, \boldsymbol{u}) - s(t, \boldsymbol{u}')]^2 \, dt \\ &= 2NE - 2\int_0^{NT} s(t, \boldsymbol{u}) s(t, \boldsymbol{u}') \, dt \end{aligned} \tag{1.57}$$

where all integrals of double-frequency terms have been neglected by assuming that the carrier frequency $\omega_0 \gg 2\pi/T$.

The distance between signals which differ only in the phase difference between them, can be represented more simply than in the general case of signals with unequal amplitude envelopes. For equal-energy signalling, minimizing the Euclidean distance is the same as maximizing the final integral in (1.57). The value of this interval, evaluated between a received waveform and one of the signals from the signal set, is the metric used by the maximum likelihood receiver to decide the likelihood that the particular signal was sent. The integral may be implemented directly as a correlation (with multipliers and integrators) or as a matched filter (because of the relationship between correlation and convolution).

Substituting the CPM signal waveforms from (1.8) and doing some trigonometric substitutions gives

$$\begin{aligned} \|s - s'\|^2 &= 2NE - \tfrac{4E}{T} \sum_{n=0}^{N-1} \int_0^T \cos[\omega_0(\tau + nT) + \psi(\tau, \boldsymbol{u}) + \varphi_0] \cdot \\ &\qquad \cos[\omega_0(\tau + nT) + \psi(\tau, \boldsymbol{u}') + \varphi_0] \, d\tau \\ &= 2NE - \frac{2E}{T} \sum_{n=0}^{N-1} \int_0^T \cos[\psi(\boldsymbol{\tau}, \boldsymbol{u}) - \psi(\boldsymbol{\tau}, \boldsymbol{u}')] \, d\tau \end{aligned}$$

$$= \frac{2E}{T} \sum_{n=0}^{N-1} \int_0^T [1 - \cos \Delta\psi(\tau)]\, d\tau \tag{1.58}$$

where double-frequency terms have again been neglected, and

$$\Delta\psi(\tau) = \psi(\tau, \boldsymbol{u}) - \psi(\tau, \boldsymbol{u}') \tag{1.59}$$

The *normalized* Euclidean distance function $d()$ is defined as

$$\begin{aligned} d^2(s, s') &= \frac{\|s - s'\|^2}{2E_b} \\ &= \frac{\log_2 M}{T} \sum_{n=0}^{N-1} \int_0^T [1 - \cos \Delta\psi(\tau)]\, d\tau \end{aligned} \tag{1.60}$$

where the bit energy $E_b$ is related to the symbol energy $E$ by

$$E_b = \frac{E}{\log_2 M} \tag{1.61}$$

(1.60) is commonly known as the NSED between two CPM signals. It depends on the number of bits per symbol ($\log_2 M$), the length of the observation period ($N$), and the phase separation between the signals ($\Delta\psi(\tau)$) during that interval. The squared Euclidean distance of (1.58) and the NSED of (1.60) are valid for any phase modulation, not only for CPM.

In coded modulation (and in this context CPM may be thought of as a form of coded modulation), the distance of interest is the *free Euclidean distance*

$$d^2_{\text{free}} = \min_{s \neq s'} d^2(s, s'), \quad t \geq 0 \tag{1.62}$$

In words, the free Euclidean distance is the smallest NSED between any two different signal trajectories of a transmission, measured over infinite time. In practice one does not have to measure for too long, once it is recognized that the smallest distances inevitably result from those trajectories which are identical for all time, except for an interval during which they part at a certain code or phase state and merge again soon thereafter.

The free Euclidean distance is important because it dominates the union bound expression for the error rate of the modulation at reasonably high signal-to-noise ratio. The union bound on the probability of wrongly deciding a received signal given that $s_i(t)$ was transmitted, is given by [90, p.265]

$$P\left[\varepsilon|s_i(t)\right] \le \sum_{k\neq i} P_2\left[s_i(t), s_k(t)\right] \tag{1.63}$$

where $P_2[s_i(t), s_k(t)]$ is the probability of deciding that signal $s_k(t)$ was received when in fact $s_i(t)$ was transmitted. The 2 in the subscript emphasizes the fact that only pairs of signals need be considered at a time for the probability calculation.

In AWGN, (1.63) becomes

$$P\left[\varepsilon|s_i(t)\right] \le \sum_{k\neq i} Q\left(\frac{\|s_i(t)-s_k(t)\|}{\sqrt{2N_0}}\right) = \sum_{k\neq i} Q\left(\sqrt{d^2(s_i(t), s_k(t))\frac{E_b}{N_0}}\right) \tag{1.64}$$

The familiar normal integral function

$$Q(x) = \int_x^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \tag{1.65}$$

is steeply monotonically descending for large $x$. For reasonably large signal-to-noise ratio $E_b/N_0$, the term with smallest Euclidean distance $d^2(s_i(t), s_k(t))$ in the summation will strongly dominate the sum.

For practical calculations of Euclidean distances, it can be shown that

$$d^2(s, s') = N \log_2 M - \rho(s, s') \tag{1.66}$$

where

$$\begin{aligned}\rho(s, s') &= \frac{1}{E_b}\int_0^{NT} s(t, \boldsymbol{u}) s(t, \boldsymbol{u}')\, dt \\ &= \frac{\log_2 M}{T}\sum_{n=0}^{N-1}\int_0^T \cos\Delta\psi(\tau)\, d\tau \end{aligned} \tag{1.67}$$

is the normalized correlation coefficient between two equal-energy signals over $N$ symbol intervals. Minimizing $\rho(s, s')$ between two signals maximizes the NSED between them. Conversely, the maximum correlation found between any two signals gives the minimum NSED, which in turn determines the error performance. It is often convenient when calculating Euclidean distances, in a computer program for example, rather to calculate $\rho(s, s')$ and then simply subtract it from $N \log_2 M$ to get the NSED.

Specializing in full-response CPFSK now, define

$$\begin{aligned}
\Delta\psi(\tau) &= \psi(\tau, \boldsymbol{u}) - \psi(\tau, \boldsymbol{u}') \\
&= 2\pi h[v_n - v_n' + (u_n - u_n')\tau/T] \\
&= 2\pi h[\Delta v_n + \Delta u_n \tau/T], \quad 0 \leq \tau < T
\end{aligned} \tag{1.68}$$

where

$$\begin{aligned}
\Delta v_n &\in \{-(P-1), -(P-2), \ldots, (P-1)\} \\
\Delta u_n &\in \{-(M-1), -(M-2), \ldots, (M-1)\}
\end{aligned} \tag{1.69}$$

It can be seen from (1.58) and (1.68) that the NSED for full-response CPFSK can be calculated directly from the sequence of phase states $v_n$ and sequence of input symbols $u_n$, $0 \leq n \leq N-1$. This is accomplished practically by defining an incremental correlation coefficient $\Delta\rho(\Delta v_n, \Delta u_n)$ as the integral in (1.67) over a single interval $0 \leq \tau < T$. It is precalculated for the particular $M$ and $h$ of interest, and the values are written in a two-dimensional look-up table as a function of $\Delta v_n$ and $\Delta u_n$. Given a particular state and input sequence, the differences $\Delta v_n = v_n - v_n'$ and $\Delta u_n = u_n - u_n'$ are calculated. The full correlation coefficient $\rho(s, s')$ is then formed by summation from the look-up table, and the NSED follows easily from there. Table 4.4 gives an example of such a look-up table.

# Chapter 2

# Problem Statement, Objectives, and Literature Review

This chapter presents the problem statement in section 2.1, elaborates it into more specific objectives in section 2.2, and concludes with a review of the published literature in section 2.3.

## 2.1 Problem Statement

The hypothesis of this thesis is that there exist coded modulation schemes, apart from minimum shift keying (MSK), with simplified receivers. MSK, as a member of the continuous phase modulation (CPM) family, is regarded here as a form of coded modulation. A coded modulation has a simplified receiver if its likelihood ratio simplifies into a single term.[1] The likelihood ratios of uncoded modulations always simplify.

To be quite clear on the hypothesis, this dissertation does *not* propose to find a construction method or design method for simplified receivers (or simple

[1]This definition is slightly ambiguous because it is possible for some coded schemes to have different likelihood ratios in different encoder states, or for the likelihood ratios to depend nonlinearly on the signals sent from a state. A clearer definition and some examples of the process of writing likelihood ratios and simplifying them are given in Chapter 3.

receivers in general), unless such a method contributes to the verification or falsification of the hypothesis. It is proposed that the discovery of *any* simplified receiver apart from MSK's, or a proof of the uniqueness of MSK in this regard, would be of considerable interest in its own right.

It is not immediately apparent how to prove or disprove the hypothesis above analytically for schemes other than MSK, which is the only one known to have a simplified receiver. (MSK is assumed here to include other closely related schemes which can be received optimally with receivers which differ from Figure 1.9 only in the choice of front-end matched filters. Such schemes include general full-response CPM with $h = 1/2$, and 4-PSK with half-rate coding.) An attempt was made early in this project to construct an existence proof for the kinds of mathematical identities that would be needed to simplify likelihood ratios, but it was abandoned when the analysis seemed intractable.

Therefore, the approach taken initially was to assume that the hypothesis is true, and to attempt to disprove it by demonstrating that no simplified receivers exist for the modulation schemes of interest; or conversely, to assume that the hypothesis is false and to attempt to find simplified receivers as counter-examples. In both cases, the problem was to attempt to find simplified receivers, by exhaustive search if necessary.

## 2.2 Objectives

In order to define the first objectives of the project, it is necessary to decide what forms the search for simplified receivers should take. As mentioned in the introduction of Chapter 1, the space of all possible receivers (and by implication the space of all possible modulation codes) is vast, and any practical search is forced to constrain its search space somehow.

Constraining the search for simplified receivers is problematic for two reasons: Firstly, the likelihood ratio is a completely general formulation of any receiver, since it is a part of the general Bayesian hypothesis test [80]. Specifying that the focus of the search will be on the likelihood ratios of all possible receivers does not limit the search at all.

Secondly, the process of simplification of a likelihood ratio depends on mathematical identities which are not known in advance. All likelihood ratios are generated from a given code, trellis, or system description, and hence all likelihood ratios share certain characteristics. But it was not easy to formulate a mathematical description of likelihood ratios which demarcates it from arbitrary mathematical expressions, and which makes it easier to see which likelihood ratios simplify in the desired way. It is scant comfort to specify that only likelihood ratios derived from trellis codes will be considered, since the majority of practical modulation codes are trellis codes.

The method followed in this project differs from the analytical approaches followed by other researchers in that no particular algebraic form (such as groups or rings) is specified, nor are any of the usual heuristic techniques employed (such as set partitioning [82], lattice cosets [17], rotational invariance [86, 87], or other design rules [33]). It was decided at the outset to keep the search as general as possible in order to maximize the probability of discovering a simplified receiver.

In order to constrain the search space, a two-fold approach to the search was taken: I decided to focus an initial search on a specific coded modulation scheme, and to focus a search on mathematical identities that may be useful for simplifying likelihood ratios. If either of these resulted in a simplified receiver, the hypothesis would have been proven; and if no simplified receiver was found, the results of the searches might have potential for generalization into a non-existence proof.

In the choice of specific modulation scheme, one might as well choose a scheme with potentially good performance, having high spectral or power efficiency or both. CPM is emphasized because the one modulation known to have a simplified receiver is a CPM. The intention was to design the *combination* of the channel encoder (CE) and continuous-phase encoder (CPE) of a CPM with coding to have a simplified receiver.

Many CPM schemes have been evaluated and compared over the years, and a comprehensive catalogue of coded and uncoded schemes to choose from has been assembled by now [7, 3, 6, 10, 9, 54, 62, 74]. In terms of choice of modulation index, the cases of $h = 1/2$ and $h = 1/4$ stand out. Schemes with $h = 1/2$ possess several implementations of simple optimal and suboptimal receivers [20, 2, 21, 69, 92, 77, 70, 78], while schemes with $h = 1/4$ have an attractive combination of good power and bandwidth efficiency, particularly for $M > 2$ and with coding [3, p.438][6].

The modulation scheme chosen was a family of modulations called quadrature minimum shift keying (QMSK), proposed by Hodgart [34]. QMSK is related to both MSK and four-level continuous-phase frequency shift keying (CPFSK) with modulation index $h = 1/4$. It is trellis coded with specific constraints on the allowed signal sequences designed to facilitate carrier synchronization in the receiver. It is called a family of modulations because, apart from these constraints, the CE is left unspecified. QMSK was chosen because of its relationship to MSK, and because its defining constraints usefully limit the space of possible codes to be searched. It is described in detail in section 4.2.

The search for mathematical identities takes as its starting point a general likelihood ratio derived from an unspecified trellis code of a given number of states and given observation period. The challenge is to design an algorithm which generates only mathematically distinct expressions from the general form, in order to avoid unnecessary attempted simplifications. (Mathematically distinct expressions are defined in Chapter 4 as expressions which can not be obtained from one another by simple substitution of variables.) The search for identities has the simplification of mathematical expressions in common with the search for simplified QMSK receivers, but it attempts to be more efficient by only examining likelihood ratios which are essentially different from one another.

The project then has three objectives:

1. To examine as many encoders for QMSK as possible, and find if any of them result in simplified receivers.

2. To examine as many mathematical expressions derived from trellis codes of a given size as possible, and find if any of them simplify into useful identities.

3. To examine the results of the work done on the first two objectives, and find if they can be used to prove or disprove the main hypothesis.

Several important issues have not been covered here, including catastrophic codes, time-shift invariance and rotational invariance. These will be brought up as necessary in the detailed description of the methods in Chapters 3 and 4.

## 2.3 Literature Review

This work falls broadly into the field of trellis-coded modulation (TCM). The idea of improving the performance of digital communication systems by combining the functions of the (error-correcting) CE and the modulator, and demodulating them together, emerged from a wealth of early work in coding theory and was probably first explicitly suggested by Massey [47] in 1974. Massey covered the case of binary transmission, and the concept was extended to the $M$-ary case by Lee [43] in 1976. Concrete proposals on how to implement these ideas followed with the work of Imai and Hirakawa [37] in 1977 and Ungerboeck [83, 82] in 1976 and 1982. It was especially Ungerboeck's famous 1982 paper [82] on set partitioning which precipitated an avalanche of new research. Of note was the analytical description of the CE by Calderbank and Mazo [16] and rotationally invariant channel coding by Wei [86, 87] in 1984, and systematic search techniques for good TCM schemes in 1990 by Pietrobon *et al* [60]. The book by Biglieri *et al* [14] gives a good introduction and overview of developments in TCM up to 1991.

A tutorial paper in 1984, again by Massey [49], proposed a formal decomposition of the modulator into a CE which embodies all the memory of the transmitter (but excluding source or secrecy encoding), and a memoryless modulator (MM). This view was taken up by Rimoldi [64, 65], who showed in 1988 that the transmitter of any CPM scheme can always be decomposed into a CPE and an MM such that the CPE is a linear, time-invariant sequential circuit and the MM is also time-invariant. Before Massey and Rimoldi's work, so-called bandwidth-efficient CPM schemes were not commonly recognized as belonging essentially in the field of TCM.

The immediate impetus for this project was Massey's 1980 paper [48] describing a simple maximum likelihood receiver for MSK. I came across it while I was involved in a project to evaluate and test an MSK receiver invented by Hodgart [35], which uses the Massey detector and integrates it with an efficient and robust carrier and symbol timing recovery. A DSP implementation of the Massey-Hodgart receiver performs remarkably well [35]. In the light of these results, the question has to be asked whether another modulation scheme may benefit from a similar simple receiver. The only other mention I have found of this possibility, is in the conclusion of Rimoldi's doctoral dissertation [64, p.129] of 1988:

> "Exploiting the fact that the tilted-phase ... leads to a time-invariant trellis, new forms of decoding algorithms might be found; an example has been given for MSK ... which is a special case of CPM."

The earliest mention of the simplification of MSK's likelihood ratio appears to be a somewhat enigmatic footnote in the well-known paper by Forney [24] on the Viterbi algorithm in 1973. I quote the footnote in full:

> "De Buda [20] actually proves that an optimum decision on the phase $x_k$at time $k$ can be made by examining the received waveform only at times $k-1$ and $k$; i.e., $(z_{0,k-1}, z_{1,k-1}), (z_{0k}, z_{1k})$. The proof is that the log likelihood ratio
>
> $$-\ln\frac{P(z_{0,k-1}, z_{1,k-1}, z_{0k}, z_{1k}|x_{k-1}, x_k = 0, x_{k+1})}{P(z_{0,k-1}, z_{1,k-1}, z_{0k}, z_{1k}|x_{k-1}, x_k = \pi, x_{k+1})}$$
>
> is proportional to $-z_{0,k-1} + z_{1,k-1} - z_{0k} - z_{1k}$ for any values of the pair of states $(x_{k-1}, x_{k+1})$. Fot this phase decision (which differs slightly from our sequence decision) the error probability is exactly $Q(\sqrt{2}/2\sigma)$." [My citation [20], corresponding to Forney's [44].]

In fact, de Buda [20] makes no mention of the likelihood ratio or its simplification, at least not in the cited paper.

In his 1980 paper [48], Massey is more explicit:

> "... the [MSK] demodulator ... is optimum in the sense of maximizing the cut-off rate, $R_0$, of the discrete channel between the modulator input and the demodulator output (and also in the sense of maximizing the capacity, $C$, of this channel).
>
> "To prove this claim, one must show that the demodulator ... *preserves the likelihood ratio* for the decision on [the transmitted data], since any operation on [the received signal] reduces $R_0$ (and also $C$) unless and only unless this likelihood ratio is preserved." [Massey's italics]

He omits the proof, but gives the simplified likelihood ratio for MSK, which I derive in detail in section 3.1.

Apart from the MSK receiver, I know of no other reference in the literature to a similar simplification of a likelihood ratio. MSK is a CPM, and the point of departure was therefore to examine other CPM schemes, especially CPFSK, and the combination of channel coding with CPFSK modulators.

MSK itself provides a rich source of interesting and sometimes surprising interpretations. It was invented in 1961 by Doelz and Heald [22] and rediscovered in 1972 by de Buda [20]. Mathwich, Balcewicz and Hecht [51] in 1974, and Gronemeyer and McBride [30] in 1976, showed that MSK can be described as a linear offset quadrature phase shift keying with sinusoidally shaped pulses. Amoroso and Kivett [2] invented the serial MSK modulator in 1977, which regards MSK as a form of antipodal signalling. Massey [48] showed that the Viterbi algorithm implementation of the MSK receiver reduces to his simple receiver in 1980. In 1994, Rimoldi [67] described a 'diversity' implementation of MSK signalling which transmits each symbol twice, staggered in time, over two orthogonal channels. A synthesis of these different views was offered most recently by Rimoldi [67].

The most general form of a receiver for binary CPM was derived by Osborne and Luntz [56] in 1974, followed by the extension to the $M$-ary case by Schonhoff [71] in 1976. The Osborne-Luntz receiver is none other than a direct application of the likelihood ratio to the composite hypothesis decision problem

[80] which results from considering a sequence of received signals and making a maximum likelihood decision on the first signal. No attempt was made to simplify the likelihood ratio (in most cases it cannot be simplified), and the general CPM 'receiver' is in fact only of use, because of its complexity, for predicting theoretical error rates.

The application of trellis coding (in addition to the CPE) to improve the error performance of CPM schemes dates almost from the same time that CPM was first recognized as a useful class of power and bandwidth efficient modulations by Anderson *et al* [7], and Aulin, Rydbeck and Sundberg [10, 9]. An early paper on trellis phase coding by Anderson and de Buda [4] in 1976, in fact anticipated the TCM revolution by emphasizing the importance of soft-decision demodulation in Euclidean space. It was followed by the invention of multi-$h$ coding of CPM in 1978 by Anderson and Taylor [8].

Since 1984, trellis-coded CPM has become an active field of research with many contributors [45, 61, 13, 33, 54, 66, 6]. In the majority of papers, however, the assumption is made that for optimal coherent reception, a Viterbi decoder is present in the receiver. Where receiver complexity is mentioned at all, the emphasis is on reducing the number of states in the Viterbi decoder (for example [54]) or approximating the signal in some way (for example [41]).

The published literature which most closely resemble my work reported here, in spirit if not in letter, fall into two groups: The first group comprise papers which assume that the receiver will be a linear time-invariant filter, and then optimize that filter for given modulation schemes, linear and nonlinear. The second group comprise papers on so-called matched coding, in which convolutional encoders are chosen for CPM schemes in such a way that the combined CE and CPE has fewer states than the product of the number of states of each separately.

In the first group, the co-authors Galko and Pasupathy considered binary modulation with symmetrical signal sets [27] in 1982, and then generalized it to find the optimum linear receiver for a generic digital modulation scheme [29] in January 1988. One month earlier, Svensson [76] published a similar paper, apparently independently, but considered only schemes with binary input. A paper by Svensson and Sundberg [77] in 1984 considered CPM with modulation index $h = 1/2$ in fading channels, which led to several papers on optimizing linear MSK-type receivers for reception of other CPM schemes [28, 78, 3, among others].

The first papers mentioned above by Galko and Pasupathy [29] and Svensson [76] are sufficiently general and similar enough to be discussed together. They take as a basic assumption that the receiver shall be a linear filter, the output of which is sampled and compared to a threshold or level slicer, for a general composite hypothesis testing problem [80] corresponding to an unspecified digital modulation in AWGN. The signal space is taken to be all the possible sequences of signals in a given observation period, which may be much longer than the symbol period, similarly to the approach taken by Osborne and Luntz [56] for CPFSK. Observing a received signal for longer than one symbol period for optimal demodulation may be necessary because of coding, including continuous-phase coding, or intersymbol interference, including partial response signalling [38]. The impulse response of the optimal filter and the optimal threshold(s) are then derived for a channel noise power which is assumed to be either infinitely large (poor signal-to-noise ratio) or vanishingly small (good signal-to-noise ratio). In a sense, this is an inverse approach to the method considered in this dissertation, which does not at the outset assume any particular receiver structure such as a linear filter.

In the second group of papers, the concept of matched codes is proposed and developed. The matched encoder for MSK was first defined by Morales-Moreno and Pasupathy [54] in 1988 as an encoder which, when paired with the MSK CPE, results in the CPE not affecting the complexity (number of states) of the Viterbi decoder in the receiver. In general, the number of states in the receiver's Viterbi decoder would be given by the product of the number of states in the CE and the number of phase states (in the CPE). The concept was then applied to a partial response MSK called tamed FM (TFM) by Morales-Moreno, Holubowicz and Pasupathy [53] in 1994 and most recently by Holubowicz and Morales-Moreno [36] to duobinary MSK with coding in 1995. In some cases, the complexity of the resulting receiver of a matched-encoded CPM is even lower than for the CPM without coding, although the coding gain is less.

Matched coding works by ensuring that some of the potential states in the transmitter are never reached, assuming that the transmitter starts in a known state. By potential states is meant the full list of combined encoder and phase states, equal in number to the product of the number of CE and CPE states. There is then no need for the receiver to make provision for all potential transmitter states, as long as the transmitter never leaves its orbit of permitted states, or at least returns to the correct orbit in a reasonable time if accidentally flushed from it. In spite of the dramatic simplification in receiver complexity, coupled with positive coding gain, that has been reported in some cases [36], no reduction to a complexity comparable to the MSK receiver has been reported.

# Chapter 3

# Likelihood Ratio Analysis of Continuous-Phase Frequency Shift Keying

It is well-known that the matched filter is the optimal detector of a communications signal over one symbol interval in additive white Gaussian noise (AWGN), in the sense of maximizing the signal-to-noise ratio at its output [90]. The sampled output level of the matched filter does not depend on the pulse shape of the received signal, only on its energy $E$. (The pulse shape is, however, important for other reasons, such as the bandwidth occupancy of the scheme or its susceptibility to intersymbol interference.)

For this reason, the error performance (also known as the energy or power efficiency) of an equal-energy modulation scheme is determined only by its signal energy $E$ (or, strictly, the signal-to-noise ratio) and the spacing between its signals or sequences of signals. It is not necessary to specify the exact pulse shapes when discussing error performance, as long as it is assumed that the receiver input is matched to the signal set.

In this chapter, therefore, continuous-phase frequency shift keying (CPFSK) is taken as conveniently representative of continuous phase modulation (CPM) more generally, and even representative of coded modulation in general.

Section 3.1 introduces likelihood ratio analysis by deriving the likelihood ratio test for minimum shift keying (MSK) and showing that it simplifies into the Massey receiver for MSK [48]. A concise and useful notation is introduced in section 3.2 for writing and simplifying likelihood ratios, called the likelihood transform. Section 3.3 returns to the MSK likelihood ratio derivation to show how the likelihood transform can be used to prove MSK's invariance over the starting phase state and over the length of the observation period. After this thorough analysis of MSK, the method is generalized to CPFSK and CPM with coding in section 3.4. Finally, catastrophic trellises are defined and discussed in section 3.5.

## 3.1 Deriving a Likelihood Ratio for MSK

In section 1.5.2, I repeated Massey's proof [48] of the optimality of his MSK detector. In section 1.7 I summarized the derivation of the likelihood parameter receiver of Osborne and Luntz [56] and Schonhoff [71]. Both these approaches yielded maximum likelihood detectors, although the Massey detector is very much simpler. An obvious question to ask is: what happens when the likelihood parameter receiver is applied to MSK? The interesting answer is that it simplifies into the Massey detector.

In this section I derive the Massey detector from the likelihood parameter receiver for general CPM, which is in effect another proof that the Massey detector makes maximum likelihood decisions. Massey states that this derivation is possible, but does not give it in his paper [48]. I assume, as elsewhere, that the channel noise is AWGN.

When considering a binary ($M = 2$) modulation, the set of $M$ likelihood parameters of (1.54) may be combined into a single likelihood ratio, which is exactly equivalent to the binary likelihood ratio test of (1.37). For the moment, we assume that the (differential) MSK signal, as defined by Figure 1.8 and Table 1.1, is in phase state $\sigma_0 = v_0 = 0$ at time $t = 0$, that the initial phase can be set arbitrarily to $\varphi_0 = 0$ because it is assumed known in coherent detection, and that the receiver observes the signal over $N = 2$ symbol intervals. These assumptions will be justified in section 3.3. We may assume that the starting time of the observation period is $t = 0$ with no loss of generality.

From (1.42) and (1.54) we have

$$\begin{aligned}\Lambda(R) &= \frac{f_{R|0}(R|v_1 = 0)}{f_{R|1}(R|v_1 = 1)} \\ &= \frac{e^{\frac{2}{N_0}\int_0^{2T} r(t)s(t,0,0)\,dt} + e^{\frac{2}{N_0}\int_0^{2T} r(t)s(t,0,1)\,dt}}{e^{\frac{2}{N_0}\int_0^{2T} r(t)s(t,1,1)\,dt} + e^{\frac{2}{N_0}\int_0^{2T} r(t)s(t,1,0)\,dt}}\end{aligned} \tag{3.1}$$

where Osborne and Luntz's convenient notation $s(t, v_1, v_2)$, $0 \leq t < 2T$ is used to represent the signal over two successive symbol intervals.

To write the integrals over single intervals, requires a return to the signal notation $s_{vu}(\tau)$, $0 \leq \tau < T$, defined previously in (1.24). It is convenient to use the frequency variable $u_n = [v_{n+1} - v_n]_{\bmod 2}$ again here. This gives

$$\begin{aligned}\Lambda(R) &= \frac{e^{\frac{2}{N_0}\int_0^{T} r(t)s_{00}(t)\,dt}\left[e^{\frac{2}{N_0}\int_T^{2T} r(t)s_{00}(t-T)\,dt} + e^{\frac{2}{N_0}\int_T^{2T} r(t)s_{01}(t-T)\,dt}\right]}{e^{\frac{2}{N_0}\int_0^{T} r(t)s_{01}(t)\,dt}\left[e^{\frac{2}{N_0}\int_T^{2T} r(t)s_{10}(t-T)\,dt} + e^{\frac{2}{N_0}\int_T^{2T} r(t)s_{11}(t-T)\,dt}\right]} \\ &= \frac{e^{\frac{2}{N_0}\lambda_0(S_{00})}\left[e^{\frac{2}{N_0}\lambda_1(S_{00})} + e^{\frac{2}{N_0}\lambda_1(S_{01})}\right]}{e^{\frac{2}{N_0}\lambda_0(S_{01})}\left[e^{\frac{2}{N_0}\lambda_1(S_{10})} + e^{\frac{2}{N_0}\lambda_1(S_{11})}\right]}\end{aligned} \tag{3.2}$$

where

$$\lambda_n(S_{vu}) = \int_{nT}^{nT+T} r(t)s_{vu}(t - nT)\,dt$$

is the metric defined previously in (1.29). Using the essential substitution (1.33)

$$\lambda_n(S_{10}) = -\lambda_n(S_{00}) \quad \text{and} \quad \lambda_n(S_{11}) = -\lambda_n(S_{01}), \qquad n = 0, 1, 2, \ldots$$

results in the simplification

$$\begin{aligned}\Lambda(R) &= \frac{e^{\frac{2}{N_0}\lambda_0(S_{00})}\left[e^{\frac{2}{N_0}\lambda_1(S_{00})} + e^{\frac{2}{N_0}\lambda_1(S_{01})}\right]}{e^{\frac{2}{N_0}\lambda_0(S_{01})}\left[e^{-\frac{2}{N_0}\lambda_1(S_{00})} + e^{-\frac{2}{N_0}\lambda_1(S_{01})}\right]} \\ &= \frac{e^{\frac{2}{N_0}\lambda_0(S_{00})}e^{\frac{2}{N_0}\lambda_1(S_{00})}e^{\frac{2}{N_0}\lambda_1(S_{01})}}{e^{\frac{2}{N_0}\lambda_0(S_{01})}} \\ &= e^{\frac{2}{N_0}[\lambda_0(S_{00})+\lambda_1(S_{00})-\lambda_0(S_{01})+\lambda_1(S_{01})]}\end{aligned} \tag{3.3}$$

Since we assume symbols transmitted with equal *a priori* probability, the likelihood ratio test — (1.37), (1.39) — for MSK is

$$\Lambda(R) \underset{1}{\overset{0}{\gtrless}} 1 \tag{3.4}$$

or, since logarithms are monotonically increasing functions,

$$\ln \Lambda(R) = \frac{2}{N_0}\left[\lambda_0(S_{00}) + \lambda_1(S_{00}) - \lambda_0(S_{01}) + \lambda_1(S_{01})\right] \underset{1}{\overset{0}{\gtrless}} 0 \tag{3.5}$$

This is exactly Massey's receiver (1.34) with $n = 0$. We derived (3.5) for $0 \leq t < 2T$, but it applies equally well for all $t \geq 0$. Note that the simplification in (3.3) and the subsequent taking of the logarithm renders the MSK receiver linear, as compared to the nonlinear general CPFSK receiver of (1.54).

## 3.2 The Likelihood Transform

### 3.2.1 Definition of the likelihood transform

The expressions (3.1) and (3.2) for MSK's likelihood ratio contain ratios of sums of exponentials. The Gelfond-Schneider theorem described in section 1.8 allows us to replace each exponential by a single variable without loss of generality. This leads to a concise and convenient notation for writing and simplifying likelihood ratios.

For example, in the simplification of the MSK likelihood ratio (3.3) above, one might as well replace each exponential by a single variable: let

$$\begin{aligned}
e^{\frac{2}{N_0}\lambda_n(S_{00})} &= a_{00} \\
e^{\frac{2}{N_0}\lambda_n(S_{01})} &= a_{01} \\
e^{\frac{2}{N_0}\lambda_{n+1}(S_{00})} &= b_{00} \\
e^{\frac{2}{N_0}\lambda_{n+1}(S_{01})} &= b_{01}
\end{aligned} \tag{3.6}$$

and $n = 0$ in this case. This choice of variables is arbitrary: here $a$ is used for the first interval and $b$ for the second, while the subscript reflects the signal. $a$,

$b$, $c$ and $d$ would have done as well.

We can now rewrite (3.3) much more simply as

$$\begin{aligned}\Lambda &= \frac{a_{00}\left[b_{00}+b_{01}\right]}{a_{01}\left[\frac{1}{b_{00}}+\frac{1}{b_{01}}\right]} \\ &= \frac{a_{00}\left[b_{00}+b_{01}\right]}{a_{01}\left[\frac{b_{00}+b_{01}}{b_{00}b_{01}}\right]} \\ &= \frac{a_{00}b_{00}b_{01}}{a_{01}} \end{aligned} \tag{3.7}$$

which translates into the result of (3.3).

If an expression such as the one we started with in (3.3) does not simplify algebraically after substitution of the exponentials, then it does not simplify at all. The Gelfond-Schneider theorem guarantees that no generality is lost in the substitution. I call this process of substituting variables for the exponentials in a likelihood ratio, a *likelihood transform.*

**Definition 1** *The expression that results from substituting single variables for every exponential in a likelihood function or ratio in AWGN is called its* likelihood transform.

Because the substitution variables replace exponentials, they may always be assumed to be positive and real. I refer to them as *likelihood transform variables* (LTVs) from here on.

### 3.2.2 The significance of the likelihood transform

The association between an LTV and its associated signal is close, as we have already seen in the choice of notation used in (3.6) above. I shall even sometimes use a notation such as $s_a(\tau)$ to refer to the signal associated with the LTV $a$ in the $n$-th interval. This close relationship makes it convenient to use LTVs in likelihood transforms without specifying the associated signals. In Chapter 4 for example, LTVs will be used without even specifying a signal set, making it possible to analyse generic receiver types without reference to specific modulation schemes. For this reason, LTVs will not always have subscripts explicitly indicating the associated signal as in (3.6).

The likelihood transform is therefore 'lossy' in that it does not preserve the definitions of the waveforms represented in the likelihood ratio. Like the matched filter, it only retains the energy content of the signal. This characteristic makes it possible to compare the maximum likelihood receivers and error performance of seemingly unrelated modulation schemes by examining the likelihood transforms of their likelihood ratios. It will not normally be necessary to write these likelihood ratios from first principles as has been done here so far: starting in section 3.3 and for the rest of the dissertation, it will be shown how easily likelihood transforms of likelihood ratios may be written directly from their code trees or trellises.

Later, an attempt will be made to simplify likelihood ratios of CPM schemes with coding, so one might as well look carefully at exactly how the simplification of the MSK likelihood ratio happened: the substitution (1.33) resulted in the cancellation of the sum $[b_{00} + b_{01}]$ from the numerator and denominator of the likelihood ratio, leaving a final expression comprising a single term.

Note that this cancellation does not happen when the likelihood parameters of (1.54) are compared separately. They have to be combined in pairs into likelihood ratios as we have done here, in order for cancellation of sums to be possible. This does not mean that only binary modulations can be analysed in this way: it works for $M$-ary schemes as long as the required simplification occurs when all $M$ likelihood parameters are combined pair-wise into ratios.

For future reference, it is useful to define exactly what is meant by the simplification of a likelihood ratio:

**Definition 2** *A likelihood ratio in AWGN is said to* simplify *if its likelihood transform can be written as a single term.*

A single term is understood to mean an expression comprising only products and quotients of variables. After taking the logarithm of a simplified likelihood ratio, a simplified maximum likelihood decision results. Achieving this for any modulation scheme instantly provides a simple maximum likelihood receiver for it. I shall call such a receiver a *simplified receiver*:

**Definition 3** *A* simplified receiver *is one whose structure is determined by the logarithm of the simplified likelihood ratio (or ratios) of its modulation scheme in AWGN.*

The usefulness of the likelihood transform will be demonstrated repeatedly throughout the rest of the dissertation, as a concise way of relating the structure of the maximum likelihood receiver to the scheme's trellis code, without explicit reference to a signal set.

## 3.3 The Assumptions in the MSK Likelihood Ratio

The rest of this chapter is devoted to illustrating some of the utility of the likelihood transform by offering an alternative analysis of the maximum likelihood MSK receiver and receivers for CPFSK more generally. The results for invariance over starting state and length of observation period are in fact well known and may be found in [3] for example.

In writing the likelihood ratio for MSK (3.1), we made two assumptions. One was that the starting phase state was $\sigma_n = v_n = 0$, and the other was that the observation period $N = 2$. In the absence of 'magic genies' [90, p.419] in practical receivers, we have to check that the same likelihood ratio simplification results when these assumptions are dropped.

### 3.3.1 Arbitrary starting phase state

To take into account different starting phase states, we add the symbol prior to the one we are detecting to the list of 'unwanted' parameters in the likelihood ratio. Previously in (3.1) we assumed $v_0 = 0$ and used the single unwanted parameter $\boldsymbol{U} = (v_2)$ while detecting $v_1$. I now move to a more general notation and specify $\boldsymbol{U} = (v_n, v_{n+2})$ while detecting $v_{n+1}$. In other words, the input symbols before *and* after the one we are detecting are regarded as unknown, with distributions given by (1.47) as before. The observation period remains at $N = 2$.

The MSK trellis diagram of Figure 1.10 is redrawn here in Figure 3.1 for reference. The branches are now labelled with their associated LTVs (3.6).

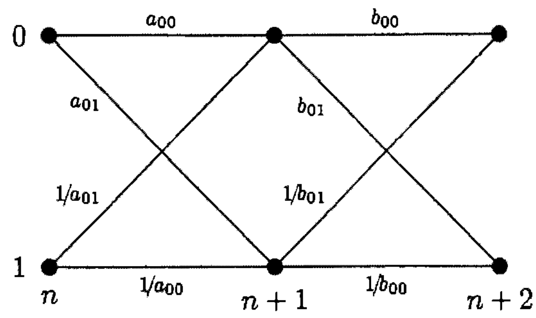



Figure 3.1: MSK trellis diagram with branches labelled with their LTVs.

It is easy to write the new likelihood ratio by inspection from the trellis of Figure 3.1. All the possible paths through phase state $v_{n+1} = 0$ appear as a sum of factors in the numerator, and all paths through $v_{n+1} = 1$ appear similarly in the denominator:

$$\begin{aligned}
\Lambda &= \frac{a_{00}b_{00} + a_{00}b_{01} + \frac{b_{00}}{a_{01}} + \frac{b_{01}}{a_{01}}}{\frac{a_{01}}{b_{00}} + \frac{a_{01}}{b_{01}} + \frac{1}{a_{00}b_{00}} + \frac{1}{a_{00}b_{01}}} \\
&= \frac{a_{00} + \frac{1}{a_{01}}}{\frac{1}{a_{00}} + a_{01}} \cdot \frac{b_{00} + b_{01}}{\frac{1}{b_{00}} + \frac{1}{b_{01}}} \\
&= \frac{a_{00}b_{00}b_{01}}{a_{01}} \qquad (3.8)
\end{aligned}$$

which is identical to (3.7). (The reader is urged at this point to examine carefully the correspondence between the expression in (3.8) and the trellis in Figure 3.1, bearing in mind that the LTVs in (3.8) are exponentials of correlation samples (metrics) of the form of (1.29). A clear understanding of this correspondence is essential for the rest of the dissertation.)

Without specifying the starting phase therefore, the likelihood ratio still simplifies to the same decision rule (3.5) obtained before. This validates the assumption of an arbitrary $v_n = 0$, or $v_0 = 0$ as we had it initially.

The derivation (3.8) of the invariance to starting phase only works because

1. the MSK trellis is invariant over phase rotations of $2\pi h$ radians; and

2. the MSK simplified likelihood ratios are identical from every starting state.

These properties might indeed have been predicted from the fact that the continuous-phase encoder (CPE) is linear and defined over the ring of integers modulo $P$ (with $P = 2$ in this case), as shown by Rimoldi [64, 65]. These conditions are not necessarily true for CPM schemes with coding.

It is interesting to note that the expressions in (3.8) factorize into two factors, each containing variables from one symbol interval only, as shown in the second step above. (The left-hand factor contains only $a$s and the right-hand factor only $b$s.) This is to be expected because the symbols in $\boldsymbol{v}$ are assumed to be statistically independent, so that one should be able to write their joint probability density function as a product of the conditional density functions associated with each symbol. Once again, this is not necessarily the case for coded schemes. When $M$ is less than the number of states of the encoder, not all the states are reachable from any given state in the previous interval. The sequence of states $\boldsymbol{v}$ therefore cannot be independent of one another as in (1.47), and the likelihood ratio will not factorize as in (3.8).

The two-fold tolerance to the starting phase state results directly in a two-fold ambiguity in the phase of the received signal, which the receiver cannot resolve. With differential MSK, the detected phase state is the estimate of the (delayed) input data, which may therefore be inverted in sign with respect to what was transmitted. If fast frequency shift keying (FFSK) had been transmitted, a differential decoding stage is needed after the (Type I) receiver in order to recover the input data, which has the effect of removing this ambiguity. The input data are not related directly to the phase state of the transmitted FFSK signal, only to changes in the phase state.

Showing that the likelihood ratio is invariant over the starting and ending phase states as we have done here, is an alternative way of showing the equivalence of (1.28) – (1.32), with the Massey receiver (1.34) as the result.

### 3.3.2 Limited observation period

What follows is an alternative proof of de Buda's assertion [20] that MSK can be optimally detected in two symbol intervals. It is well known that the optimal observation period for full-response CPM is two symbol intervals if the initial

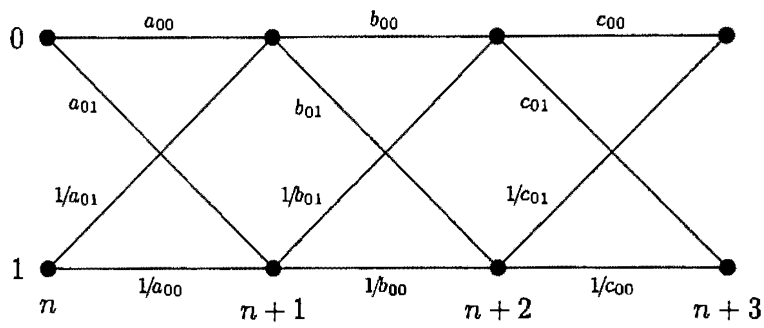



Figure 3.2: MSK trellis with the observation period extended to three symbols. Branches are labelled with LTVs.

state is known, and three symbol intervals if the initial state is not known or erroneous [3]. This result is reproven here for the special case of MSK as an illustration of how the likelihood transform simplifies such proofs.

The validation of the assumption of an observation period of $N = 2$ in the MSK likelihood ratio (3.1) proceeds similarly to the starting phase validation. We may now safely assume an arbitrary $v_n = 0$, and extend the MSK trellis to a third observation interval as shown in Figure 3.2.

The choice of LTVs introduced in (3.6) may be extended to more than two symbol intervals, by adding $c_{vu}$ and so forth to the $a_{vu}$ and $b_{vu}$ already defined:

$$\begin{aligned} e^{\frac{2}{N_0}\lambda_{n+2}(S_{00})} &= c_{00} \\ e^{\frac{2}{N_0}\lambda_{n+2}(S_{01})} &= c_{01} \end{aligned} \tag{3.9}$$

From Figure 3.2, the MSK likelihood ratio over three symbols is

$$\begin{aligned} \Lambda &= \frac{a_{00}\left[b_{00}(c_{00}+c_{01}) + b_{01}\left(\frac{1}{c_{00}} + \frac{1}{c_{01}}\right)\right]}{a_{01}\left[\frac{1}{b_{00}}\left(\frac{1}{c_{00}} + \frac{1}{c_{01}}\right) + \frac{1}{b_{01}}(c_{00}+c_{01})\right]} \\ &= \frac{a_{00}\left[b_{00} + \frac{b_{01}}{c_{00}c_{01}}\right]}{a_{01}\left[\frac{1}{b_{00}c_{00}c_{01}} + \frac{1}{b_{01}}\right]} \\ &= \frac{a_{00}b_{00}b_{01}}{a_{01}} \end{aligned} \tag{3.10}$$

which gives the same result as before, and is independent of the signal in the third symbol interval. Further extending the observation period $N > 3$ yields the same result: there is nothing to be gained in observing the MSK signal for more than two symbol intervals, by which time a maximum likelihood decision can be made.

Rimoldi [65] proves the same result and obtains the Massey receiver by couching the problem in terms of the Viterbi decoder: if the survivors at depth $n + 1$ of the trellis go through the same state at depth $n$, then the survivors at depth $n + 2$ go through the same state at depth $n + 1$. The proof is almost identical to Massey's proof [48], which I summarized in section 1.5.2 above.

## 3.4 Generalizing the MSK Analysis

With the example of MSK now well understood, it is possible to generalize some of the ideas used there to CPFSK and the rest of CPM. This section defines the concepts that will be needed in the next chapter where the search for a simplified receiver is described.

### 3.4.1 Arbitrary starting phase states for full-response CPFSK

Like the special case of MSK, the receiver for $M$-ary full-response CPFSK without coding is also invariant over starting phase state [3]. The starting phase state is the current phase state, assumed known, from which the likelihood ratio is written in order to detect the current symbol; or for differential CPFSK, the next phase state. The likelihood ratios of CPM generally do not simplify, and are therefore not identical when calculated from different starting states. The proof cannot rely on likelihood ratio simplification as in the case of MSK.

That the $M$-ary CPFSK receiver is invariant over starting phase state can be seen by reasoning that a different starting phase state $v_0 + k$, $k$ an integer, corresponds to a different starting phase $\varphi_0 = 2\pi hk$, as long as the trellis is invariant over phase state rotations. Since coherent detection is assumed, the starting phase $\varphi_0$ is assumed known, and therefore the detector is indifferent to the starting phase state $v_0$.

More formally, the above argument is valid if we can show in any symbol interval that

$$s_{(v+k),u}(\tau, \varphi_0) = s_{vu}(\tau, \varphi_0 + 2\pi hk), \quad 0 \le \tau < T, \;\; k \in \mathbb{Z} \tag{3.11}$$

where $\mathbb{Z}$ is the set of integers and the notation is abused somewhat to include the starting phase $\varphi_0$ in the argument list of $s_v u()$. For full-response CPFSK, the phase equation (1.23) corresponding to $s_{(v+k),u}(\tau, \varphi_0)$ is

$$\begin{aligned} \psi_{(v+k),u}(\tau) &= 2\pi h\left[(v_n + k) + u_n \tau / T\right] \\ &= \psi_{vu}(\tau) + 2\pi hk, \quad 0 \le \tau < T \end{aligned} \tag{3.12}$$

where $\psi_{vu}(\tau)$ is defined similarly to $s_{vu}(\tau)$ in (1.24). (3.11) is therefore true for all integer $k$.

The invariance over starting phase means that a coherent receiver cannot distinguish between the $P$ different transmitted waveforms that result from each of the $P$ initial and unspecified states of the CPE at time $t = 0$. In practical terms, the receiver's carrier synchronization circuit will lock on to one of $P$ phases of the received signal. If the receiver demodulates phase state estimates (like the MSK maximum likelihood receiver above), then the demodulated data too will have a $P$-fold ambiguity. In practice, this difficulty is overcome in one of two ways [20]:

1. Each transmitted message contains some known data, in a message header for example, so that the ambiguity can be resolved by subtracting $(\bmod P)$ the measured phase state difference from the detected phase states; or

2. Non-differential conventional CPM is transmitted and differential decoding is performed on the detected phase states, which removes the ambiguity at the cost of slightly degraded error performance.

### 3.4.2 The significance of starting phase states in CPM with coding

This subsection offers some speculation on the kinds of receiver structure for CPM with coding that we may look for. If a likelihood ratio is found to simplify, what will be the effect of an unknown starting phase on the receiver structure and performance?

The blindness of the full-response CPM receiver without coding to the starting phase state is a decided advantage, when one considers the problems faced by receivers for CPM with coding. If the trellis is no longer regular,[1] detected states are no longer simply related to the phase states through a constant offset modulo $P$. The receiver has to have exact knowledge of the signal phase state at the start of every symbol interval, else it will make demodulation errors even in the absence of noise [14].

Two well-known papers by Wei [86, 87] propose design rules for making trellis-coded modulation (TCM) 'transparent' to phase rotations of the signal set in increments of $\pi/2$ radians. Because the CPE trellis has a direct mapping, via the memoryless modulator (MM), to the signal constellation, the phase rotational invariance or otherwise of CPFSK with coding can be discussed using the same notation and design rules as those customarily used for rotationally invariant TCM.

From a likelihood ratio point of view, the reason for rotational variance is that a simplified receiver for CPM with coding generally will have a different set of likelihood functions from every starting state. The encoder imposes limitations on which transitions are possible from which states. The receiver has to be a compound receiver with a different structure 'switched in' depending on the current starting state. It will either need a 'magic genie' to tell it what each starting phase state is, so that it can use the correct subreceiver to detect the next symbol; or else the code will have to incorporate some form of differential encoding so that the receiver becomes insensitive to phase rotations of the received signal [86].

As an example, Figure 3.3 shows a partial trellis labelled with some LTVs. Two pairs of likelihood parameters $l_{vu}$ over three symbols can be written from the two starting states 0 and 1, and it is seen that their likelihood ratios $\Lambda_0$

[1]The meaning of the word *regular* will be defined on page 63.

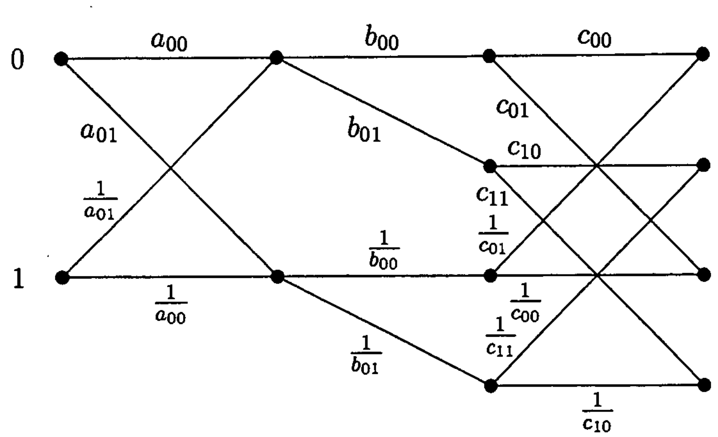



Figure 3.3: Example of a partial trellis which is invariant over the starting phase state. The branches are labelled with LTVs.

and $\Lambda_1$ are identical:

$$
\begin{aligned}
l_{00} &= a_{00}\left[b_{00}(c_{00}+c_{01})+b_{01}(c_{10}+c_{11})\right] \\
l_{01} &= a_{01}\left[\frac{1}{b_{00}}\left(\frac{1}{c_{00}}+\frac{1}{c_{01}}\right)+\frac{1}{b_{01}}\left(\frac{1}{c_{10}}+\frac{1}{c_{11}}\right)\right] \\
l_{10} &= \frac{1}{a_{00}}\left[\frac{1}{b_{00}}\left(\frac{1}{c_{00}}+\frac{1}{c_{01}}\right)+\frac{1}{b_{01}}\left(\frac{1}{c_{10}}+\frac{1}{c_{11}}\right)\right] \\
l_{11} &= \frac{1}{a_{01}}\left[b_{00}(c_{00}+c_{01})+b_{01}(c_{10}+c_{11})\right] \\
\Lambda_0 &= \frac{l_{00}}{l_{01}} = \frac{l_{11}}{l_{10}} = \Lambda_1
\end{aligned}
\tag{3.13}
$$

This property means that only one receiver structure is needed, and it will make a maximum likelihood decision irrespective of the starting state. If the likelihood ratios themselves simplify (which was not the case in our example above), then a simplified receiver results. If the ratios do not simplify, the general likelihood parameter receiver is the best we can do without using a Viterbi or sequential decoder, but the receiver structure itself will be invariant over rotations of the starting phase state. This likelihood transform approach is an alternative way of examining the problem of rotational invariance, and it could bear further investigation.

If the example above is changed slightly, by exchanging the two signals starting from starting state 1 for example, the likelihood functions from state 1 become

$$
\begin{aligned}
l_{10} &= \frac{1}{a_{01}}\left[\frac{1}{b_{00}}\left(\frac{1}{c_{00}}+\frac{1}{c_{01}}\right)+\frac{1}{b_{01}}\left(\frac{1}{c_{10}}+\frac{1}{c_{11}}\right)\right] \\
l_{11} &= \frac{1}{a_{00}}\left[b_{00}(c_{00}+c_{01})+b_{01}(c_{10}+c_{11})\right] \\
\Lambda_1 &= \frac{l_{11}}{l_{10}} \neq \Lambda_0 \qquad (3.14)
\end{aligned}
$$

and the likelihood ratios are no longer the same. If each likelihood ratio on its own simplifies, one may construct a compound receiver with each of the simplified likelihood ratios defining a subreceiver.

One way of deciding which subreceiver to use is to reason that, in the absence of a 'magic genie', the best estimate the receiver has of the starting phase state is its own previous maximum likelihood estimate. (We are assuming as before that the receiver detects phase states rather than frequency values.) This leads naturally to a decision-feedback receiver, which brings its own problems. Decision feedback always has the potential for losing the 'track' of the signal and generating long error bursts, sometimes catastrophically.

Differential CPFSK schemes without coding (and some CPFSK schemes with coding and with some regularity in the encoder) have trellises that are invariant over time shifts and invariant over phase state rotations. I shall call such trellises *regular* trellises for reasons of conciseness. Regular trellises are associated with linear time-invariant encoders.

**Definition 4** *A* regular *trellis is one which has the same likelihood ratio from every state within each symbol interval and which is invariant over time shifts of multiples of the symbol period.*

For a scheme with a regular trellis, it suffices to write a single likelihood ratio from any starting phase state. Likelihood ratios written from different states are identical, as I have shown in the case of MSK.

Encoders considered in this dissertation will generally not be restricted to the class of time-invariant linear sequential circuits. A time-varying channel encoder which cycles its encoding rule over a cycle length of $J$ may however be regarded as a time-invariant block encoder with $J$ inputs, which provides the modulator with $J/R_c$ symbols at a time, where $R_c$ is the code rate. The block encoder may still be nonlinear.

### 3.4.3 Extending to multi-level signalling

MSK is a binary modulation, and it was natural to write a decision rule for its optimal receiver in terms of a likelihood ratio. When dealing with an arbitrary $M$-ary modulation, it is not as clear how the decision rule should be formulated.

The likelihood parameter receiver of (1.54) evaluated $M$ likelihood functions on an observation, and chose the symbol corresponding to the largest parameter as the most likely. The simplified receiver came about through the simplification of the ratio of two likelihood functions.

The natural extension of these ideas is to require that all $M$ of the likelihood functions of an $M$-ary scheme should have common factors, and that these common factors include all the sums of terms in the functions. In this way, a ratio of any two of the likelihood functions will result in a simplified likelihood ratio. The logarithm of each likelihood function is taken, and the common factor discarded, to determine the maximum log-likelihood decision rule: choose the symbol corresponding to the largest simplified-likelihood parameter. Of course the calculation of the likelihood parameter should now entail nothing more than adding or subtracting a few matched filter or correlator output samples.

The immediate difficulty with this method is that $M$-ary schemes generally have more than two states, and in the presence of coding the sets of simplified likelihood functions from each state are not necessarily the same. One may have to implement a compound receiver as mentioned before, which hopefully would still be less complex than a full Viterbi decoder and have reasonable error performance.

MSK is unique not only in that the likelihood functions from each state simplify in their likelihood ratio, but also that the two likelihood ratios from the two states are identical.

### 3.4.4 The choice of observation period

In the process of looking for simplified receivers, we need to know over what interval to write the likelihood functions of each trellis that we examine. Choosing too short an interval is sure to lead to a suboptimal receiver because not all the information in the signal sequence is used, and choosing too long an interval dramatically multiplies the number of trellises to be examined without further benefit.

It was shown in section 3.3.2 that extending the observation period for MSK beyond two symbols does not give the receiver any useful additional information. The same kind of proof is possible for other schemes with simplified receivers. For regular schemes such as full-response CPM, optimal observation periods can be deduced directly from the trellises.

The observation period should be exactly long enough to observe what happens to two or more paths through the trellis, from the moment that they diverge from a common state, until they merge again at the first opportunity in a common state. This interval is sometimes called the free distance between two paths.

In the coding theory for feedforward encoders, the concept of constraint length is defined as the number of current and future output symbols that are affected by the current input symbol [44, p.11].[2]

If the channel encoder (CE) and CPE are combined in a common trellis representation, as will be done in the next chapter, the idea of a constraint length is seen to be closely related to the diverging and merging of paths through the trellis. Having gone different ways from a certain state, the longest interval that any two paths can keep apart before merging at the first opportunity, is the constraint length of the scheme. This is also the observation period that an optimal receiver will have to use.

[2]For encoders with $n_o$ outputs and memory order $m_o$, the constraint length is actually defined as $n_A = (m_o + 1)n_o$, which takes into account the number of outputs in assuming that they are to be interleaved. In the binary single-input rate-1/2 encoders considered in the next chapter, I recombine the two outputs into a single value modulo 4, and therefore prefer to use the term constraint length more loosely to refer simply to the number of output symbols affected.

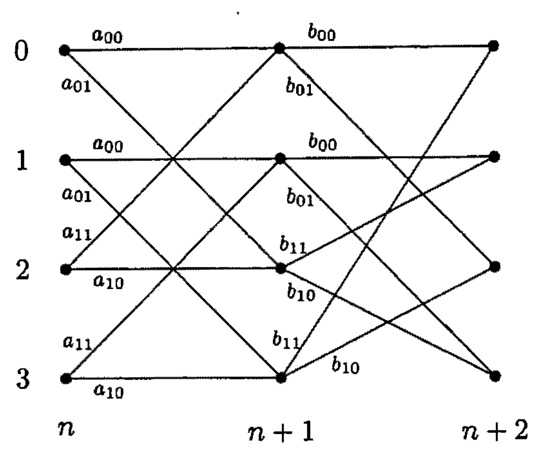



Figure 3.4: Example of a catastrophic trellis.

Depending on the coding scheme, not all pairs of paths necessarily achieve the constraint length of the scheme. It is possible that diverging paths from some states may merge sooner than those from other states. The free Euclidean distance of any scheme, and hence its error performance, is often determined by the shortest free distance in the trellis. It follows that, in order to optimize error performance for a given observation period, the shortest free distance should be the same as the constraint length of the scheme.

## 3.5 Catastrophic Trellises

If a trellis has simplifying likelihood ratios from every state, it may still not represent a practical modulation scheme. The time-varying trellis in Figure 3.4 has likelihood ratios which simplify from every state and a free distance of three symbol intervals. Yet something is suspicious when one examines its actual likelihood ratio from, for example, the first state:

$$\begin{aligned} \Lambda_{00} &= \frac{a_{00}\left[b_{00}\left(c_{00}+c_{01}\right)+b_{01}\left(c_{10}+c_{11}\right)\right]}{a_{01}\left[b_{10}\left(c_{10}+c_{11}\right)+b_{11}\left(c_{00}+c_{01}\right)\right]} \\ &= \frac{a_{00}b_{00}b_{01}}{a_{01}} \end{aligned} \tag{3.15}$$

All the $c$ terms from the third interval have cancelled, leaving a receiver which pretends to be optimal while observing the received signal for only two symbol intervals.

Thinking back to (3.7), this is the likelihood ratio of MSK. Furthermore, the pairs of likelihood ratios from pairs of adjacent states, that is, $\Lambda_{00}$ and $\Lambda_{01}$, $\Lambda_{20}$ and $\Lambda_{21}$, and so on, are identical. It is not that the trellis comprises two disjoint sub-trellises, but rather that it is possible to follow two separate but parallel trajectories through the trellis indefinitely without accumulating any incremental Euclidean distance between them.

This phenomenon has certainly been observed before in some TCM schemes [13]. Rimoldi and Li [68] define a catastrophic CPM scheme as one in which there exists a pair of input sequences, say $\boldsymbol{u}$ and $\boldsymbol{u}'$, which differ in an infinite number of places such that the corresponding transmitted signals $s(t, \boldsymbol{u})$ and $s(t, \boldsymbol{u}')$ have finite Euclidean distance between them. This is a direct extension of the classical definition of catastrophic codes [50][85, p.250][44, p.308]: a code is catastrophic if and only if the state transition diagram contains a loop of zero weight, other than the self-loop around the all-zero state.

The definition of a catastrophic CPM scheme above can be extended directly to define a catastrophic coded modulation scheme. In this context, a modulation is generally regarded as being 'coded' if the transmitter possesses memory, and the contents of this memory defines the transmitter's state at any given time:

**Definition 5** *A coded modulation scheme is said to be* catastrophic *if there exists a pair of transmitter state sequences which differ in an infinite number of places while the corresponding transmitted signals differ by a finite Euclidean distance.*

The definition is given in terms of transmitter states rather than input symbols, because the states are what a maximum likelihood receiver detects. The mapping from input symbol to next transmitter state may be left unspecified, as long as it is assumed to be a one-to-one mapping.

It follows from Figure 3.4 that states 0 and 1 are associated with the same point in signal space, as are states 2 and 3, being differentiated by the CE state only. Each pair of states may for example form a phase state of a CPM scheme. In effect one looks for pairs of parallel but non-touching paths which wrap around the trellis indefinitely with zero Euclidean distance between them.

I call such pairs *zero-distance loops.*

This situation is found from every pair of states in the trellis of Figure 3.4. There may be a CE, and it may be in various states at various times, but the receiver is given no information as to what those states are. It might as well be MSK without coding, which is why an MSK receiver resulted from (3.15).

Simplification of the likelihood ratio, aided by the likelihood transform, is a useful tool with which to check the validity of estimations of free distance: if not all the intervals within the free distance contribute LTVs to the simplified likelihood ratio, the scheme is catastrophic or otherwise degenerate.

# Chapter 4

# Searches for Simplified Receivers

This chapter covers the methods and results of two explorations of the search space of possible simplified receivers. These searches are in principle not any different from the code searches commonly conducted to find good modulation codes for additive white Gaussian noise channels. The chief difference is in the stated goal of the search: In trellis coded modulation the goal is generally to maximize the free Euclidean distance subject to practical constraints on the number of code states. Here, the main goal is to find a practical simplified receiver, with maximization of the free Euclidean distance as a secondary goal.

The main result of the search is that no simplified receiver has been found which is not in some way based on minimum shift keying's (MSK's) receiver. This is not a particularly surprising result, since there is no mention of a simplified receiver apart from Massey's in the published literature. However, some of the results described in this chapter lead to interesting formulations of coded modulations related to MSK, and pointed the way for further investigation. These are discussed more fully in Chapter refch5.

One obvious way of ensuring that a likelihood ratio simplifies into a single term, is to start with a single term and attempt to elaborate it into a sensible likelihood ratio. This reverse process was not systematically pursued, but section 4.1 gives an example of the obstacles such a technique would have to overcome. This method is taken up again in the next chapter when the results of the various searches are discussed.

The first search method generated various likelihood ratios which describe useful modulation schemes, and then checked whether they simplify. If a likelihood ratio had been found to simplify, the simplified receiver would have followed directly. The task was to choose a specific coded modulation scheme, or rather a class of coded modulation schemes with the channel encoder (CE) left unspecified. This narrowed the search sufficiently for reasonably complex schemes to be considered. Section 4.2 therefore considers a four-level continuous-phase frequency shift keying (CPFSK) with $h = 1/4$ and a scheme called quadrature minimum shift keying (QMSK).

The second search method produced mathematical expressions which may be interpreted as likelihood ratios or parts of likelihood ratios, and tried to simplify these. The explicit goal was to find mathematical identities which may aid in simplifying likelihood ratios. Section 4.3 describes an attempt at finding identities similar to the MSK identity, and a wider search based on generating all possible likelihood ratios for a given input alphabet size $M$ and observation period $N$. The number of possible likelihood ratios becomes very large for moderately large $M$ and $N$, and only the simpler cases were tractable. A number of identities were found which all bear some similarity to the MSK identity.

For the sake of convenience, all expressions for likelihood ratios and likelihood functions will be written in terms of their likelihood transforms.

## 4.1 Constructing Receivers From Simplified Likelihood Ratios

The likelihood ratio for MSK simplified (see (3.7)) because of the algebraic identity

$$\frac{a+b}{\frac{1}{a}+\frac{1}{b}} = ab, \quad a, b > 0 \tag{4.1}$$

I call it the *MSK identity*.

This identity was the only tool initially available for simplifying likelihood ratios of schemes other than MSK. This section illustrates the construction of a maximum likelihood receiver directly from a pair of likelihood functions whose quotient is known to simplify. The nature of the corresponding modulation scheme, continuous-phase or not, may then be derived from the receiver.

MSK can be detected optimally over two symbol intervals. As an attempt to do something similar over three intervals, we work backwards from the likelihood transform of a general simplified likelihood ratio, without specifying the signals of the scheme:

$$\begin{aligned}\Lambda &= a_0 a_1 b_0 b_1 c_0 c_1 \\ &= \frac{a_0(b_0+b_1)(c_0+c_1)}{\frac{1}{a_1}\left(\frac{1}{b_0}+\frac{1}{b_1}\right)\left(\frac{1}{c_0}+\frac{1}{c_1}\right)} \\ &= \frac{a_0\left[b_0(c_0+c_1)+b_1(c_0+c_1)\right]}{a_2\left[b_2(c_2+c_3)+b_3(c_2+c_3)\right]}\end{aligned} \tag{4.2}$$

where we have substituted

$$\begin{array}{lll} a_2 = \frac{1}{a_1} & b_2 = \frac{1}{b_0} & c_2 = \frac{1}{c_0} \\ & b_3 = \frac{1}{b_1} & c_3 = \frac{1}{c_1} \end{array} \tag{4.3}$$

(4.2) is in the form of a ratio of two likelihood functions, each one defining a tree of possible trajectories from a common state. The corresponding maximum likelihood receiver is shown in Figure 4.1. The signals have not been identified and separate correlators are shown for each one, but one can see that correlations may be combined if the same signal occurs in more than one interval. Compare this diagram with the MSK detector in Figure 1.9.

Now the question is how to define the signals corresponding to each variable in (4.2). Figure 4.2 shows one possible trellis that fits the description, where the two sub-trees corresponding to the two likelihood functions have been merged in the final interval.

A difficulty with this approach now becomes apparent: The signals corresponding to $c_0$ and $c_1$ are duplicated in the last interval, and $c_2$ and $c_3$ also. This duplication seems to be an inevitable result of starting with a simplifying

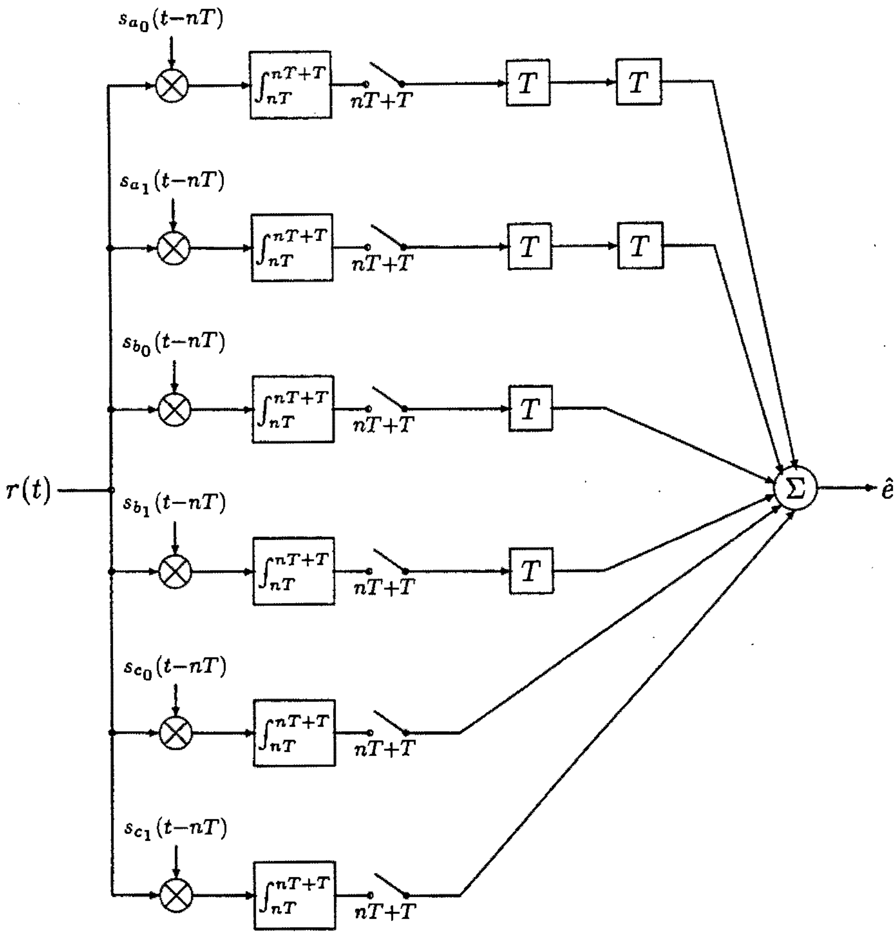



Figure 4.1: A general maximum likelihood receiver derived from a simplified likelihood ratio over three symbol intervals. The correlations may be combined if the same signal occurs in more than one interval.

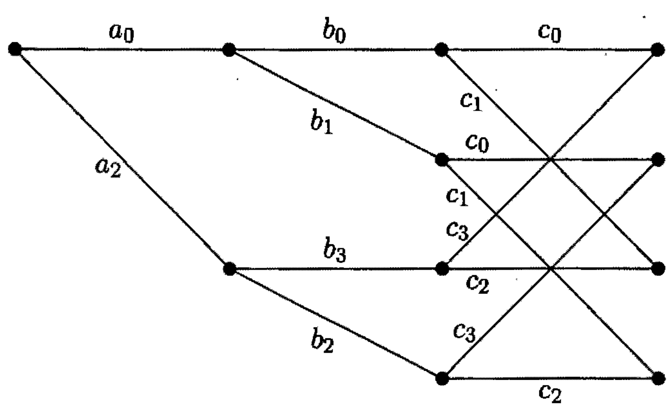



Figure 4.2: Trellis derived from two likelihood functions, each of which defines a sub-tree from a common node.

likelihood ratio, also for longer observation periods.

We distinguish between two cases: one arises if we insist that the modulation is a continuous phase modulation (CPM) without coding, and the other if we allow coding. In the case of CPM without coding, each state in this trellis is a phase state. It then becomes impossible to reconcile the trellis, or indeed any similar trellis constructed from (4.2), with the continuous-phase requirement. Signals such as $c_0$ which start in different phases cannot have the same name or, put differently, identical signals cannot start in different phase states.

One way of resolving the difficulty is by making parallel branches, as shown in Figure 4.3. The duplicated signals can now be merged, but the parallel transitions in the second interval can have a normalized squared Euclidean distance (NSED) between them of at most 1 for CPFSK and at best approaching 2 for any CPM. This in turn limits the free Euclidean distance of the scheme to those values. Furthermore, the parallel transitions require pairs of signals which are separated in frequency by multiples of the symbol rate, so that they may start and end in the same phase. For all of the schemes considered in this dissertation, and most practical schemes in the literature, such a requirement means expanding the signal set at the expense of bandwidth. The use of parallel transitions to resolve the signals produced from a simplifying likelihood ratio is not attractive.

In the second case we allow coding, which has the effect of introducing

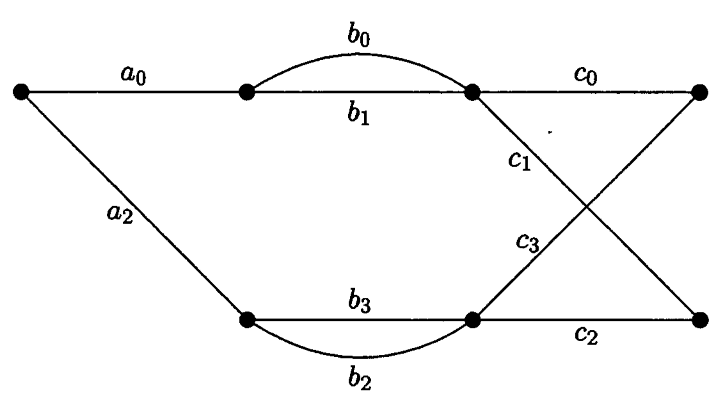



Figure 4.3: Making CPM with parallel branches, which requires at least one pair of signals separated in frequency by the symbol rate.

code states into the trellis which are independent of the phase states, assuming that we combine code and phase states in one trellis. Now the duplicated signals in Figure 4.2 may be thought of as being physically identical (having the same starting phases and frequencies), but being distinguished by different code states. Tracing back to the second interval, the pairs of signals corresponding to $b_0$ and $b_1$ (and $b_2$ and $b_3$) again have to be either physically identical, or else differ in frequency by multiples of the symbol rate, because they end in the same phase.

Taking first the case where the signals in the second interval are physically identical, the best free Euclidean distance that can be achieved in this case is with an MSK with coding (assuming that the modulation index is limited to $h < 1$). It is only nominally coded because the encoder turns out in every case to be a simple delay, with no coding gain. The free Euclidean distance is limited to $d^2_{\text{free}} = 2$, the same as MSK without coding. (In constructing candidate trellises, care should be taken to avoid catastrophic trellises, as defined in section 3.5.)

If the pairs of signals in the second interval are allowed to accumulate multiples of $2\pi$ radians between them over one symbol interval, an interesting situation is found. Such pairs of signals are orthogonal (assuming a sufficiently high carrier frequency) and can therefore contribute a NSED of 1 to the free squared Euclidean distance. With a constraint length of three symbol intervals, a free Euclidean distance of $d^2_{\text{free}} = 3$ ought to be achievable, and correspondingly more with longer constraint lengths.

The obvious candidate would be a coded form of Sunde's FSK (1REC with $h = 1$) [75], but that limits the trellis to a single phase state. Anderson *et al* have also shown that $h = 1$ is a particularly weak modulation index in terms of free Euclidean distance, especially when combined with coding [3]. It is possible to construct a scheme with a modulation index like $h = 1/2$ and with either $M = 3$ or $M = 4$ signals in its alphabet. This generates signals spaced in frequency by half the symbol rate, including at least one pair which differs in frequency by the symbol rate. This scheme can no longer be treated as a differential CPM, because $M > P$, but that should not introduce any special difficulty. It is known that such schemes have some coding gain, albeit at the expense of bandwidth efficiency [3].

The *ad hoc* investigation above is followed up in Chapter 6 with a much more structured analysis.

## 4.2 Search for Simplified Receivers for Specific Modulations

The most general possible search would generate all possible likelihood ratios and attempt to simplify them, with the aim of assigning a simplified likelihood ratio to an equal-energy modulation scheme once it has been found. The enormous number of possible likelihood ratios makes it impossible to do an exhaustive search for simplifications.

A more focused approach was therefore tried: a specific coded modulation scheme was chosen with the CE as yet unspecified. The search now became a search of possible modulation codes, by generating the likelihood ratio of each CE combined with the modulator, and attempting simplification. In this way, it was hoped to extend the search to longer observation periods and more states.

This section describes a search for a simplified receiver for QMSK, which is a four-level CPFSK with coding and with modulation index $h = 1/4$. Before describing this search, an example is given of a proof of the non-existence of simplified receivers for CPM schemes without coding.

### 4.2.1 Four-level CPFSK with $h = 1/4$

One possible place to start looking for a CPFSK receiver with a simplifying likelihood ratio, is four-level full-response CPFSK (1REC) with modulation index $h = 1/4$. I shall simply call it 4-CPFSK hereafter. It is a bandwidth-efficient scheme, with much potential for excellent error performance when combined with a convolutional encoder to increase its free Euclidean distance [45][3, p.438]. A simplified receiver for it would obviously be of interest.

Using the simple likelihood transform notation, it is straightforward to show that a simplified receiver does not exist for 4-CPFSK. Its trellis was illustrated in Figure 1.5, and is repeated in Figure 4.4 for easy reference. From the arbitrary

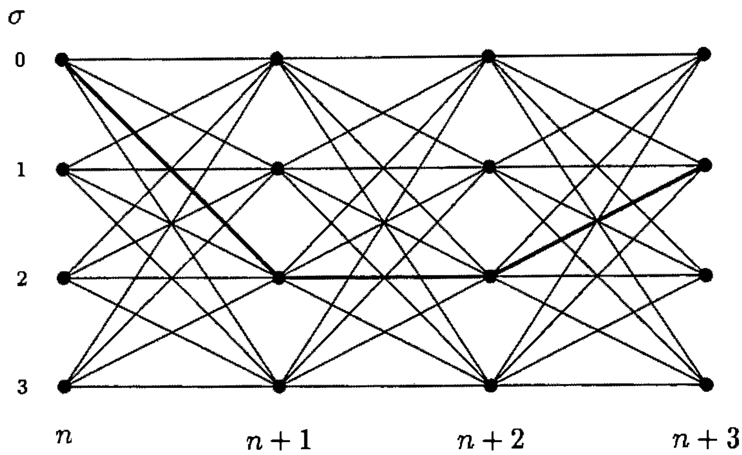



Figure 4.4: Trellis diagram for four-level full-response CPFSK with $h = 1/4$. The signal trajectory for the data input sequence $(2, 0, 3)$ is shown in a heavy line, assuming $\sigma_n = 0$ to start with.

starting state $\sigma_n = 0$, the four likelihood functions over two symbol intervals are

$$\begin{aligned} l_{00} &= a_{00}(b_{00} + b_{01} + b_{02} + b_{03}) \\ l_{01} &= a_{01}(b_{10} + b_{11} + b_{12} + b_{13}) \\ l_{02} &= a_{02}\left(\frac{1}{b_{00}} + \frac{1}{b_{01}} + \frac{1}{b_{02}} + \frac{1}{b_{03}}\right) \\ l_{03} &= a_{03}\left(\frac{1}{b_{10}} + \frac{1}{b_{11}} + \frac{1}{b_{12}} + \frac{1}{b_{13}}\right) \end{aligned} \tag{4.4}$$

because

$$\begin{aligned} b_{20} &= \tfrac{1}{b_{00}} & b_{30} &= \tfrac{1}{b_{10}} \\ b_{21} &= \tfrac{1}{b_{01}} & b_{31} &= \tfrac{1}{b_{11}} \\ b_{22} &= \tfrac{1}{b_{02}} & b_{32} &= \tfrac{1}{b_{12}} \\ b_{23} &= \tfrac{1}{b_{03}} & b_{33} &= \tfrac{1}{b_{13}} \end{aligned} \tag{4.5}$$

The likelihood transform variables (LTVs) here are named according to the convention previously used in (3.6). Taken in pairs and combined into ratios, no likelihood ratio simplification occurs among the likelihood functions in (4.4). Extending the observation period to three or four symbol intervals simply makes it worse, the likelihood functions have no common factors.

It is also possible, but more tedious, to prove the same result by following the 'magic genie' [90, p.419] or Viterbi decoder argument of Massey [48] or Rimoldi [65]: the various decision functions obtained depend on the starting and ending phase.[1]

One may conjecture that the regular trellis associated with 4-CPFSK is the reason why it does not have a simplified receiver. The trellis is fully connected, meaning that every phase state can be reached in one symbol interval from any other state. This is a result of having $M = P$. If the lack of a simplified receiver for 4-CPFSK is representative of what may be expected of similar schemes, the next step is to look at schemes with trellises which are not fully connected. If we choose $M < P$, for example binary CPFSK with $h = 1/4$, we are given some freedom regarding which states to connect to which in the trellis. It is then conceivable that a trellis may be constructed which might have likelihood functions which simplify in pairs.

### 4.2.2 Quadrature Minimum Shift Keying

The Costas loop synchronization circuits of the Massey-Hodgart receiver [35], mentioned in section 1.5.2 and illustrated in Figure 1.11, are closely integrated with the data detection circuit, resulting in a simple and robust receiver. It is tempting to try to use Costas loops for other full-response CPFSK schemes as well. The effect is to place a constraint on the design of the trellis to facilitate easy synchronization.

Hodgart [34] has proposed a family of four-level CPFSK schemes with coding and with $h = 1/4$, calling it QMSK. The choice of four-level modulation with coding and $h = 1/4$ was inspired by the good performance reported for this class of modulations when combined with convolutional encoders [45][3, p.438]. With proper encoder choice, QMSK should have some coding gain over MSK, while requiring about the same bandwidth. No receiver has been proposed for QMSK, but the proposal for the signal design seems to lend itself to easy synchronization in a possible practical receiver.

The QMSK definition does not specify a receiver, because apart from the

[1] This result also contradicts the finding of Kritzinger [40], who claims to have found a simplified 4-CPFSK receiver by Massey and Rimoldi's methods.

required continuous phase and the general Costas constraint placed on the trellis design, the encoder is left unspecified. In particular, Hodgart stresses that one does not have to limit oneself to linear time-invariant convolutional encoders, as most researchers have done [3, 6].

### QMSK description

The QMSK signal is defined as $M = 4$ CPFSK with $h = 1/4$, fed by a rate-$1/2$ CE. The input to the CE is therefore binary. The transmitted signal, which has one of four frequencies during any symbol interval, is constrained always to be either in phase or exactly anti-phase to any one of a set of four continuous reference waveforms. The reference waveforms are sine waves, one at each signalling frequency, and phased so that they are all exactly in phase once every four symbol intervals at a symbol boundary. These are the four signal frequencies that have to be recovered at the receiver, compared to the two in the Massey-Hodgart MSK receiver. The aim is to have an 'all in-phase' structure in the receiver, similar to the Massey-Hodgart MSK receiver (see Figure 1.11), where the quadrature arms of the Costas loops are dedicated to synchronization only, and the in-phase arms detect the signal.

If one should drop this phasing requirement and use all eight oscillators (four frequencies, two quadrature oscillators at each frequency) to detect the signal, this is sure to complicate both the detection and the synchronization.

Figure 4.5 shows the four reference waveforms near baseband, as well as a possible QMSK signal which satisfies the QMSK constraints. Note that although the reference waveforms are expected to coincide in phase once every four symbol intervals on a symbol boundary, the actual phase at that moment is an arbitrary $\varphi_0$. This phase coincidence results from the value of the modulation index $h = 1/4$, and the requirement that the QMSK signal should at every symbol boundary be in a position to switch frequencies phase-continuously and follow a new phase trajectory either in phase or anti-phase to one of the reference oscillators.

QMSK differs from 4-CPFSK without coding in that the input to the CE is binary, not quaternary, in that it has a rate-$1/2$ CE preceding the modulator

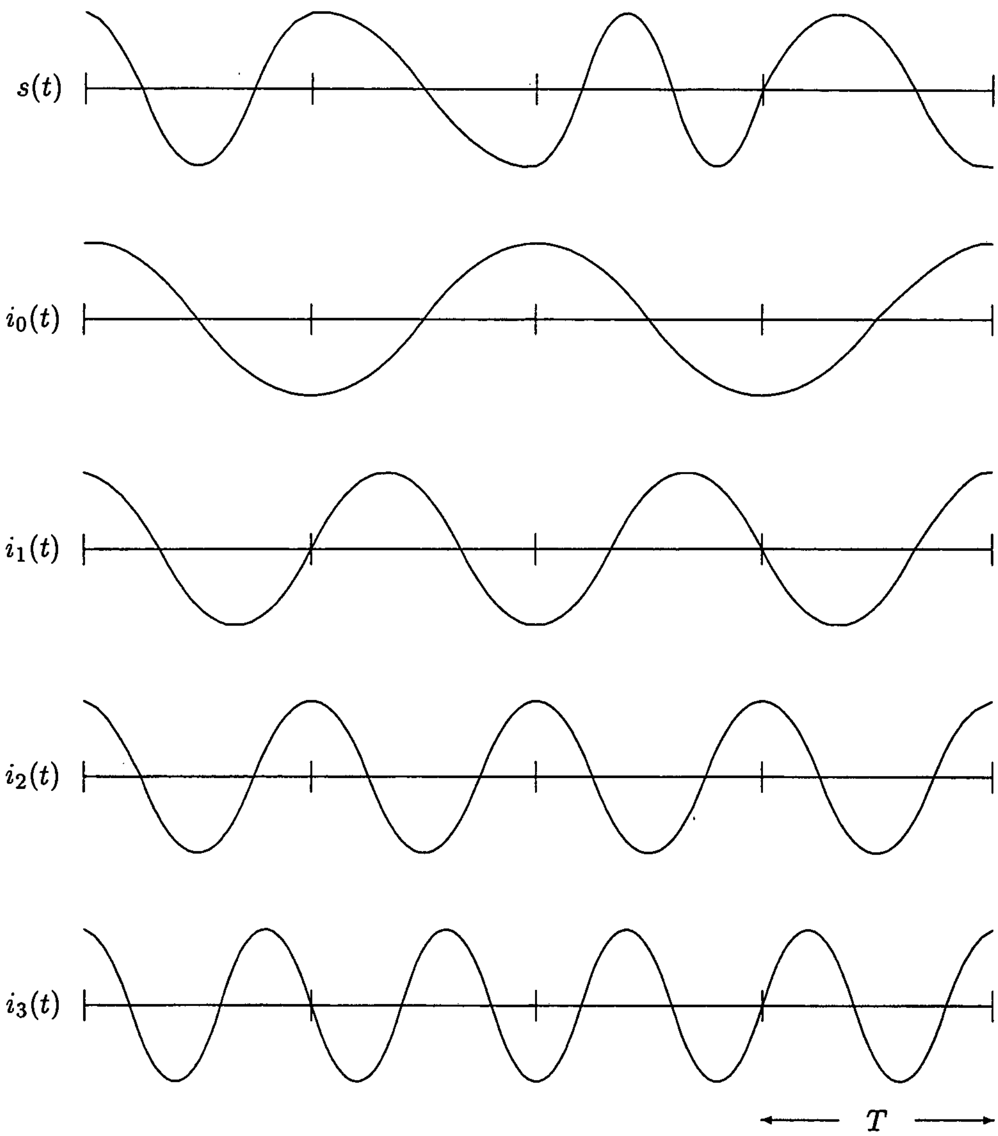



Figure 4.5: QMSK modulator waveforms over four symbol intervals, showing a valid signal (top) and four reference oscillations, one at each signalling frequency.

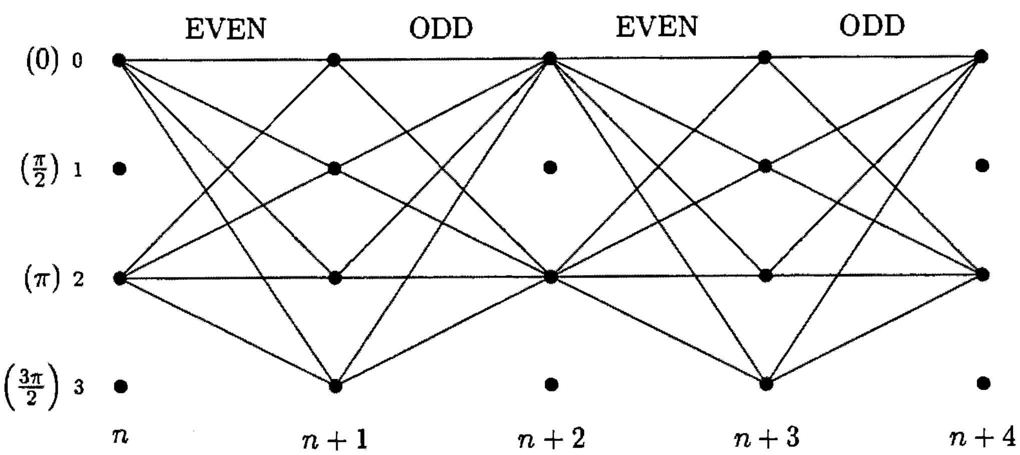



Figure 4.6: QMSK phase trellis, showing the phases of the four reference oscillators at the symbol boundaries, or alternatively, showing the combined CE/CPE code trellis. The symbol intervals are designated 'EVEN' and 'ODD' as shown. The CE ensures that the CPE/MM combination selects trajectories from this trellis only.

(to increase the number of levels from two to four), and that certain phase trajectories are excluded by the encoder from appearing in the transmitted signal, in order to meet the QMSK carrier synchronization constraint mentioned above. I am not aware of any encoders in the literature which have this last property.

The QMSK phase trellis is shown in Figure 4.6. Comparing this trellis with Figure 4.4, one can see immediately that the branches from some states are now binary, while others are quaternary, and that the trellis is time-varying, in the sense that even and odd symbol intervals do not have the same signal set. Looking at Figure 4.5, one can see that at every second symbol boundary, the signal has a choice of all four reference oscillators (or their inverses) to follow, while at the intervening boundaries it has a choice of only two, the other two oscillators being in phase quadrature at those moments.

One may see this trellis as either the phases of the reference oscillators at the symbol boundaries, or as the code trellis of the continuous-phase encoder (CPE), with its input data constrained by a CE. Note that there are unused phase states in the even intervals.

The input to a 4-CPFSK modulator without coding is an arbitrary four-level symbol sequence. It is the duty of the QMSK CE to ensure that no symbol

| EVEN | | | | ODD | | | |
|---|---|---|---|---|---|---|---|
| State $\zeta_n$ | Input $y_n$ | Output $u_n$ | Next $\zeta_{n+1}$ | State $\zeta_n$ | Input $y_n$ | Output $u_n$ | Next $\zeta_{n+1}$ |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 | 0 | 1 | 2 | 0 |
| 1 | 0 | 2 | 0 | 1 | 0 | 1 | 1 |
| 1 | 1 | 3 | 1 | 1 | 1 | 3 | 1 |

Table 4.1: Hodgart's two-state QMSK encoder A definition.

reaches the QMSK CPE and memoryless modulator (MM) which causes the signal to depart from the trellis of Figure 4.6. Simultaneously, it is hoped that the CE will provide a greater free Euclidean distance.

The time-varying nature of the QMSK trellis means that the CE too will be time-varying: it changes its encoding rule every symbol interval. Presumably one may specify that it has one rule for even symbol intervals and another for odd intervals, but in general its rules may rotate with a cycle of integers modulo $J$, where $J$ is an even positive integer.

Despite a superficial resemblance, QMSK is not the same as multi-$h$ modulation [52, 8], which switches the modulation index cyclically. In QMSK the modulation index is fixed.

### The Hodgart encoders

Hodgart [34] proposed two two-state CE definitions and a four-state definition. The two-state encoders are called 'A' and 'B', and are defined in Tables 4.1 and 4.2. The trellis associated with each encoder follows its table. The two-state encoders yield a free Euclidean distance of $d^2_{\text{free}} = 3$, or 1.8 dB improvement over MSK. The four-state encoder, which I have called 'C', is a slightly modified combination of the previous two, and has a free Euclidean distance of $d^2_{\text{free}} = 4$, or 3 dB gain over MSK. It is defined by its trellis diagram in Figure 4.9.

The CE output $u_n$ is fed into the CPE of a conventional 4-CPFSK, in other words, the $u_n$ refer to frequency values, not next phase states. The trellises in Figures 4.7 to 4.9 generally have the input data unspecified. It is not necessary for likelihood ratio simplification to specify whether the upper branch leaving

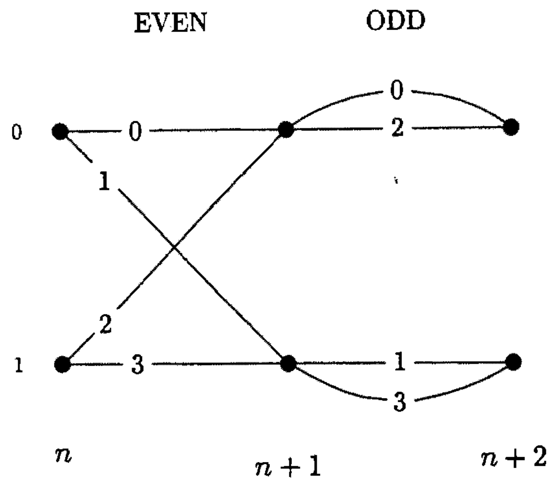



Figure 4.7: Trellis for Hodgart's two-state QMSK encoder A, with 1.8 dB gain over MSK. The branches are labelled with the CE output $u_n$.

| EVEN | | | | ODD | | | |
|---|---|---|---|---|---|---|---|
| State $\zeta_n$ | Input $y_n$ | Output $u_n$ | Next $\zeta_{n+1}$ | State $\zeta_n$ | Input $y_n$ | Output $u_n$ | Next $\zeta_{n+1}$ |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 3 | 1 | 0 | 1 | 2 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 |

Table 4.2: Hodgart's two-state QMSK encoder B definition.

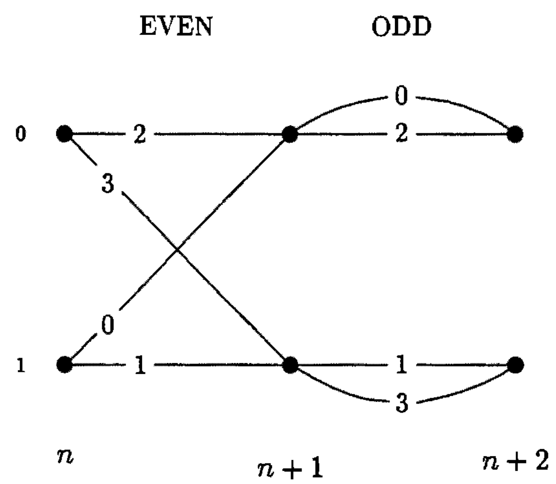



Figure 4.8: Trellis for Hodgart's two-state QMSK encoder B, with 1.8 dB gain over MSK. The branches are labelled with the CE output $u_n$.

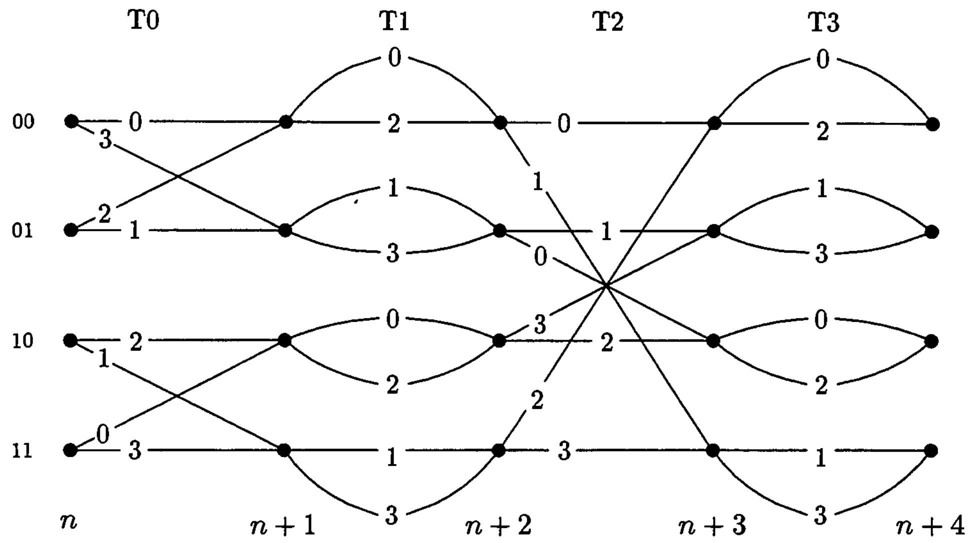



Figure 4.9: Trellis for Hodgart's four-state QMSK encoder C, with 3 dB gain over MSK. The branches are labelled with the CE output $u_n$.

a state represents input data 0 or 1, as long as one input level selects one valid signal trajectory and the other level the other trajectory, merely affecting as it does the data-to-signal mapping. The CE definition in QMSK may be regarded as a table look-up throughout.

I shall call the QMSK schemes that result from the use of encoder A, B and C, scheme A, B and C respectively. One may readily verify that the encoders meet the QMSK phase trellis constraint. The four-state encoder and trellis is time-varying with a cycle period of $J = 4$, as compared to two (even/odd) for the two-state encoders.

To formalize the CE notation slightly, it helps to conceptualize the CE as having one input and two outputs. It can then be described in the conventional way as a rate-1/2 encoder with binary inputs and outputs. The CE outputs are combined linearly through

$$u_n = \left[u_n^{(1)}, u_n^{(0)}\right] \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{4.6}$$

where $u_n^{(1)}$ is the most significant bit of the CE output and $u_n^{(0)}$ is the least significant bit, before being fed into the CPE as a number modulo 4. Figure 4.10

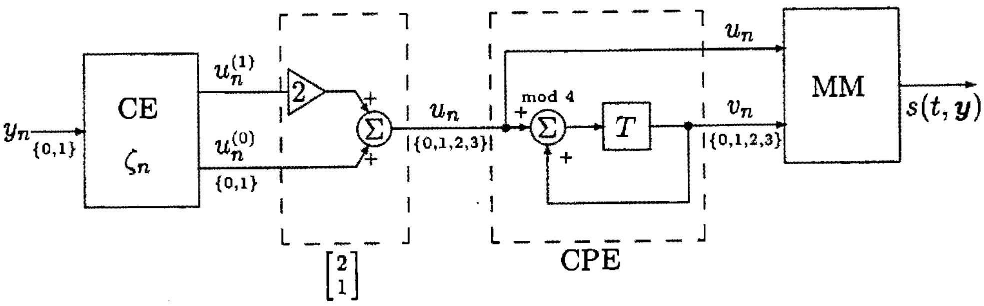



Figure 4.10: QMSK modulator block diagram, showing the cascade of CE, CPE and MM.

shows the block diagram.

It is interesting to note that the three different schemes have different power spectra, even though all signalling frequencies are used with equal probability in each one.

I illustrated the Hodgart QMSK encoders as an introduction to QMSK, and to show what performance may be possible with the scheme. No receiver has yet been designed for it, neither in terms of signal detection nor synchronization. My intention was to try to find a simplified receiver for it.

### 4.2.3 Likelihood ratio analysis of QMSK

The Hodgart encoders were not designed with simplified receivers in mind, and it can be shown that they do not have simplified receivers. To do so, I introduce a combined CE and CPE trellis diagram as shown in Figure 4.11. For a two-state encoder and four-state CPE, as we had for Hodgart's scheme A, the combined trellis has eight states. Each pair of adjacent states form a phase state, and each state within a pair is a CE state. Similarly, if the CE has four states, then the combined trellis states are grouped into groups of four CE states, and each group forms one phase state.

The combined trellis now has the useful characteristic of having binary branches (binary input to the CE) from every state except the unused ones. Previously, the QMSK CPE trellis had various numbers of branches emerging from and merging into each used state.

The trellis is *irregular* (in the sense of regular trellises defined in Chapter 3),

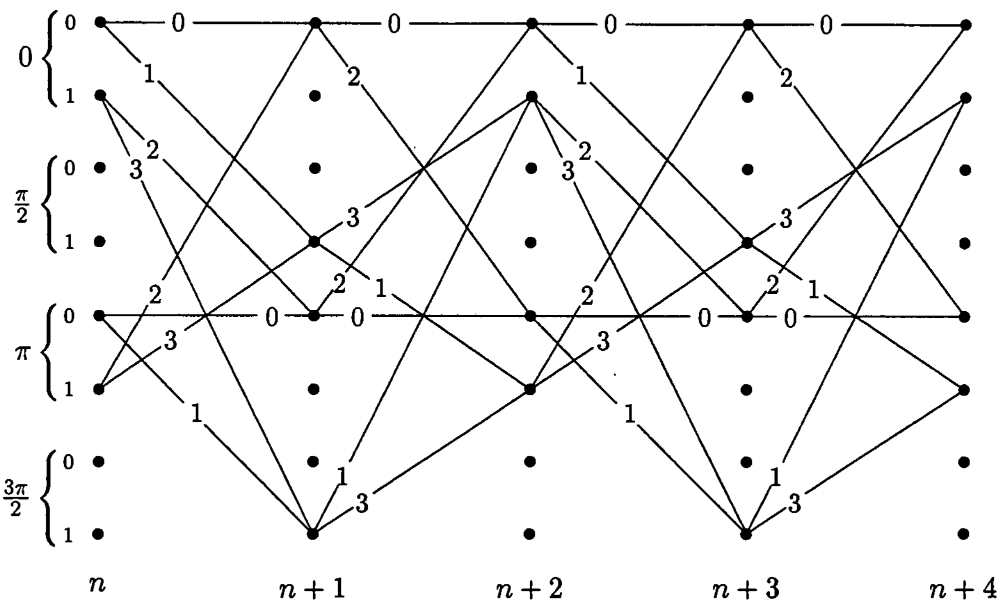



Figure 4.11: Combined CE and CPE trellis for Hodgart's QMSK scheme A, over four symbol intervals. The states are labelled with their CE state and the associated signal phase. The branches are labelled with the CE output or CPE input.

because it is not invariant over state rotations. The practical implication of this is that the likelihood ratio now has to be calculated from every state. If it should simplify every time, then a simplified receiver would exist from every state. These simplified receivers would generally not be the same from every state, however. This difficulty was discussed in more detail in section 3.4.2.

Referring to Figure 4.11, likelihood functions from each state are now labelled $l_{vzy}$, where the subscript $v$ refers to the CPE starting state $\sigma_n = v_n$, the $z$ refers to the CE starting state $\zeta_n$ and the $y$ refers to the input data in the first interval $y_n$. For an observation period of three symbol intervals, from phase state 0, encoder state 0, they are

$$
\begin{aligned}
l_{000} &= a_{00}\left[b_{00}(c_{00}+c_{01}) + b_{02}(c_{20}+c_{21})\right] \\
&= a_{00}\left[b_{00}(c_{00}+c_{01}) + b_{02}\left(\frac{1}{c_{00}}+\frac{1}{c_{01}}\right)\right], \qquad (4.7) \\
l_{001} &= a_{01}\left[b_{11}(c_{22}+c_{23}) + b_{13}(c_{02}+c_{03})\right] \\
&= a_{01}\left[b_{11}\left(\frac{1}{c_{02}}+\frac{1}{c_{03}}\right) + b_{13}(c_{02}+c_{03})\right] \qquad (4.8)
\end{aligned}
$$

where the LTV naming convention follows that of (3.6) as before. The likelihood

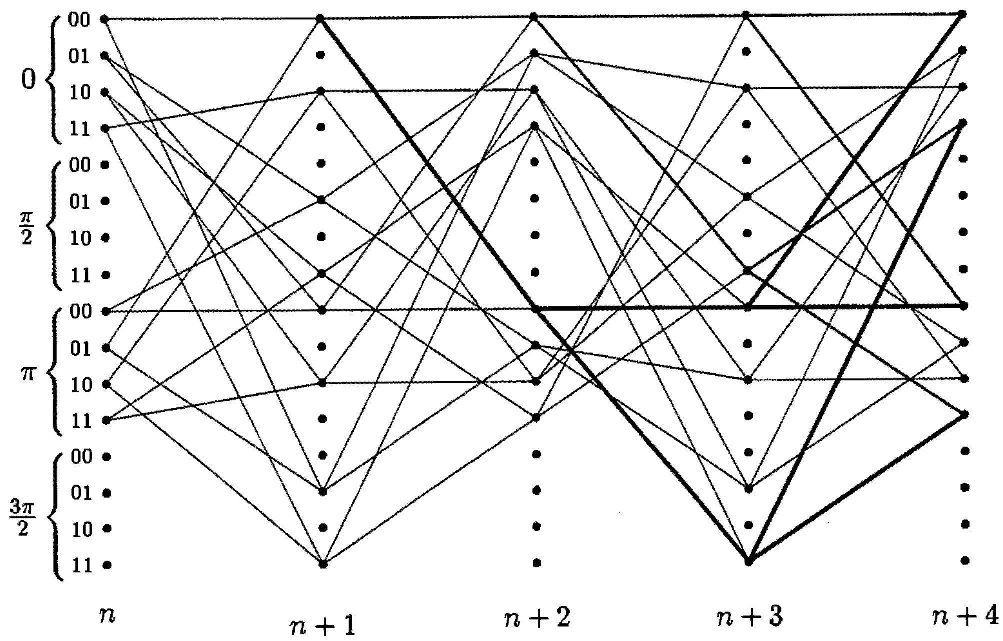



Figure 4.12: Combined CE and CPE trellis for Hodgart's QMSK scheme C.

ratio

$$\Lambda_{00} = \frac{l_{000}}{l_{001}} = \frac{a_{00}c_{02}c_{03}(b_{00}c_{00}c_{01} + b_{02})(c_{00} + c_{01})}{a_{01}c_{00}c_{01}(b_{11} + b_{13}c_{02}c_{03})(c_{02} + c_{03})} \tag{4.9}$$

does not simplify further. One may readily verify that the likelihood ratios from every other state in the trellis also do not simplify. This kind of checking is greatly speeded by the use of a computer program which does symbolic simplification of expressions. It is only necessary to find one state from which the likelihood ratio does not simplify to pronounce the trellis unusable for a simplified receiver.

Extending the observation period to four or more intervals also does not produce a simplification. Indeed, an examination of the trellis shows that a longer observation period than three need not be considered, because the shortest free distance in the trellis is three. It is always possible to find pairs of paths through a trellis which never merge, but that is not what is meant here. It is the first possible merging of two paths which initially separated from a common state, which determines the modulation's free distance and hence its error performance [3]. Section 3.4.4 dealt with free distance and constraint length in more detail.

Hodgart's QMSK schemes B and C similarly do not simplify. The combined trellis of scheme C, for example, is shown in Figure 4.12. The paths from the first (top) state in the second interval are shown in heavier lines. In this trellis, some paths merge in three intervals after splitting, while others need five. This is probably not a desirable property, both from the viewpoint of maximizing free Euclidean distance, and for defining a receiver. From the figure, the likelihood ratio from the given state over the paths shown in heavy lines, is

$$\begin{aligned}\Lambda_{00} &= \frac{b_{00}\left[c_{00}(d_{00}+d_{02})+c_{01}(d_{11}+d_{13})\right]}{b_{02}\left[c_{20}(d_{20}+d_{22})+c_{21}(d_{31}+d_{33})\right]} \\ &= \frac{b_{00}c_{00}c_{01}d_{00}d_{02}d_{11}d_{13}\left[c_{00}(d_{00}+d_{02})+c_{01}(d_{11}+d_{13})\right]}{b_{02}\left[c_{00}d_{00}d_{02}(d_{11}+d_{13})+c_{01}d_{11}d_{13}(d_{00}+d_{02})\right]} \qquad (4.10)\end{aligned}$$

It is clear from the figure why it is not necessary to have an observation period longer than three symbols from this state. From other states, an observation period of five symbols may be necessary.

### 4.2.4 The search for a simplified receiver for QMSK

The preceding analysis of the Hodgart schemes suggests a way of automating the search for trellises with simplifying likelihood ratios. A computer program can generate all allowable trellises and check each one for simplification of its likelihood ratios, free Euclidean distance, and catastrophic trellises. Each of these tests will be discussed individually below, but first one needs an idea of how many trellises are to be generated and examined.

#### Minimizing the number of QMSK trellises

Looking at the combined trellises in Figure 4.11 or Figure 4.12, it is apparent that only half the states are used in any one interval. If the CE has four states, for example, then the actual number of used states is $n_s = 4 \times 4/2 = 8$.

Calculating the number of possible QMSK trellis codes is done in two steps. First it is assumed that it is known which used states are chosen: these are fixed over a single interval, so that the number of possible trellises $T(n_s)$ with a fixed number of states $n_s = ZP/2$ may be calculated. $Z$ is the number of CE

| $n_s$ | $T(n_s)$ |
|---|---|
| 2 | 1 |
| 3 | 6 |
| 4 | 90 |
| 5 | 2040 |
| 6 | 67950 |
| 7 | 3110940 |
| 8 | 187530840 |

Table 4.3: Number of possible trellises as a function of the number of states.

states and $P$ is the number of CPE states.

$T(n_s)$ is difficult to enumerate, and the problem was only solved recently [57], apparently for the first time. Appendix A describes the problem in more detail, and Table 4.3 summarizes the number of trellises over one interval $T(n_s)$ as a function of $n_s$ up to $n_s = 8$. With the available computing resources, a practical limit of $Z \leq 4$ is placed on the number of CE states.

In the second step, the number of ways of choosing $n_s$ used states from a total of $2n_s$ in one interval is calculated as

$$\binom{2n_s}{n_s} = \frac{(2n_s)!}{(n_s!)^2} \tag{4.11}$$

For $n_s = 8$ states, there are 12870 combinations of used (or unused) states in every odd symbol interval. In even intervals, the choice of used states is fixed. If we want to examine all possible $Z = 4$ trellises over a CE cycle length of $J = 4$, for example, we shall need to generate

$$\binom{2n_s}{n_s}^{\frac{J}{2}} T^J(n_s) \approx 2 \times 10^{41} \tag{4.12}$$

trellises for $n_s = ZP/2 = 8$. If we could process a thousand million trellises per second on a massively parallel supercomputer, it would take $6 \times 10^{24}$ years to complete the search. The brute-force approach needs to be tempered with some ingenuity.

Various rules may be formulated for pruning these numbers effectively, as described in Appendix A. These pruning rules were encapsulated in a program called USED which generated all possible combinations of used states, and then applied various shifting and mirroring transformations to prune trellises which may be derived from one another. The result was a more manageable number of combinations of used states.

The output of USED was a file of candidate used state combinations, to be fed into the next stage of processing: a program called ALL. ALL generated all possible trellises with the given combination of used states and ran each one through the battery of tests described in the subsections to follow.

To minimize the number of trellises examined, ALL had to try to generate as few trellises as possible. This was not as easy to do as in the case of USED, because it is not so clear from the outset which trellises may be safely pruned without discarding a potential simplified receiver. Parallel transitions were not generated from the outset, that is, branches from the same state which merge again immediately in a common next state. A parallel transition in a QMSK trellis cannot be differentiated at the receiver because its branches differ by a Euclidean distance of zero.

Appendix A gives more information on how the algorithms limit the number of trellises.

### Likelihood ratio simplification

As each possible trellis was generated, the fastest selection criterion was in fact the symbolic simplification of its likelihood ratios from each state in the trellis. It was important for the efficiency of the algorithm at this level of iteration, that trellises be discarded as soon as possible if they turned out to be unsuitable. The likelihood ratio simplification, being the strictest of the tests, caused the vast majority of trellises to be discarded on the first state tested. Only those trellises which had likelihood ratios that simplified in every state of the trellis, were passed on to the next two tests.

Symbolic simplification of a likelihood ratio-like expression is relatively easy to implement as a computer program, because of the regularity of the expression. The two likelihood functions from each state can be generated in the form of, for example,

$$\begin{aligned} l &= b_{00}c_{00}d_{00} + b_{00}c_{00}d_{02} + b_{00}c_{01}d_{11} + b_{00}c_{01}d_{13} + \\ &\quad b_{02}\frac{1}{c_{00}}\frac{1}{d_{00}} + b_{02}\frac{1}{c_{00}}\frac{1}{d_{02}} + b_{02}\frac{1}{c_{01}}\frac{1}{d_{11}} + b_{02}\frac{1}{c_{01}}\frac{1}{d_{13}} \end{aligned} \tag{4.13}$$

which was taken from Hodgart's scheme A. Such an expression may be represented by an array of integers (which is what I used) or by a more sophisticated data structure such as a linked list or tree. Each likelihood function is then simplified separately, by removing from it its lowest common denominator and any other factors common to all terms. If the remaining sum of terms is the same for the two likelihood functions, then their likelihood ratio will simplify by cancellation of the sums. In (4.13) above, the $a_{00}$ factor from the first interval was not included, because it would have been removed anyway as a common factor during the simplifying process.

A detailed description of the simplification algorithm is given in Appendix B, sections B.1 and B.8. Note that if the two likelihood functions from a state are identical (so that their likelihood ratio is unity), the trellis is degenerate and should be discarded.

### Catastrophic trellises

If a trellis passed the likelihood ratio simplification test, it may still be catastrophic. The $Z = 2$, $J = 4$ trellis in Figure 4.13 has likelihood ratios which simplify from every state, it uses legal QMSK signals, it even has a free Euclidean distance of $d^2_{\text{free}} = 4$, which is a 3 dB improvement over MSK. Yet its likelihood ratio is, from the the first state for example:

$$\begin{aligned} \Lambda_{00} &= \frac{a_{00}\left[b_{00}\left(c_{01}+c_{03}\right)+b_{02}\left(c_{21}+c_{23}\right)\right]}{a_{02}\left[b_{20}\left(c_{21}+c_{23}\right)+b_{22}\left(c_{01}+c_{03}\right)\right]} \\ &= \frac{a_{00}b_{00}b_{02}}{a_{02}} \end{aligned} \tag{4.14}$$

This is MSK's receiver, even though the trellis has a constraint length of three as one may verify from Figure 4.13.

Referring to the definition of catastrophic codes in Chapter 3, the trellis of

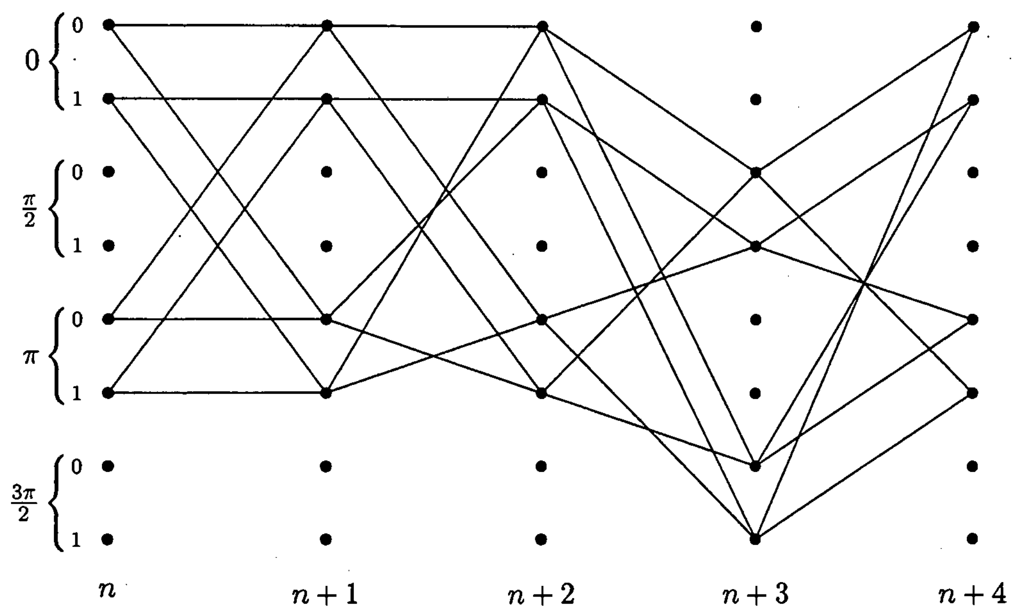



Figure 4.13: Example of a catastrophic QMSK trellis with $Z = 2$ and $J = 4$, which otherwise seems valid.

Figure 4.13 was clearly catastrophic. For example, starting from state 00 (meaning phase state 0, encoder state 0), the infinite CPE input sequences (frequencies) $\boldsymbol{u} = (0,0,1,3,0,0,1,3,0,0,\ldots)$ and $\boldsymbol{u}' = (2,2,0,0,1,3,0,0,1,3,0,0,\ldots)$ accumulated a finite NSED of 2 between them, while the corresponding sequences of CE states differed in every position except the first: they were $(0,0,0,\ldots)$ and $(0,0,1,1,1,\ldots)$ respectively.

The frequency shift in the third and fourth intervals of the trellis is just that: MSK with lower signalling frequency $\omega_0$ for the first two intervals, followed by MSK at a slightly higher frequency $\omega_0 + \pi/2T$ for the next two intervals, and so on.

Checking for catastrophic trellises in ALL was therefore necessary, if somewhat tedious. All potential zero-distance loops had to be traced through the trellis recursively until they either diverged or remerged with themselves to form a closed loop. Note that such loops may span more than $J$ intervals in a figure-of-eight fashion, which made an exhaustive search expensive in terms of processing time. The algorithm in ALL was carefully optimized to avoid redundant searching.

The strict definition of a catastrophic scheme was later relaxed somewhat to examine the possibility that some catastrophic schemes may be practically useful. Section 5.1.1 of Chapter 5 gives more detail.

### Free Euclidean distance

The final qualifying check done on a trellis which passed the two previous tests, was to measure its free Euclidean distance, and discard the trellis if it was below a certain threshold. The threshold was usually set at a free Euclidean distance of $d^2_{\mathrm{free}} = 3$, which is a modest 1.8 dB improvement on MSK. If this was not done, ALL found many non-catastrophic QMSK trellises with simplified receivers, but with no coding gain over MSK. Examples are given in the next subsection.

The automated calculation of Euclidean distance in the algorithm was facilitated by the precalculation of a table of incremental normalized correlation coefficients $\Delta\rho(\Delta v_n, \Delta u_n)$ for $M = 2,\ h = 1/4$, as described in section 1.9. Table 4.4 shows the result of calculating, from (1.67) and (1.23),

$$\begin{aligned}
\Delta\rho(\Delta v_n, \Delta u_n) &= \frac{1}{T}\int_0^T \cos\left[\frac{\pi}{2}\left(\Delta v_n + \Delta u_n \tau / T\right)\right]\, d\tau \\
&= \cos\frac{\pi}{2}\Delta v_n \int_0^1 \cos\frac{\pi}{2}\Delta u_n \tau\, d\tau - \sin\frac{\pi}{2}\Delta v_n \int_0^1 \sin\frac{\pi}{2}\Delta u_n \tau\, d\tau \\
&= \begin{cases} \cos\frac{\pi}{2}\Delta v_n\,, & \Delta u_n = 0 \\ \frac{2}{\pi \Delta u_n}\left[\cos\frac{\pi}{2}\Delta v_n \sin\frac{\pi}{2}\Delta u_n - \sin\frac{\pi}{2}\Delta v_n\left(1-\cos\frac{\pi}{2}u_n\right)\right], & \\ & \Delta u_n \neq 0 \end{cases}
\end{aligned} \tag{4.15}$$

As an example of the use of Table 4.4, the NSED between two paths in Figure 4.13 are calculated in Table 4.5.

### Examples of candidate QMSK trellises

The main result of the search was that it seems likely that there is no simplified receiver for QMSK, as the exhaustive searches turned up nothing. How does one generate some confidence in this result? How do we know that the negative result is not due to a logical fault or a programming error? One way is to relax

| $\Delta v_n$ \ $\Delta u_n$ | $-3$ | $-2$ | $-1$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|---|---|---|
| $-3$ | $\frac{2}{3\pi}$ | $\frac{2}{\pi}$ | $\frac{2}{\pi}$ | $0$ | $-\frac{2}{\pi}$ | $-\frac{2}{\pi}$ | $-\frac{2}{3\pi}$ |
| $-2$ | $\frac{2}{3\pi}$ | $0$ | $-\frac{2}{\pi}$ | $-1$ | $-\frac{2}{\pi}$ | $0$ | $\frac{2}{3\pi}$ |
| $-1$ | $-\frac{2}{3\pi}$ | $-\frac{2}{\pi}$ | $-\frac{2}{\pi}$ | $0$ | $\frac{2}{\pi}$ | $\frac{2}{\pi}$ | $\frac{2}{3\pi}$ |
| $0$ | $-\frac{2}{3\pi}$ | $0$ | $\frac{2}{\pi}$ | $1$ | $\frac{2}{\pi}$ | $0$ | $-\frac{2}{3\pi}$ |
| $1$ | $\frac{2}{3\pi}$ | $\frac{2}{\pi}$ | $\frac{2}{\pi}$ | $0$ | $-\frac{2}{\pi}$ | $-\frac{2}{\pi}$ | $-\frac{2}{3\pi}$ |
| $2$ | $\frac{2}{3\pi}$ | $0$ | $-\frac{2}{\pi}$ | $-1$ | $-\frac{2}{\pi}$ | $0$ | $\frac{2}{3\pi}$ |
| $3$ | $-\frac{2}{3\pi}$ | $-\frac{2}{\pi}$ | $-\frac{2}{\pi}$ | $0$ | $\frac{2}{\pi}$ | $\frac{2}{\pi}$ | $\frac{2}{3\pi}$ |

Table 4.4: Incremental normalized correlation coefficients $\Delta\rho(\Delta v_n, \Delta u_n)$ for $M = 2$ and $h = 1/4$.

| | $n$ | $n+1$ | $n+2$ |
|---|---|---|---|
| $\Delta v$: | $0$ | $-2$ | $-2$ |
| $\Delta u$: | $-2$ | $0$ | $-2$ |
| $\Delta\rho$: | $0$ | $-1$ | $0$ |

From (1.66): $\text{NSED} = 3 - (0 - 1 + 0) = 4$

Table 4.5: Example calculation of the NSED between two paths in Figure 4.13 for the common starting state 00 and ending state 10, and frequencies $\boldsymbol{u} = (0, 0, 1)$ for one path and $\boldsymbol{u}' = (2, 0, 3)$ for the other.

| | | Left out | | | |
|---|---|---|---|---|---|
| $Z$ | $J$ | LR simplification | Not catastrophic | Free ED | Trellises tested |
| 2 | 2 | 24 (3) | 16 (4) | 256 (2) | 10800 |
| 2 | 4 | 1344 (3) | 512 (4) | 72704 (2) | 58786560 |

Table 4.6: Results of relaxing the qualifying tests for QMSK trellises. The entries are the numbers of trellises found for each combination of $Z$ and $J$, with the free Euclidean distance in brackets.

some of the selection criteria for the trellises in order to see that ALL comes up with plausible candidate schemes. The three qualifying tests were

1. likelihood ratio simplification from every state;

2. not catastrophic; and

3. free squared Euclidean distance of at least 3.

The results of relaxing one of these at a time are summarized in Table 4.6 for $Z = 2$ CE states and CE cycle length $J = 2$ and 4. The first trellis found in each case for $J = 4$ is shown in Figure 4.13 on page 91, and Figures 4.14 and 4.15 below. It is interesting to verify how each of them fails to meet the particular omitted requirement, and meets the rest.

### 4.2.5 Number of trellises tested

With the computing resources available, the only exhaustive searches achieved were for $Z = 2$ CE states and $J = 2$ and $J = 4$ CE cycle lengths. The number of trellises examined in these two cases were 10800 and 58786560 respectively.

Partial searches were done for $Z = 2,\ J = 8$, and for $Z = 4,\ J = 2$ and 4. The total numbers of trellises examined in each case are shown in Table 4.7. Most of the searches were done on relatively fast 'workstation' computers, but the partial search for $Z = 4,\ J = 2$ was started simultaneously on 22 'personal' computers, one for each of the 22 most promising combinations of used states generated by USED. This parallel search averaged 300 million trellises per combination. In addition, the most promising combination was searched to a depth of 1600 million trellises.

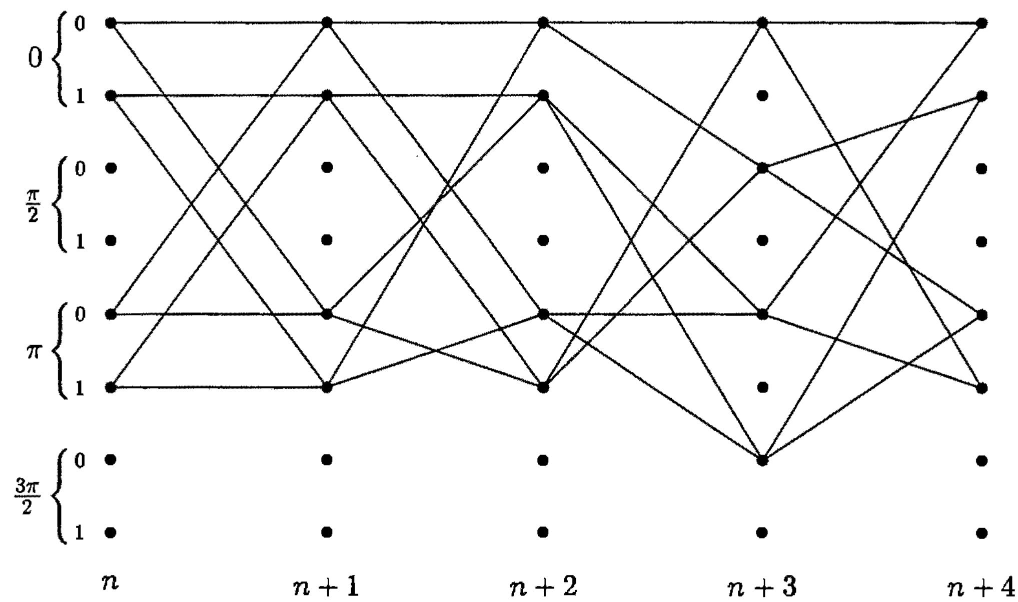



Figure 4.14: QMSK trellis found by suppressing the likelihood ratio simplification test.

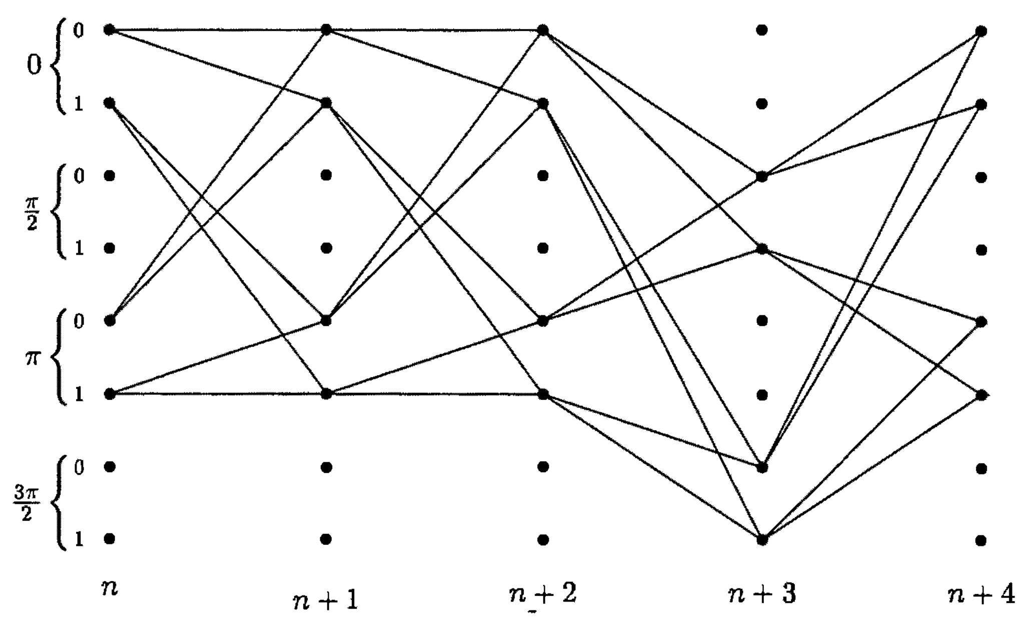



Figure 4.15: QMSK trellis found by lowering the minimum free Euclidean distance threshold.

| $Z$ | $J$ | Trellises | Completion |
|---|---|---|---|
| 2 | 2 | 10800 | exhaustive |
| 2 | 4 | 58786560 | exhaustive |
| 2 | 8 | $3 \times 10^9$ | partial |
| 4 | 2 | $13 \times 10^9$ | partial |
| 4 | 4 | $10 \times 10^9$ | partial |

Table 4.7: Total number of trellises generated and tested, arranged by number of CE states $Z$ and CE cycle length $J$.

## 4.3 Search for Identities to Aid Simplification

This section describes the search method for finding mathematical identities which may help in simplifying likelihood ratios of trellis-coded modulations. It starts by describing an exploration of variations of the MSK identity (4.1) in section 4.3.1. Some identities were found, but their usefulness is in question because they do not seem to be of the form of likelihood ratio expressions.

Section 4.3.2 therefore describes a search based on generating expressions which resemble likelihood ratios. The expressions are generated and simplified with the help of a computer program, to see which ones simplify into single terms, thus forming identities. However, the method of generating all possible likelihood ratio expressions was far too general in this approach, and only the trivial case of binary modulation over two symbol intervals could be covered.

A more sophisticated program was then written, as described in section 4.3.3 and Appendix B, to generate and simplify likelihood ratio expressions based on actual trellises. Care was taken only to generate mathematically distinct expressions, in which the program was reasonably successful. This method yielded some interesting identities up to an observation period of three intervals.

The MSK identity (4.1) is a quotient of Laurent polynomials with all coefficients unity. It may be difficult to prove the existence or not of more identities of this kind. There is a restricted class of identities which follows from the magnitudes of complex numbers:

$$\begin{aligned} (a+jb)(c+jd) &= (ac-bd)+j(ad+bc) \\ \Rightarrow\quad (a^2+b^2)(c^2+d^2) &= (ac-bd)^2+(ad+bc)^2 \qquad (4.16) \end{aligned}$$

but these are not likely to be of help here.[2]

### 4.3.1 Search for MSK-like identities

The process of simplifying a likelihood ratio may be formulated in terms of the likelihood transform of its numerator $f()$, its denominator $g()$ and the simplified result $h()$. For example, in four variables we may write

$$\frac{f(a,b,c,d)}{g(a,b,c,d)} = h(a,b,c,d), \quad a,b,c,d > 0, \tag{4.17}$$

where $f()$ and $g()$ are of the form

$$\begin{aligned} f(a,b,c,d) &= f_a(a) + f_b(b) + f_c(c) + f_d(d) \\ g(a,b,c,d) &= g_a(a) + g_b(b) + g_c(c) + g_d(d) \end{aligned} \tag{4.18}$$

where

$$\begin{aligned} f_p(p) &= p^i, \\ g_p(p) &= p^j, \quad i,j \in \mathbb{Z}, \ p > 0 \end{aligned} \tag{4.19}$$

and $\mathbb{Z}$ is the field of integers. In other words, the numerator and denominator of the likelihood ratio are assumed to be sums of variables to integer powers. The result $h$ is a product of integer powers of its variables, containing no sums at all:

$$h(a,b,c,d) = a^i b^j c^k d^l, \quad i,j,k,l \in \mathbb{Z} \tag{4.20}$$

Only powers of 1 and $-1$ have been encountered so far, and these are likely to be the most important cases, but the search was widened later to include other powers as well in case an identity was found. The first search was done by simplifying all combinations of $f()$ and $g()$ with powers of 1, 0 and $-1$ for two to eight terms each, with the restriction

$$g_p(p) = \frac{1}{f_p(p)}, \quad p > 0 \tag{4.21}$$

[2]I am indebted to Drs. Ken Hughes and Zaid Kimmie, both of the Department of Mathematics, University of Cape Town, for useful discussions on this subject.

and checking to see if the result is of the form of (4.20), in other words, comprising a single term. I wrote computer programs to generate the expressions and write them in text files. They were then simplified with a symbolic mathematics program.

For example, with four terms in $f()$ and $g()$:

$$\begin{aligned}
\frac{a+b+c+d}{a^{-1}+b^{-1}+c^{-1}+d^{-1}} &= \frac{abcd(a+b+c+d)}{abc+abd+acd+bcd} \\
\frac{a+b+c+d^{0}}{a^{-1}+b^{-1}+c^{-1}+d^{0}} &= \frac{abc(a+b+c+d)}{abc+ab+ac+bc} \\
\frac{a+b+c+d^{-1}}{a^{-1}+b^{-1}+c^{-1}+d} &= \frac{abc(ad+bd+cd+1)}{d(abcd+ab+ac+bc)} \\
\vdots \ \text{etc}
\end{aligned} \tag{4.22}$$

With two terms the search produced the identity (4.1) as expected, but nothing else. To cover some more possibilities the numerator $f()$ was fixed as a sum of powers of 1, while the denominator $g()$ was varied through all combinations of powers of 1, 0 and $-1$. Again the search was done seven times with two to eight terms in $f()$ and $g()$. For example, for four terms the search generated:

$$\begin{aligned}
\frac{a+b+c+d}{a^{-1}+b^{-1}+c^{-1}+d^{-1}} &= \frac{abcd(a+b+c+d)}{abc+abd+acd+bcd} \\
\frac{a+b+c+d}{a^{-1}+b^{-1}+c^{-1}+d^{0}} &= \frac{abc(a+b+c+d)}{abc+ab+ac+bc} \\
\frac{a+b+c+d}{a^{-1}+b^{-1}+c^{-1}+d^{1}} &= \frac{abc(a+b+c+d)}{abcd+ab+ac+bc} \\
\vdots \ \text{etc}
\end{aligned} \tag{4.23}$$

Another approach was to reorganize the simplification equation (4.17) somewhat as

$$\frac{f(a,b,c,d)}{h(a,b,c,d)} = g(a,b,c,d) \tag{4.24}$$

and to check whether a simplification produces something of the form of (4.18), that is, a sum of terms. Fixing the numerator $f()$ as before and trying all combinations of powers of 1, 0 and $-1$ as before, the search with four terms

| Number of terms | Range of powers | Number of combinations |
|---|---|---|
| 2 | $-7$ to 7 | 120 |
| 3 | $-3$ to 3 | 84 |
| 4 | $-2$ to 2 | 70 |
| 5 to 8 | $-1$ to 1 | |

Table 4.8: Powers and numbers of terms searched in all combinations for likelihood ratio simplification.

covered

$$\begin{aligned}
\frac{a+b+c+d}{a^{-1}b^{-1}c^{-1}d^{-1}} &= a^2bcd + ab^2cd + abc^2d + abcd^2 \\
\frac{a+b+c+d}{a^{-1}b^{-1}c^{-1}d^{0}} &= a^2bc + ab^2c + abc^2 + abcd \\
\frac{a+b+c+d}{a^{-1}b^{-1}c^{-1}d^{1}} &= \frac{a^2bc}{d} + \frac{ab^2c}{d} + \frac{abc^2}{d} + abc \\
\vdots \ \text{etc}
\end{aligned} \tag{4.25}$$

This was iterated with two to eight terms as before.

The search was then widened to include powers other than 1, 0 and $-1$, for each of the three variants described above. The equations are not explicitly listed here: they are similar to the ones seen before. Table 4.8 summarizes the ranges of powers covered for various numbers of terms. The ranges of powers were limited by the large number of possible combinations of powers and terms that comprised each expression, for which simplification had to be attempted.

The widened search produced a class of identities with two terms:

$$\frac{a^i + b^j}{a^{-i} + b^{-j}} = a^i b^j, \quad a, b > 0, \ \ i, j \in \mathbb{Z} \tag{4.26}$$

for example

$$\frac{a^3 + \frac{1}{b}}{\frac{1}{a^3} + b} = \frac{a^3}{b} \tag{4.27}$$

These are not likely to occur in the likelihood ratios of equal-energy modulations, but they may well be useful in the analysis of coded $M$-ary pulse amplitude modulation (PAM) schemes.

### 4.3.2 A more general search for identities

#### General likelihood ratios

A general expression for any trellis-based likelihood ratio is

$$\begin{aligned}\Lambda &= \frac{a(c(g(\cdots)+h(\cdots)+\cdots)+d(\cdots)+\cdots)}{b(e(i(\cdots)+j(\cdots)+\cdots)+f(\cdots)+\cdots)}\\ &= \frac{acg\cdots+ach\cdots+ad\cdots+\cdots}{bei\cdots+bej\cdots+bf\cdots+\cdots}\end{aligned} \tag{4.28}$$

where $a, b, c, \cdots$ are arbitrary positive real variables. Apart from the common factors $a$ and $b$, representing the two branches being compared, the numerator and the denominator are recursive sums to a depth $N-1$. The variables are not necessarily distinct: any two of them may be related to one another or even equal. The different variables in (4.28) may therefore be thought of as place-holders for LTVs chosen from a smaller set corresponding to the signal set. The likelihood ratio is used to decide between the sequences starting with branch $a$ and those starting with branch $b$.

For example, any binary trellis ($M=2$) over an observation period of three intervals ($N=3$) can be written in the form

$$\begin{aligned}\Lambda &= \frac{a(c(g+h)+d(i+j))}{b(e(k+l)+f(m+n))}\\ &= \frac{acg+ach+adi+adj}{bek+bel+bfm+bfn}\end{aligned} \tag{4.29}$$

One possible trellis which corresponds to this expression is shown in Figure 4.16 as an example.

The number of branches from each state is equal to the size of the input alphabet $M$, so that every distinct input follows a different branch. The length of the observation period $N$ determines the number of variables in each term of (4.28) or (4.29). (4.28) may be compared to (1.54) as a way of seeing the one-to-one correspondence between the likelihood transform expression and its likelihood parameter receiver. Each LTV represents an exponential of a

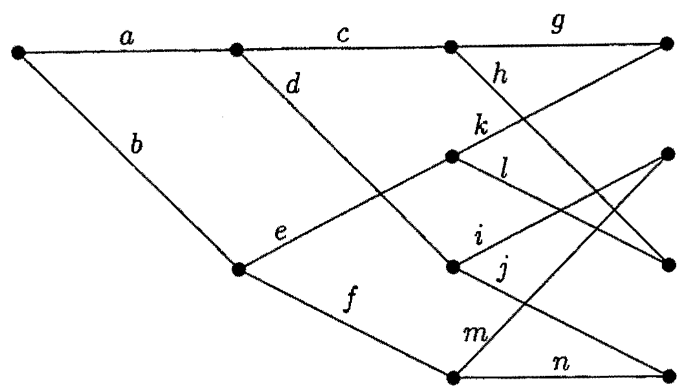



Figure 4.16: A partial trellis corresponding to the general likelihood ratio of (4.29).

correlation over one symbol interval.

### Abbreviated likelihood ratios

For the purpose of determining whether a likelihood ratio simplifies, the first-interval common factors $a$ and $b$ may be left out from the outset, and that is what will be done for the rest of the dissertation. The expressions that result when the common factors from the first interval of a likelihood ratio are left out, will be called *abbreviated likelihood ratios.* The numerator and denominator of an abbreviated likelihood ratio are regular recursive sums to a depth of $N-1$. Continuing the present example (4.29), the abbreviated likelihood ratio is

$$\begin{aligned}\Lambda' &= \frac{b}{a}\Lambda \\ &= \frac{c(g+h)+d(i+j)}{e(k+l)+f(m+n)} \\ &= \frac{cg+ch+di+dj}{ek+el+fm+fn} \qquad (4.30)\end{aligned}$$

The variables of the first interval may be arbitrarily chosen, as $a$ and $b$ for example, because they are always present in any likelihood ratio as representing the first pair of branches from the starting node. The use of abbreviated likelihood ratios is for convenience only and should be understood to imply the presence of these two variables even if they are not explicitly mentioned. An abbreviated likelihood ratio explicitly contains variables from one fewer interval than its full counterpart.

### The search

A first, naïve approach to generating all possible likelihood ratios of this form, is to regard each of the variables in an abbreviated likelihood ratio such as (4.30) as a place-holder, and to fill the place-holders with all possible combinations of variables. Each expression so generated is then simplified to check whether it results in an expression comprising a single term.

Still assuming $M = 2$, $N = 3$, one would start, for example, with the trivial expression

$$\frac{aa + aa + aa + aa}{aa + aa + aa + aa} = 1$$

followed by

$$\begin{aligned} \frac{aa + aa + aa + aa}{aa + aa + aa + ab} &= \frac{4a}{3a + b} \\ \frac{aa + aa + aa + aa}{aa + aa + aa + ac} &= \frac{4a}{3a + c} \end{aligned} \tag{4.31}$$

and so on. The number of variables to choose from is restricted to the number of place-holders in the numerator (or denominator). A place-holder may be filled either with a variable or its reciprocal. In the present example, up to eight variables $(a, b, \cdots, h)$ and their eight reciprocals $(1/a, 1/b, \cdots, 1/h)$ may be used.

In order to avoid unnecessary duplication due to the commutative properties of addition and multiplication, the first place-holder in the numerator (and in the denominator) is filled from the set of variables $\{a, 1/a\}$, the second place-holder is filled from $\{a, 1/a, b, 1/b\}$, the third from $\{a, 1/a, b, 1/b, c, 1/c\}$ and so on. In addition, the first place-holder in the numerator is fixed as $a$, all the other variables being relative to it anyway.

Note that the expressions so generated are not necessarily realistic, in other words, only a small fraction are likely to form any part of a likelihood ratio generated from an actual code trellis. Different expressions are also not necessarily mathematically distinct, in that one may often be obtained from another

| $M$ | $N$ | $V$ | $T_e$ |
|---|---|---|---|
| 2 | 2 | 2 | 32 |
| 2 | 3 | 8 | $5.3 \times 10^{13}$ |
| 2 | 4 | 24 | $5.4 \times 10^{61}$ |
| 3 | 2 | 9 | $1.7 \times 10^{16}$ |

Table 4.9: Number of possible expressions in the first attempt at finding useful identities.

by simple substitution of variables (as in the case of the last two expressions in (4.31) above).

### The number of expressions

To calculate the total number of expressions generated, the numerator and denominator are found each have to have $Q$ terms,

$$Q = M^{N-1} \tag{4.32}$$

each term comprising a product of $N-1$ variables. The total number of place-holders in the numerator (or the denominator) is then

$$V = (N-1)Q = (N-1)M^{N-1} \tag{4.33}$$

The denominator therefore takes one of $2 \times 4 \times 6 \times \cdots \times 2V = 2^V \cdot V!$ forms, and the numerator half as many because the variable in the first place-holder is fixed at $a$. The total number of expressions to test is then

$$T_e = 2^{2V-1}(V!)^2 = 2^{2(N-1)M^{N-1}-1}\left(\left((N-1)M^{N-1}\right)!\right)^2 \tag{4.34}$$

which grows extremely fast with increasing $V$, or $M$ and $N$. Table 4.9 tabulates the first few values of $T_e$.

### Results

None of the cases other than the first one ($M = 2$, $N = 2$) in Table 4.9 were tractable with the computing resources at my disposal. Clearly, a more selective way of generating plausible expressions was needed.

The following expressions were found to simplify for input alphabet size $M = 2$ and observation period $N = 2$:

$$\begin{aligned}
\frac{a+b}{\frac{1}{a}+\frac{1}{b}} &= ab \\
\frac{a+a}{\frac{1}{a}+\frac{1}{a}} &= a^2 \\
\frac{a+\frac{1}{a}}{\frac{1}{a}+a} &= 1 \\
\frac{a+\frac{1}{b}}{\frac{1}{a}+b} &= \frac{a}{b}
\end{aligned} \tag{4.35}$$

The first and last equations are recognizable as two forms of the MSK identity, while the middle two are not useful.

### 4.3.3 Search for identities derived from trellises

Instead of an undirected search for identities that may possibly be of use in the simplification of likelihood ratios, the search is now focused on likelihood ratios that can reasonably be expected to follow from actual trellises.

#### Assumption of half-rate coding

The first difference between the current search and the previous one, is that the assumption is made that, if the coded modulation scheme has an input alphabet of $M$ symbols, and a modulator capable of producing one of $m$ channel waveforms, then

$$m = 2M \tag{4.36}$$

In other words, the code rate is $R_c = 1/2$. The reason for this assumption is that, apart from reciprocity ($a \leftrightarrow 1/a$), there is no other mathematical relationship between two positive real variables, each of which is an exponential of an equal-energy signal correlation, that may aid in simplification of an expression. More formally, if $x$ and $y$ are real numbers representing correlations with

$$|x| = |y|, \quad x \neq y \tag{4.37}$$

related by some $y = f(x)$, then the only algebraic function which relates $e^x$ and $e^y$ is

$$y = -x \Leftrightarrow e^y = \frac{1}{e^x}, \quad x, y > 0 \tag{4.38}$$

because the equation on the left is the only solution of (4.37). This will be formally proved in Chapter 6.

The signal set may typically be bi-orthogonal [90], although orthogonality is not strictly necessary. The set comprises $m$ signals which may be grouped into $M$ pairs, each member of a pair being the inverse of the other: $s_{1/a}(\tau) = -s_a(\tau)$. The correlation of each member of such a pair of signals with the received signal, followed by the exponentiation of each, results in a pair of positive real variables which are reciprocals of one another. An expression such as '$1/a$' will henceforth be treated as a single variable which refers to the signal which is the inverse of the signal referred to by the variable '$a$'.

### Standard graphical representation of likelihood ratios

The correspondence between a likelihood ratio and its trellis is not unique. For example, the abbreviated likelihood ratio ($M = 2$, $N = 3$)

$$\Lambda = \frac{a_1(a_2 + b_2) + b_1(a_2 + b_2)}{\frac{1}{a_1}\left(\frac{1}{a_2} + \frac{1}{b_2}\right) + \frac{1}{b_1}\left(\frac{1}{a_2} + \frac{1}{b_2}\right)} \tag{4.39}$$

describes any of the trellises in Figure 4.17 equally well. (4.39) introduces a change in notation: subscripts of LTVs now refer to the interval number $n$. LTVs may in general be defined in any convenient way, and the new notation places less emphasis on the physical signal set.

The first 'trellis' (a) in Figure 4.17 is in fact a tree, and this is the graphical representation which follows most readily from the likelihood ratio itself. Likelihood ratios can always be represented as trees even when they are derived from trellises. When the number and labelling of states are not relevant to the current discussion, likelihood ratios will be represented as trees rather than trellises, because trees have an unambiguous correspondence with their likelihood ratios. It certainly does not matter to the simplified receiver what particular graphical

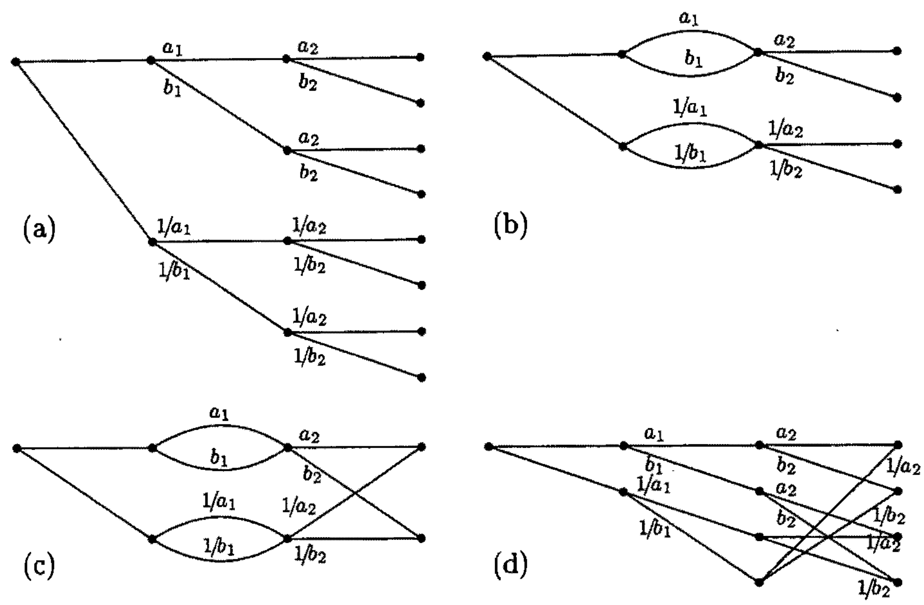



Figure 4.17: Different partial trellises with the same $M = 2$, $N = 3$ abbreviated likelihood ratio (4.39). (a) is the standard tree representation.

representation is associated with its likelihood ratio.

Once a tree has been drawn, it is easy to identify the recurring parts of the likelihood ratio which indicate where branches may merge to form a trellis. Figure 4.17 (b) to (d) shows a few possible trellises for (4.39). These trellises need not be regular, but may have odd numbers of merging branches, be time-varying, and so forth.

### The more directed search

It is easy to show that all realistic likelihood ratios must be of the form of (4.28):

$$\begin{aligned}\Lambda &= \frac{a_0(a_1(a_2(\cdots) + b_2(\cdots) + \cdots) + b_1(\cdots) + \cdots)}{b_0(c_1(c_2(\cdots) + d_2(\cdots) + \cdots) + d_1(\cdots) + \cdots)} \\ &= \frac{a_0a_1a_2\cdots + a_0a_1b_2\cdots + a_0b_1\cdots + \cdots}{b_0c_1c_2\cdots + b_0c_1d_2\cdots + b_0d_1\cdots + \cdots} \qquad (4.40)\end{aligned}$$

with all variables positive real. These arbitrary variables may be regarded as place-holders for the likelihood transform variables, which are chosen from a smaller set of $M$ reciprocal pairs of variables as discussed before. In this

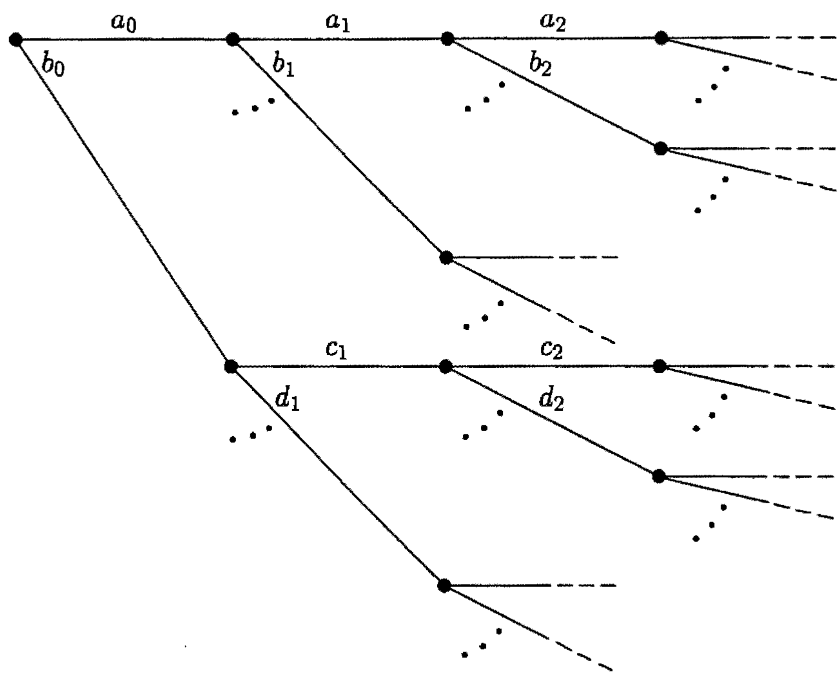



Figure 4.18: The tree representing the general likelihood ratio of (4.40).

general form, (4.40) may be thought of as a template, with its place-holders still to be filled with likelihood transform variables. It is also simplest to regard it as representing a tree rather than any particular trellis, as illustrated in Figure 4.18.

### Separating the intervals

The next difference from the previous search is now introduced: the same variable can never appear in different signalling intervals. This must be the case because, even if the same signal is transmitted in two different intervals, the assumption of additive white noise in the channel means that the random variables representing the outputs of the matched filter for that signal are uncorrelated from one interval to the next.

When generating all possible likelihood ratio expressions (for a given $M$ and $N$), one must therefore choose variables for each signalling interval from disjoint sets of variables. Each one of these disjoint sets may contain up to $m = 2M$ variables as before. For the sake of generality, a set need not contain all $m$ possible variables, nor is it necessary for each variable's reciprocal to be present.

This new restriction is achieved simply by continuing the use of the present notation in which variables from different intervals are given different subscripts, together with the understanding that variables with different subscripts are unrelated.

The question arises of whether one may further restrict the choice of variables in any given interval without losing essential generality. Such restrictions may reduce the number of expressions that have to be evaluated. One may for example require that no two branches leaving a state may have the same name (variable). This would however preclude an encoder in which some input symbols cause changes in the encoder state without immediately selecting different signals. At this point, no further restrictions were placed on the choice of variables.

The assumptions and restrictions discussed so far usefully limit the number of variables that may be inserted into the likelihood ratio place-holders, as compared to the previous search, and therefore restrict the number of expressions that need to be evaluated.

### The search algorithm

The algorithm used for generating and evaluating all possible likelihood ratios derived from trellises or trees with given input alphabet size $M$ and observation period $N$, is described in detail in Appendix B.

The most restrictive factor limiting the performance of the search algorithm is the number of ordered partitions of multi-sets[3] of variables. It is necessary to give a brief outline of the algorithm here in order to explain why the number of ordered partitions is significant to the performance of the search.

The search algorithm generates all possible mathematically distinct abbreviated likelihood ratio expressions for a given $M$ and $N$. By 'mathematically distinct' is meant that one expression should not be obtainable from another

[3] A *multi-set* is a set with multiplicity, that is, a set in which some of the elements occur more than once. The normal definition of a set only allows distinct elements, with multiplicity of one. The order of the elements of both ordinary sets and multi-sets is not significant. The term 'multi-set' follows the usage by G.C. Rota. I thank Dr. Ken Hughes of the University of Cape Town for providing the information in this footnote.

by simple substitution of variables. This is not easy to achieve, because of the commutative properties of addition and multiplication in expressions such as (4.40): an expression such as $ab+cd$ is not mathematically distinct from $cd+ab$ or $ba+cd$, but all of these are distinct from $ad+cb$.

Multi-sets are used to hold the collection of LTVs in each symbol interval. The place-holders in the expression are 'filled in' recursively with LTVs one interval at a time. Within each interval $n$, $n = 1, 2, \ldots, N-1$, a multi-set of $B = 2M^n$ LTVs is formed, representing all the labels of the branches of a tree like the one in Figure 4.18. Because the size of the signal set, $m = 2M$, is generally smaller than the number of branches $B$ in one interval of the trellis, some or all of the LTVs from the signal set occur more than once in the multi-set.

Next, the branches have to be clustered together $M$ at a time by partitioning the multi-set into disjoint subsets called cells, each containing $M$ variables. For example, the ordered partitions of $\{a,b,c,d\}$ into cells of two elements are $\{\{a,b\},\{c,d\}\}$, $\{\{a,c\},\{b,d\}\}$, $\{\{a,d\},\{b,c\}\}$, $\{\{b,c\},\{a,d\}\}$, $\{\{b,d\},\{a,c\}\}$, and $\{\{c,d\},\{a,b\}\}$. Each cell is itself a multi-set. The order of the cells within each partition is significant, and therefore the partition of the original multi-set is an ordered partition.

To find all the ordered partitions of a multi-set, it is treated at first as if it were a set of unique elements. A list of all the possible partitions of the set is generated and stored. Because the multi-set has many duplicated elements, many of the partitions in the list will be identical. (This can be seen in the partition example above by setting $c = d$ for example.) The duplicate entries are removed from the list by sorting the list and identifying adjacent partitions which are the same.

Choosing different partitions from the list is equivalent to shuffling the LTVs within one interval in the expression. Each interval has its own list of partitions. The likelihood ratio expression itself is generated by recursively selecting a partition in each interval. In this way all possible expressions, hopefully all mathematically distinct, are formed by working through all the entries in each list.

| $M$ | $N$ | $B$ | $P_o$ |
|---|---|---|---|
| 2 | 2 | 4 | 6 |
| 2 | 3 | 8 | 2520 |
| 2 | 4 | 16 | 81729648000 |
| 3 | 2 | 6 | 20 |
| 3 | 3 | 18 | 137225088000 |
| 4 | 2 | 8 | 70 |
| 5 | 2 | 10 | 252 |
| 6 | 2 | 12 | 924 |
| 7 | 2 | 14 | 3432 |
| 8 | 2 | 16 | 12870 |
| 16 | 2 | 32 | 601080390 |

Table 4.10: Number of ordered partitions $P_o$ of a set of $B$ branch labels in the last interval of a code tree, as a function of the branching factor $M$ and the observation interval $N$.

### The number of ordered partitions

This subsection is devoted to a brief analysis of the number of ordered partitions that have to be generated and stored in the computer implementation of the search algorithm. This number will give an indication of both the amount of memory needed for storing all the partitions, and of the relative speed of the algorithm, as a function of $M$ and $N$. In this way, the values of $M$ and $N$ that may be used in a practical search may be determined.

The last interval $n = N - 1$ in the trellis, contains the most branches $B = 2M^{N-1}$, and will therefore have a multi-set with many more partitions than any other interval. The number of ordered partitions $P_o$ of a set of $B$ elements into $B/M$ cells of equal size $M$ (assuming that $B$ is an integer multiple of $M$) is [46]

$$P_o = \frac{B!}{(M!)^{B/M}} \tag{4.41}$$

Table 4.10 lists the number of partitions for some values of $M$ and $N$.

It is clear from the table that for the sets of cases $\{M = 2, N \geq 4\}$ and $\{M \geq 3, N \geq 3\}$, this algorithm requires more memory for storing the lists of partitions than is available on any known computer at present. The remaining cases, $\{M \leq 8, N = 2\}$ and $\{M = 2, N = 3\}$, can be done on a small 'personal' computer.

### Results

The complete table of results of the search for identities derived from realistic likelihood ratios is listed in Appendix C. It was only possible to search eight cases $\{2 \leq M \leq 8, N = 2\}$ and $\{M = 2, N = 3\}$ exhaustively, as determined in the previous subsection.

Although many identities are listed in Appendix C, some of them can be derived from one another by simple substitution. The search algorithm was therefore not perfect in its ability to generate only mathematically distinct likelihood ratios. Extracting the generic forms only, Appendix C is summarized here:

1. For $M = 2$ and $N = 2$, the MSK identity

$$\frac{a_1 + b_1}{\frac{1}{a_1} + \frac{1}{b_1}} = a_1 b_1 \tag{4.42}$$

   was found.

2. For $3 \leq M \leq 8$ and $N = 2$, the family of MSK-like identities

$$\frac{(M-k)a_1 + kb_1}{k\frac{1}{a_1} + (M-k)\frac{1}{b_1}} = a_1 b_1, \quad k = 1, 2, \ldots, M-1 \tag{4.43}$$

   was found. For example, with $M = 4$ and $k = 1$ we have

$$\frac{3a_1 + b_1}{\frac{1}{a_1} + 3\frac{1}{b_1}} = a_1 b_1$$

3. For $M = 2$ and $N = 3$, three types of identity were found: those which use information from all the symbols in the observation interval

$$\frac{a_1(a_2 + a_2) + b_1(a_2 + a_2)}{\frac{1}{a_1}\left(\frac{1}{b_2} + \frac{1}{b_2}\right) + \frac{1}{b_1}\left(\frac{1}{b_2} + \frac{1}{b_2}\right)} = a_1 b_1 a_2 b_2 \tag{4.44}$$

$$\frac{a_1(a_2 + a_2) + b_1(a_2 + b_2)}{\frac{1}{a_1}\left(\frac{1}{a_2} + \frac{1}{b_2}\right) + \frac{1}{b_1}\left(\frac{1}{b_2} + \frac{1}{b_2}\right)} = a_1 b_1 a_2 b_2 \tag{4.45}$$

$$\frac{a_1(a_2 + a_2) + b_1(b_2 + b_2)}{\frac{1}{a_1}\left(\frac{1}{a_2} + \frac{1}{a_2}\right) + \frac{1}{b_1}\left(\frac{1}{b_2} + \frac{1}{b_2}\right)} = a_1 b_1 a_2 b_2 \tag{4.46}$$

$$\frac{a_1(a_2 + b_2) + b_1(a_2 + b_2)}{\frac{1}{a_1}\left(\frac{1}{a_2} + \frac{1}{b_2}\right) + \frac{1}{b_1}\left(\frac{1}{a_2} + \frac{1}{b_2}\right)} = a_1 b_1 a_2 b_2 \tag{4.47}$$

those which drop information from the third interval ($n = 2$)

$$\frac{a_1 (w_2 + x_2) + b_1 (y_2 + z_2)}{\frac{1}{a_1} (y_2 + z_2) + \frac{1}{b_1} (w_2 + x_2)} = a_1 b_1 \tag{4.48}$$

where $w_2, x_2, y_2, z_2 \in \{a_2, b_2, \frac{1}{a_2}, \frac{1}{b_2}\}$, and those which drop information from the second interval ($n = 1$)

$$\frac{a_1 (a_2 + a_2) + b_1 (a_2 + a_2)}{a_1 \left(\frac{1}{b_2} + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{b_2} + \frac{1}{b_2}\right)} = a_2 b_2 \tag{4.49}$$

$$\frac{a_1 (a_2 + a_2) + b_1 (a_2 + b_2)}{a_1 \left(\frac{1}{b_2} + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right)} = a_2 b_2 \tag{4.50}$$

$$\frac{a_1 (a_2 + a_2) + b_1 (b_2 + b_2)}{a_1 \left(\frac{1}{b_2} + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2} + \frac{1}{a_2}\right)} = a_2 b_2 \tag{4.51}$$

$$\frac{a_1 (a_2 + b_2) + b_1 (a_2 + b_2)}{a_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right)} = a_2 b_2 \tag{4.52}$$

Even if the searches of point 2 above were only done for $2 \leq M \leq 8$, the family of identities (4.43) seems to apply to any $M \geq 2$ with $N = 2$. Thus (4.43) may be regarded as the more general form of the MSK identity (4.1), extended to any $M$.

The use and interpretation of all the identities found is discussed in the next chapter.

# Chapter 5

# Interpretation of the Search Results

The previous chapter concentrated on describing the various searches for simplified receivers and the results of the searches, without analysis or interpretation of the results. This chapter examines those results with the goal of emphasizing their pertinent properties. This lays the foundation for the mathematical analysis which follows in Chapter 6.

This project has used few of the standard techniques of trellis design and searching for 'best' modulation codes. Instead a new paradigm based directly on the likelihood function was proposed, with the aim of evaluating maximum likelihood receivers which have the potential for possessing the same order of complexity as the Massey receiver [48] for minimum shift keying (MSK).

The search for a simplified receiver for quadrature minimum shift keying (QMSK) pointed out the prevalence and significance of catastrophic trellises. They are re-examined in section 5.1.1, also because knowledge of their properties has wider application than the current context.

Section 5.2 presents an important conclusion of this chapter: Over two signalling intervals, and three intervals for binary input, no mathematical identities or likelihood ratios exist which may lead to a simplified receiver, except for the MSK identity and variations of it. This strongly suggests that the MSK identity is the only one that may be useful for synthesizing new simplifying likelihood ratios. For observation periods of three or more symbols, variations of the MSK identity appear, but the most promising ones always appear to be related to the MSK identity.

## 5.1 Interpreting the Results of the QMSK Trellis Search

It is significant that the direct search for a simplifying likelihood ratio for the QMSK modulation scheme did not yield a trellis which outperforms MSK. The definition of the QMSK channel encoder (CE) was quite loosely constrained, and yet the best simplified receivers found either had the same free Euclidean distance as MSK, or had catastrophic encoders. Some of the search space (for number of CE states $Z = 2$ and CE cycle length $J \leq 4$) was completely covered, and some of the rest of the space (for $Z = 2, J = 8$ and $Z = 4, J \leq 4$) was extensively probed.

It could be surmised that the reason for this negative result might be that the CPE constrains the search too much: it is possible that there is no intersection between the hypothetical set of simplified receivers and the set of receivers for CPM. The next step would be to drop the requirement of continuous phase and search more generally for simplifying likelihood ratios, as was done in section 4.3 for example.

Another hypothetical explanation for the lack of simplified receivers is the fact that the search was done with a strict definition of what constitutes a catastrophic scheme. It leaves open the possibility that some of the catastrophic trellises may have been useful. Rimoldi [68] has analysed partial response continuous phase modulation (CPM) schemes for a slightly different definition of 'catastrophic' and has come to the conclusion that the probability that a CPM scheme will be seriously affected by its catastrophic status is vanishingly small. The question of whether the same could be said of the catastrophic QMSK schemes is addressed below.

### 5.1.1 A Re-examination of Catastrophic Schemes

Catastrophic coded modulation schemes were defined in section 3.5 and illustrated for the case of QMSK in section 4.2.4. This section re-examines catastrophic codes with the intention of loosening the constraints of searches for simplified receivers, by allowing catastrophic schemes or at least some types of catastrophic schemes.

It turned out to be unnecessary to redo the extensive searches described in Chapter 4 because much tighter bounds on possible simplified receivers have been derived, as described in Chapter 6. A further discussion of catastrophic schemes is however included here because it has a wider bearing than the current context. The CPM schemes discussed by Rimoldi [68], for example, may be analysed and classified in the same way as the schemes considered here.

#### The average length of error events

It is undesirable for a scheme to be catastrophic because it is possible in principle for a single decoding error, perhaps coupled with an unfortunate sequence of symbols, to cause the signal path through the trellis to follow a route parallel to the one decoded by the receiver. The receiver may have no information (in the form of Euclidean distance metrics) to aid its recovery from what could be a long run of decoding errors.

The question now arises of whether there are differences between catastrophic schemes which may make some preferable to others. The difference of interest is the average run length of receiver errors. If some catastrophic schemes exhibit short average error runs, they may be preferable to schemes with long error events.

A run of errors occurs when a decoding error due to noise in the channel causes the receiver to assume a different path through the trellis than the actual transmitted sequence. The receiver then continues to report incorrect decoded states until either

- another decoding error due to channel noise occurs which remerges the

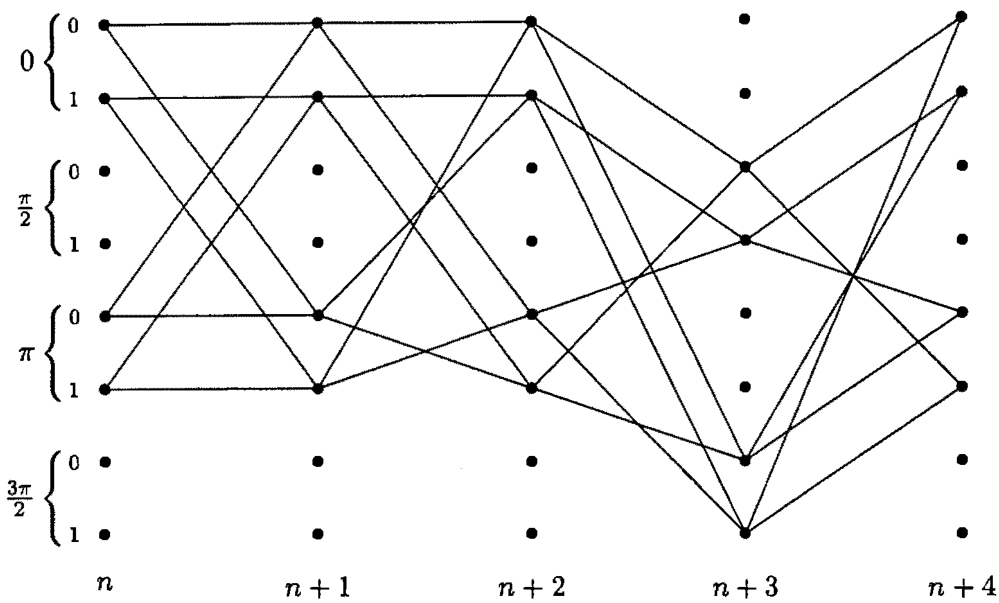



Figure 5.1: The catastrophic QMSK trellis of Figure 4.13, reproduced here for ease of reference. Pairs of paths which share the same phase state in every interval have zero Euclidean distance between them.

receiver state with the transmitter state; or

- the trellis itself merges at least one pair of zero-distance loops. (Zero-distance loops were defined in section 3.5 as pairs of distinct paths which wrap around the trellis indefinitely with no Euclidean distance between them.)

The first scenario is clearly undesirable, since long or even asymptotically infinitely long error events can be anticipated when the error rate is low (at high signal-to-noise ratio). The catastrophic QMSK trellis of Figure 4.13 (reproduced here as Figure 5.1) is an example of such a scheme. The second scenario has more potential: if the receiver can of its own accord resynchronize with the transmitter despite the occasional decoding error, a practical scheme may be possible.

It is worth remembering that a scheme is considered catastrophic if even a single pair of zero-distance loops can be found in its trellis. This may not be the best criterion to use for discarding candidate trellises if it is surmised that some catastrophic schemes may none the less be practical. Instead of requiring that CPFSK schemes not be catastrophic, it may be sufficient to require merely

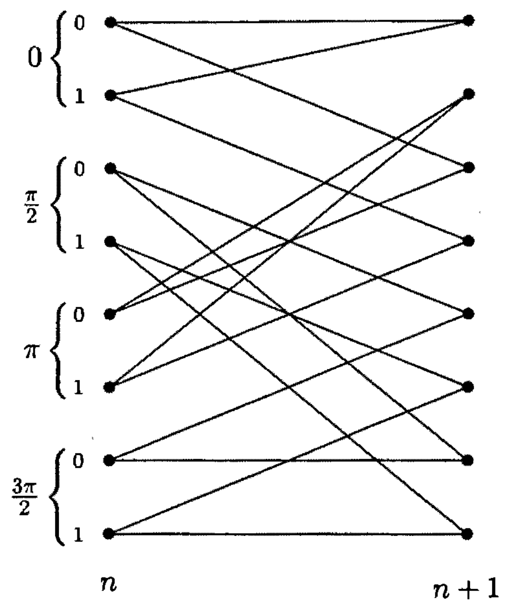



Figure 5.2: The trellis of a hypothetical limited catastrophic scheme. There are several zero-distance loops, but the upper branches from phase states 0 and $\pi$ merge.

that it should be possible to find two merging branches which start in the same phase state but different CE states. Then, assuming random data and a fully interconnected trellis, any pair of zero-distance paths will eventually merge at this point. Figure 5.2 illustrates the concept.

I call a scheme with at least one such merging a *limited catastrophic scheme*, or simply *limited*, because the average length of error events is limited by the trellis at high signal-to-noise ratio. The more such mergings occur in a catastrophic trellis, the shorter the average error event will be, assuming random data. The average length of error events of the scheme in Figure 5.2 is, for example,

$$\frac{1}{P(\text{merge})} = \frac{1}{2/8} = 4 \tag{5.1}$$

Examination of the trellis in Figure 5.1 shows that no pairs of branches have this characteristic, and hence it is not limited.

### Revisiting the QMSK trellis searches

In the searches done on QMSK trellises, described in section 4.2.4, none of the catastrophic trellises found in the exhaustive searches ($Z = 2, J = 2$ and 4) were limited. In other words, even when the criterion which excluded catastrophic trellises from consideration was relaxed to exclude only trellises which are not limited, no simplified QMSK receiver was found.

The QMSK trellis searches were done with a minimum free squared Euclidean distance threshold of $d^2_{\text{free}} = 3$, a modest increase over the performance of MSK. When this limit was removed and the search redone for the cases of two CE states ($Z = 2$) and cycle lengths of two and four ($J = 2$ and 4), many trellises were found, but the best free Euclidean distance was $d^2_{\text{free}} = 2$, the same as before.

### Performance of limited catastrophic schemes

To obtain a limited run length of errors in a catastrophic trellis, it is necessary to have at least one pair of signals (branches) which start from the same point in Euclidean signal space (for example in the same phase state) and which merge in a common transmitter state at the end of the symbol interval. The Euclidean distance between these two branches must of necessity be zero. This means that all limited catastrophic schemes are likely to have a performance penalty of at least one zero-distance interval within their constraint lengths.

In the case of CPM in general, the theoretical maximum achievable free squared Euclidean distance is

$$d^2_{\text{free}} < 2 \times (\text{free distance}) \tag{5.2}$$

which is reduced in the case of limited catastrophic schemes to

$$d^2_{\text{limited}} < 2 \times (\text{free distance} - 1) \tag{5.3}$$

For CPFSK, the theoretical maximum free squared Euclidean distance is

$$d^2_{\text{free}} \leq 2 \times (\text{free distance} - 1) \tag{5.4}$$

which is reduced to

$$d^2_{\text{limited}} \leq 2 \times (\text{free distance} - 2) \tag{5.5}$$

in the case of limited catastrophic schemes. Where previously we might have hoped to find a QMSK scheme with $d^2_{\text{free}} = 3$ or 4 over a constraint length of three intervals, the observation period now has to be increased to four intervals to obtain the same performance. The number of CE states has to increase accordingly from $Z = 2$ to $Z = 4$, which causes a big increase in the size of the search space of possible trellises.

This explains why none of the limited QMSK schemes in the exhaustively searched $Z = 2$ space achieved $d^2_{\text{free}} > 2$.

## 5.2 Interpreting the Identities Found

### 5.2.1 The MSK identity

The structure of the optimal receiver of MSK, when expressed as a likelihood ratio

$$\Lambda = \frac{a_0\,(a_1 + b_1)}{b_0\left(\frac{1}{a_1} + \frac{1}{b_1}\right)} \tag{5.6}$$

may be applied to any other coded equal-energy modulation with the same likelihood ratio. All that is required is to match the likelihood transform variables in (5.6) with an appropriate signal set. In this case, that modulation has a binary input (because there are two terms in each of the numerator and denominator of (5.6)), and it has a half-rate channel encoder (because it has a signal set of $m = 4$ signals). If the four signals are bi-orthogonal as in the case of MSK, it will have the same bit error rate as MSK. The only change to the MSK receiver is to match the pair of matched filters at the receiver input to the signal set.

Consider for example a differential offset QPSK with the 'channel encoder' transfer function $[1, D]$, the same as that of differential (Type I) MSK. This scheme was explicitly mentioned by Massey [48] as having the identical optimal receiver structure as MSK. The signal set is given in Table 5.1. Table 1.1 gives the MSK signal set for comparison. Apart from the signal mapping, the transmitter is identical to the MSK transmitter of Figure 1.8.

Such a bi-orthogonal $m = 4$ modulation scheme has the same free Euclidean

| $v_n$ | $v_{n+1}$ | $s_{vu}(\tau)$ |
|---|---|---|
| 0 | 0 | $s_{00}(\tau) = \sqrt{\frac{2E}{T}} \cos \omega_0 \tau$ |
| 0 | 1 | $s_{01}(\tau) = \sqrt{\frac{2E}{T}} \cos(\omega_0 \tau + \frac{\pi}{2})$ |
| 1 | 0 | $s_{11}(\tau) = -s_{01}(\tau)$ |
| 1 | 1 | $s_{10}(\tau) = -s_{00}(\tau)$ |

Table 5.1: The signal set of differential offset QPSK. $v_{n+1}$ is the input data at time $t = nT$, and $v_n$ is the corresponding encoder state. $\tau = t - nT$.

distance $d^2_{\text{free}} = 2$ as MSK, and hence the same bit error rate. Its null-to-null bandwidth (normalized to binary $M = 2$ input, same as BPSK) is 33% larger than MSK's, and it would therefore not normally be considered for practical implementation over MSK.

Similarly, any binary FSK, with its signal set extended to include the inverse of each signal, is compatible with an MSK receiver in this way. Only the orthogonal schemes will have optimal error performance though. FSK is orthogonal for modulation indices $h = k/2$, $k = 1, 2, \ldots$

This exhausts most of the possibilities for direct application of the MSK likelihood ratio (5.6), since PSK and FSK are the two major classes of equal-energy modulations. The remaining options are variations of these basic schemes, and include various pulse shaping techniques for controlling the power spectrum of the modulation, or attempting to use the MSK identity as a building block in a hierarchical channel encoder with the aim of simplifying its Viterbi decoder in the receiver. Such possibilities are not further explored here.

### 5.2.2 The family of MSK identities

The MSK identity is extended to any input alphabet size by the family of identities (4.43):

$$\frac{(M-k)a_1 + kb_1}{k\frac{1}{a_1} + (M-k)\frac{1}{b_1}} = a_1 b_1, \quad k = 1, 2, \ldots, M-1 \tag{5.7}$$

Their left hand sides can all be interpreted directly as abbreviated likelihood ratios, or as likelihood ratios if variables are added for the first interval. The

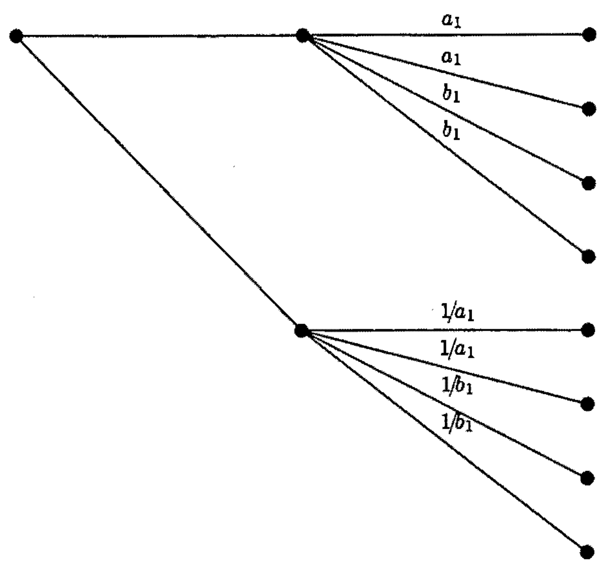



Figure 5.3: The tree resulting from the MSK-like identity (5.8).

right hand side gives the same simplified receiver for all of them.

It is interesting to note that even when $M > 2$, only four signals, $a_1$, $b_1$, $1/a_1$ and $1/b_1$, are used in the interval $n = 1$. Some branches from the same state will transmit the same signal irrespective of the input symbol, causing only a change in the state of the encoder.

For example, for $M = 4$ and $N = 2$, the identity

$$\frac{2a_1 + 2b_1}{2\frac{1}{a_1} + 2\frac{1}{b_1}} = a_1 b_1 \qquad (5.8)$$

from this family results in the tree in Figure 5.3 when it is interpreted directly as an abbreviated likelihood ratio.

The family of MSK-like identities (4.43) is the only one found for two intervals and $2 \leq M \leq 8$. These cases were exhaustively searched, and it can be stated with certainty that no other identities exist. This is most likely also true for $N = 2$ and $M > 8$.

The fact that none of the identities of the MSK family extend the signal set beyond the MSK signal set, limits its usefulness. This result further suggests that no other identities, which do extend the signal set, exist for an observation period of two symbol intervals.

### 5.2.3 The identities over three symbols

The results of the identity search over three intervals for binary input were summarized in (4.44) to (4.52). The complete results are listed in Appendix C.

The identities (4.44) – (4.52) are of three types:

- Those which simplify into a term containing likelihood transform variables from all the intervals in question;
- those which do not have any variables from the last interval in the simplified term; and
- those which do not have variables from the middle interval.

The latter two types can occur because some of the identities correspond to trellises that are labelled in such a way that the maximum likelihood detector cannot use any information from certain intervals to decide which of the two branches in the first interval were followed.

For example, the left hand side of the identity

$$\frac{a_1\left(a_2+b_2\right)+b_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)}{\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(a_2+b_2\right)}=a_1 b_1 \tag{5.9}$$

may seem a promising candidate of the second type (4.48). If it is interpreted as an abbreviated likelihood ratio, its corresponding tree is shown in Figure 5.4(a), and possible trellises in (b) and (c). The normalized squared Euclidean distance between all possible sequences from the starting node of the trellis in (b) is at least $d^2(s,s')=5$ for bi-orthogonal signalling. And yet the third interval contributes no information to the decision on which of the two branches in the first interval was followed.

Redrawing the trellis of (5.9) as in Figure 5.4(c), shows another reason why the third interval is dropped: as a two-state code, it is equivalent to MSK. MSK's optimal receiver has already been shown to make its optimal decision in only two symbol intervals in section 1.5.2 and section 3.3.2. Even if an attempt is made to redraw the trellis as a four-state trellis, the resulting scheme can be

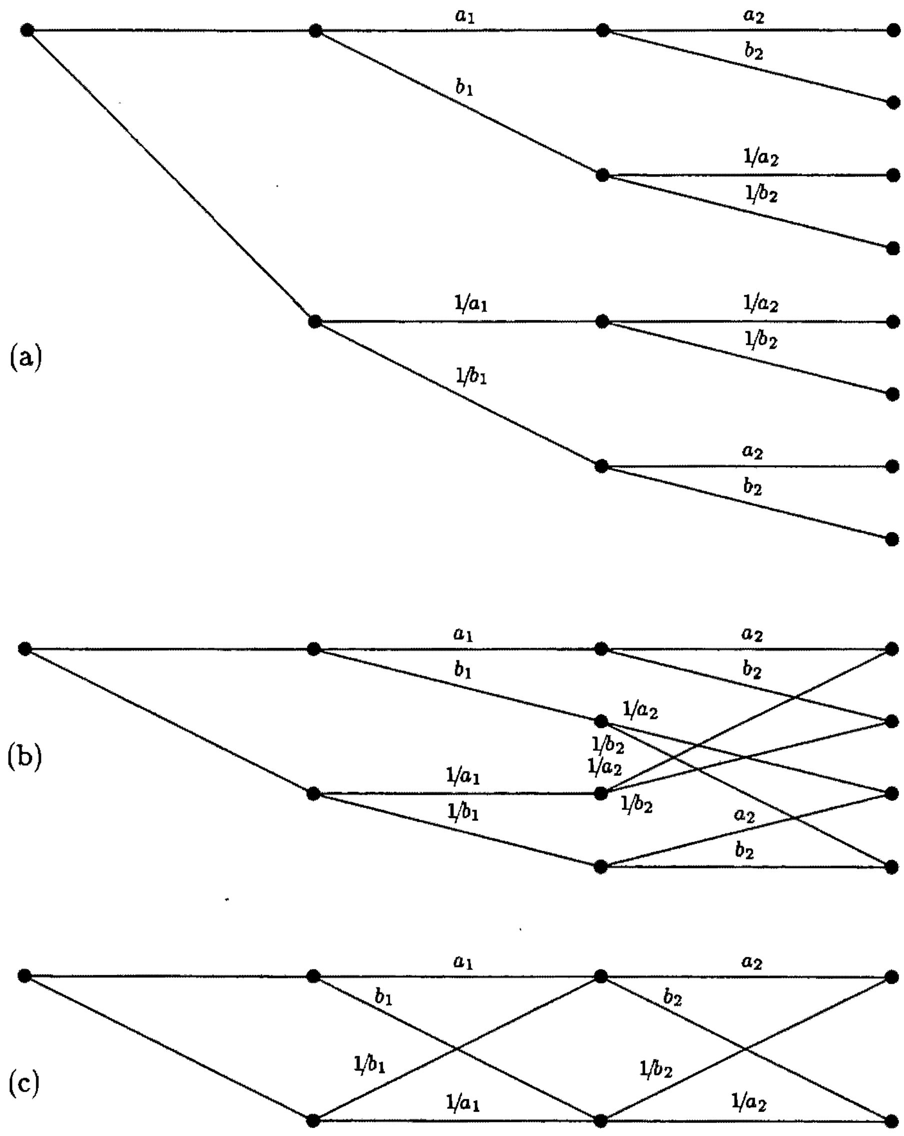



Figure 5.4: The tree (a) and a possible trellis (b) from (5.9), which drops all information from the last symbol of the observation period. The two-state trellis (c) turns out to be the MSK trellis.

shown in every case to reduce to pure MSK with a redundant delay element in its encoder.

It appears that the simplification of the likelihood ratio and the existence of a simplified receiver has sometimes been missed in the published literature. Fig. 3 of a paper by Anderson and de Buda [4] shows a state trellis for a convolutionally encoded QPSK, reproduced below as Figure 5.5. The likelihood

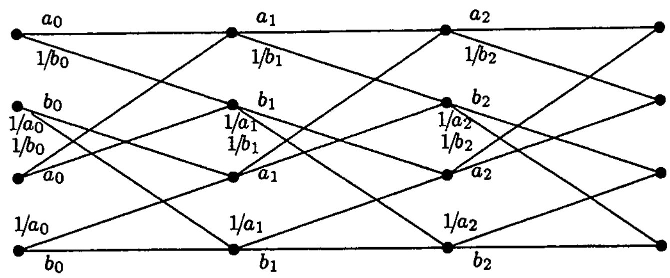



Figure 5.5: Trellis diagram for convolutionally coded QPSK, after Fig. 3 of [4].

ratio in the top state follows from the diagram:

$$\begin{aligned}\Lambda_{00} &= \frac{a_0\left[a_1\left(a_2+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+b_2\right)\right]}{\frac{1}{b_0}\left[\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)+b_1\left(a_2+\frac{1}{b_2}\right)\right]} \\ &= \frac{a_0 b_0 a_1}{b_1} \qquad (5.10)\end{aligned}$$

which is a version of (4.48). The same likelihood ratio can be generated from the other states. If an observation period of three symbol intervals s is used, the last symbol in the window is not used and the error performance can be no better than that of MSK.

Turning now to the identities of the first type, there are four of them summarized in (4.44) to (4.47). These identities are the most interesting because they use all possible information available during the observation interval that they cover. The last one, (4.47), reduces to

$$\begin{aligned}\frac{a_1(a_2+b_2)+b_1(a_2+b_2)}{\frac{1}{a_1}(\frac{1}{a_2}+\frac{1}{b_2})+\frac{1}{b_1}(\frac{1}{a_2}+\frac{1}{b_2})} &= \frac{a_1+b_1}{\frac{1}{a_1}+\frac{1}{b_1}}\cdot\frac{a_2+b_2}{\frac{1}{a_2}+\frac{1}{b_2}} \\ &= a_1 b_1 a_2 b_2 \qquad (5.11)\end{aligned}$$

This is simply the product of two MSK identities, one for each interval. The corresponding tree and a possible four-state trellis are shown in Figure 5.6.

All possible pairs of sequences from the starting state differ by at least $d^2(s,s') = 5$ for bi-orthogonal signalling as before, a gain of about 4 dB over MSK. But this promise of good performance is based on analysis from a single starting state. The problem is that the trellis shown in Figure 5.6(b) is not regular, and one therefore has to show that its good properties are maintained

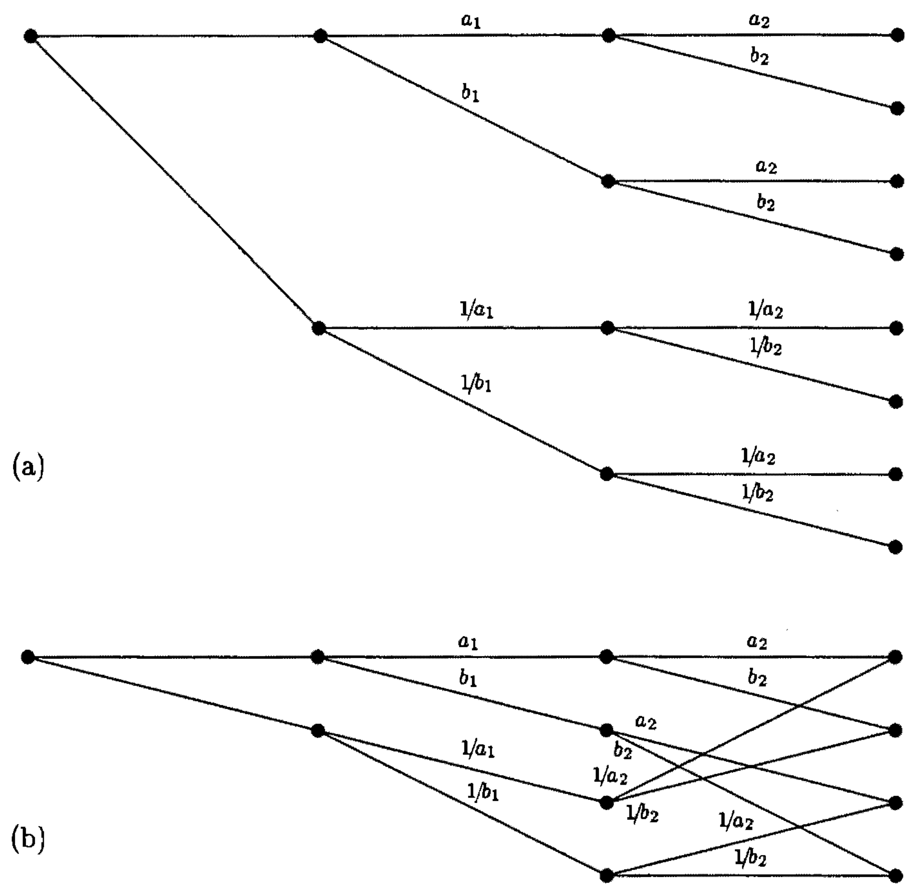



Figure 5.6: The tree (a) and a possible trellis (b) derived from an apparently optimal identity (5.11), which uses information from all the symbols in the observation period.

over time shifts and different starting states. In fact they are not maintained.

Keeping the same scheme of Figure 5.6(a) and shifting one symbol interval into the future (see Figure 5.7), it is not possible to label the branches of the next symbol interval in such a way that the likelihood ratio continues to simplify, *and* the simplified term continues to use variables from all the intervals. Information from the second observation interval is inevitably dropped during the simplification. This happens because the second interval $n = 2$ contains a repetition of the labelling of the branches from each state: $a_2, b_2$ and $a_2, b_2$. The only simplification possible is of the third type of identity (4.49) to (4.52) mentioned above, and that type does not use information from the middle interval.

It therefore does not seem to be possible to use (5.11) in such a way that

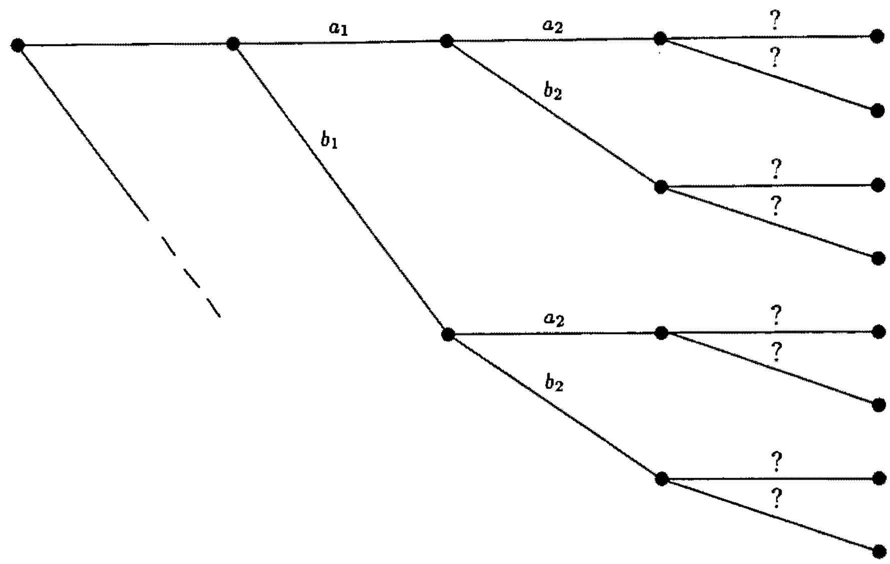



Figure 5.7: The degenerate tree that results after the tree of Figure 5.6 is continued from state 0, one interval into the future.

its likelihood ratios consistently use information from all three symbols in the observation period, from any starting state and in every symbol interval.

Looking now at some of the other identities of the first type, (4.44) and (4.46) both infer a transmitter which transmits a fixed symbol regardless of the input data in the third interval. The third symbol effectively combines with either the first or second symbol into a single extended symbol, which does improve the error performance of the scheme at the cost of halving the information rate.

For example, (4.46) may be written as

$$\begin{aligned} \frac{a_1\left(a_2+a_2\right)+b_1\left(b_2+b_2\right)}{\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)} &= \frac{a_1a_2+b_1b_2}{\frac{1}{a_1a_2}+\frac{1}{b_1b_2}} \\ &= (a_1a_2)(b_1b_2) \qquad (5.12) \end{aligned}$$

which clearly shows how the last two symbols are combined and detected together over two intervals instead of one.

The remaining identity of the first type which uses information from all three intervals, is (4.45), which is merely an odd hybrid of the others. It combines symbols in some intervals from some states without gaining any further advantage over the other schemes.

These results show that no identity exists which can be interpreted directly as a simplifying abbreviated likelihood ratio which optimally detects a binary-input coded modulation over three intervals, nor could any be found for multi-level ($M > 2$) inputs.

### 5.2.4 Extending identities for longer observation periods

Since it appears to be impossible to detect binary-input coded modulation optimally over more than two symbols with a simplified receiver, one may ask whether there exists some suboptimal receiver in that case. Such a receiver would still have the almost trivial simplicity of a simplified receiver and operate over many symbol intervals, which may be an attractive option in fading channels.

Returning for the moment to the promising identity (5.11), it was seen to be the product of two MSK identities in two consecutive intervals. It at first appears that (5.11) may be used to extend the observation period and increase the number of states indefinitely, by multiplying with the MSK identity for each additional interval needed. It seems that the free Euclidean distance increases linearly with each interval added. Such claims can obviously not be sustained, because there is no limit placed on the free Euclidean distance, while the modulation is assumed to be equal-energy.

For example, over four symbols a free Euclidean distance of $d^2_{\text{free}} = 7$ seems achievable. An abbreviated likelihood ratio over four symbol intervals may be constructed from (5.11) as follows:

$$\begin{aligned}
\Lambda_0 &= \frac{a_1+b_1}{\frac{1}{a_1}+\frac{1}{b_1}} \cdot \frac{a_2+b_2}{\frac{1}{a_2}+\frac{1}{b_2}} \cdot \frac{a_3+b_3}{\frac{1}{a_3}+\frac{1}{b_3}} \\
&= \frac{a_1\left[a_2\left(a_3+b_3\right)+b_2\left(a_3+b_3\right)\right]+b_1\left[a_2\left(a_3+b_3\right)+b_2\left(a_3+b_3\right)\right]}{\frac{1}{a_1}\left[\frac{1}{a_2}\left(\frac{1}{a_3}+\frac{1}{b_3}\right)+\frac{1}{b_2}\left(\frac{1}{a_3}+\frac{1}{b_3}\right)\right]+\frac{1}{b_1}\left[\frac{1}{a_2}\left(\frac{1}{a_3}+\frac{1}{b_3}\right)+\frac{1}{b_2}\left(\frac{1}{a_3}+\frac{1}{b_3}\right)\right]} \\
&= a_1b_1a_2b_2a_3b_3 \qquad (5.13)
\end{aligned}$$

where the subscript of $\Lambda_0$ now refers to the first interval $n = 0$. Figure 5.8 shows the associated tree.

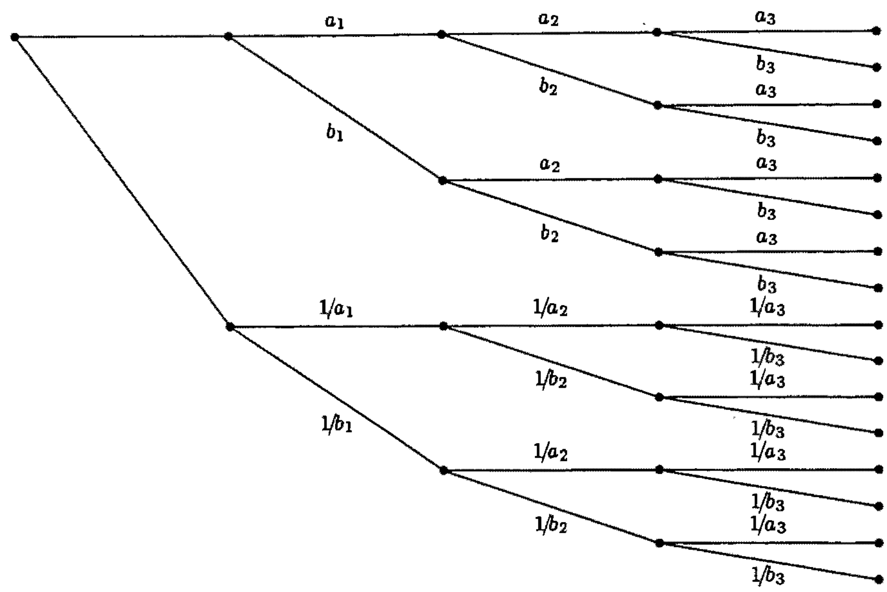



Figure 5.8: The tree that results from the repeated application of the MSK identity over four symbol intervals.

Moving on to the next symbol interval and assuming that a fifth interval has been added by multiplying again by the MSK identity, the following abbreviated likelihood ratio from state 0, $n = 1$ results:

$$\begin{aligned} \Lambda_1 &= \frac{a_2\left[a_3\left(a_4+b_4\right)+b_3\left(a_4+b_4\right)\right]+b_2\left[a_3\left(a_4+b_4\right)+b_3\left(a_4+b_4\right)\right]}{a_2\left[a_3\left(a_4+b_4\right)+b_3\left(a_4+b_4\right)\right]+b_2\left[a_3\left(a_4+b_4\right)+b_3\left(a_4+b_4\right)\right]} \\ &= 1 \end{aligned} \tag{5.14}$$

This decision therefore becomes completely degenerate, and only the first interval $n = 1$ (not shown here in the abbreviated likelihood ratio) may be used for detection.

It is however possible to design likelihood ratios over more than two intervals from some of the other identities, such that the simplified receiver uses information from the first and last symbols in the observation window. One proceeds from an identity like (4.52):

$$\begin{aligned} \frac{a_1\left(a_2+b_2\right)+b_1\left(a_2+b_2\right)}{a_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)} &= \frac{a_1+b_1}{a_1+b_1}\cdot\frac{a_2+b_2}{\frac{1}{a_2}+\frac{1}{b_2}} \\ &= a_2 b_2 \end{aligned} \tag{5.15}$$

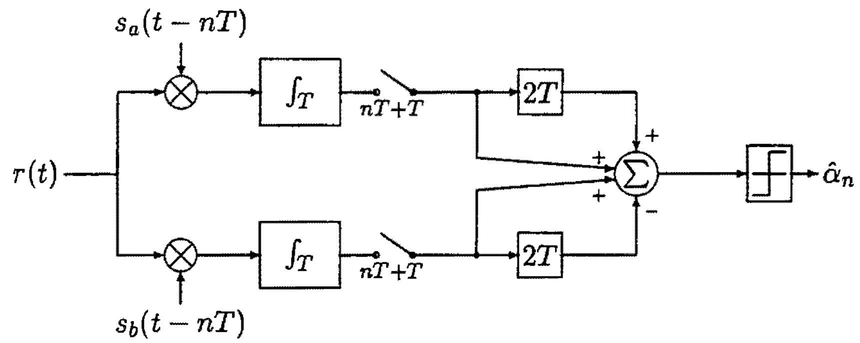



Figure 5.9: Block diagram of the simplified receiver for (5.16), which only uses information from the first and last symbols out of three.

to construct a likelihood ratio over three symbols

$$\begin{aligned} \Lambda &= \frac{a_0 \left[a_1 \left(a_2 + b_2\right) + b_1 \left(a_2 + b_2\right)\right]}{b_0 \left[a_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right)\right]} \\ &= \frac{a_0 a_2 b_2}{b_0} \end{aligned} \tag{5.16}$$

The simplified term in the last line represents the simplified receiver, which uses only the first and last received symbols in the observation window of three symbol intervals. Its block diagram is shown in Figure 5.9. A regular trellis can be constructed from (5.16) as shown in Figure 5.10. With bi-orthogonal signals, it has a free Euclidean distance of $d^2_{\text{free}} = 2$, the same as MSK.

This scheme has no coding gain over MSK, and would not normally be considered for practical use. It may however be usefully extended to longer observation periods $N \geq 3$ to provide time diversity in a fading channel, as

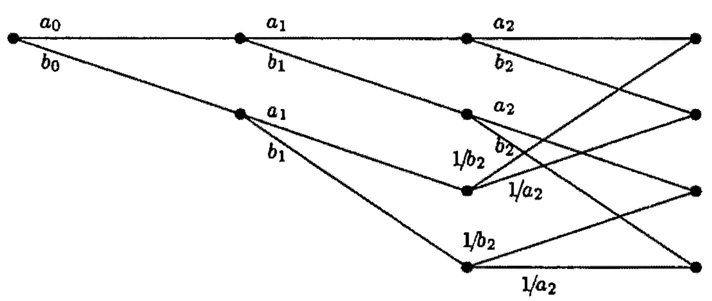



Figure 5.10: A regular trellis derived from (5.16) which uses two out of the three symbols in its observation period.

follows:

$$\begin{aligned}\Lambda &= \frac{a_0}{b_0} \cdot \prod_{n=1}^{N-2} \frac{a_n + b_n}{a_n + b_n} \cdot \frac{a_{N-1} + b_{N-1}}{\frac{1}{a_{N-1}} + \frac{1}{b_{N-1}}} \\ &= \frac{a_0 a_{N-1} b_{N-1}}{b_0} \end{aligned} \tag{5.17}$$

This still only uses information from the first and the last symbols in the observation window, but these may be spaced as far apart in time as desired. The optimal receiver is identical to the one in Figure 5.9 with the $2T$ delays replaced by $(N-1)T$.

# Chapter 6

# A Formal Analysis of Simplified Receivers

This chapter exploits the information gained from the searches over code space and identity space to develop formal proofs of the main results of the thesis: likelihood ratios are constrained by specific bounds on their structure in order to achieve simplification, and these bounds lead to the restriction of the performance of simplified receivers to that of MSK's maximum likelihood receiver.

The chapter starts in section 6.1 by summarizing the identities found in Chapter 4. These lead to a construction procedure for arbitrary simplifying likelihood ratios. The construction procedure is validated in section 6.2 (Theorem 1), which establishes rigorous bounds on the performance of simplified receivers.

In summary, simplified receivers which use information from all the symbols in their observation period are optimal only

1. for antipodal bi-orthogonal signal sets with at most four signals (Theorem 2);

2. for binary modulations (Theorem 3); and

3. for observation periods of no more than two symbol intervals (Theorem 4 and Corollary 1).

If it is not required that information from all the symbols in the observation period is used, the last condition (observation period of two intervals) may be relaxed. However, the number of symbol intervals that a simplified receiver can use for its decoding decisions remains limited to two, irrespective of the length of the observation period. The additional symbol intervals in the observation period (apart from the two which contirbute information to the decision) appear as neutral 'padding' in the simplified likelihood ratio.

## 6.1 Generalizing the Found Identities

The search for mathematical identities which may be useful in the simplification of likelihood ratios, described in section 4.3.3, resulted in the family of MSK identities (4.43) and identities over an observation period of three symbol intervals (4.44) to (4.52). This section uses these identities as a starting point to identify a general procedure for constructing likelihood ratios of any size which are guaranteed to simplify.

For reference, the found identities are summarized below. They have been slightly reorganized to make their generalization easier to see. They were: the family of MSK-like identities

$$\frac{(M-k)a_1 + kb_1}{(M-k)\frac{1}{b_1} + k\frac{1}{a_1}} = a_1 b_1, \quad k = 1, 2, \ldots, M-1 \tag{6.1}$$

identities which use information from all the symbols in the observation interval

$$\frac{a_1\left(a_2+a_2\right)+b_1\left(a_2+a_2\right)}{\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)} \quad = \quad a_1b_1a_2b_2 \tag{6.2}$$

$$\frac{a_1\left(a_2+a_2\right)+b_1\left(a_2+b_2\right)}{\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{a_2}\right)} \quad = \quad a_1b_1a_2b_2 \tag{6.3}$$

$$\frac{a_1\left(a_2+a_2\right)+b_1\left(b_2+b_2\right)}{\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)} \quad = \quad a_1b_1a_2b_2 \tag{6.4}$$

$$\frac{a_1\left(a_2+b_2\right)+b_1\left(a_2+b_2\right)}{\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{a_2}\right)+\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{a_2}\right)} \quad = \quad a_1b_1a_2b_2 \tag{6.5}$$

those which drop information from the third interval

$$\frac{a_1\left(w_2+x_2\right)+b_1\left(y_2+z_2\right)}{\frac{1}{b_1}\left(w_2+x_2\right)+\frac{1}{a_1}\left(y_2+z_2\right)} = a_1b_1 \tag{6.6}$$

where $w_2, x_2, y_2, z_2 \in \{a_2, b_2, \frac{1}{a_2}, \frac{1}{b_2}\}$, and those which drop information from the second interval

$$\frac{a_1(a_2+a_2)+b_1(a_2+a_2)}{a_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)} = a_2b_2 \tag{6.7}$$

$$\frac{a_1(a_2+a_2)+b_1(a_2+b_2)}{a_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{b_2}+\frac{1}{a_2}\right)} = a_2b_2 \tag{6.8}$$

$$\frac{a_1(a_2+a_2)+b_1(b_2+b_2)}{a_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)} = a_2b_2 \tag{6.9}$$

$$\frac{a_1(a_2+b_2)+b_1(a_2+b_2)}{a_1\left(\frac{1}{b_2}+\frac{1}{a_2}\right)+b_1\left(\frac{1}{b_2}+\frac{1}{a_2}\right)} = a_2b_2 \tag{6.10}$$

The discussion of search results in Chapter 5 highlighted several interesting, and sometimes quirky, points:

- No simplified noncatastrophic receivers were found for QMSK which exceeded $d^2_{\text{free}} = 2$.
- Those QMSK trellises which did have $d^2_{\text{free}} > 2$ were catastrophic and contributed information to the simplified likelihood ratio from only two signalling intervals.
- Over an observation period of two symbol intervals, only identities similar to the MSK identity were found.
- Over longer observation periods, some identities used information from more than two symbol intervals, but in every case this characteristic could either not be maintained over time shifts of the associated likelihood ratio, or else signals were combined in order to reduce the effective information rate.

These qualities will now be clarified in a unified description of the necessary limits which must be imposed on a likelihood ratio in order to ensure its simplification. These limits are embodied in a generalized construction procedure for simplifying likelihood ratios.

The fact that the searches which produced the identities were exhaustive, is important for the derivation of the construction procedure. The procedure would be of most use if it is capable of generating all the above identities (subject to the specified limits on input alphabet size $M$ and observation period $N$), and it generates no other expressions than these.

### 6.1.1 Construction procedure for arbitrary simplifying likelihood ratios

All trellis-based likelihood ratios are in the form of (4.28):

$$\Lambda = \frac{a(c_1(e_1(\cdots) + e_2(\cdots) + \cdots) + c_2(f_1(\cdots) + f_2(\cdots) + \cdots) + \cdots)}{b(d_1(g_1(\cdots) + g_2(\cdots) + \cdots) + d_2(h_1(\cdots) + h_2(\cdots) + \cdots) + \cdots)} \quad (6.11)$$

where the variables are place-holders for positive real likelihood transform variables (LTVs). The place-holders $a$ and $b$ are filled with LTVs from the first symbol interval, $c_1, c_2, \ldots$ and $d_1, d_2, \ldots$ with LTVs from the second interval, $e_1, e_2, \ldots$, $f_1, f_2, \ldots$, $g_1, g_2, \ldots$, and $h_1, h_2, \ldots$ from the third and so on. LTVs from different intervals are statistically independent and cannot have any direct functional relationship. LTVs from the same interval may be related either by being identical or reciprocal.

I now state the construction procedure for an arbitrary simplifying likelihood ratio over an observation period of $N$ intervals without derivation, leaving its justification for section 6.2:

1. Choose an observation period $N$ and a size of input alphabet $M_n$ for each interval after the first, $n = 1, 2, \ldots, N-1$. The input alphabet size $M_0$ in the first interval $n = 0$ is irrelevant. (The input alphabet size is usually invariant in modulation schemes, but need not be for likelihood ratio simplification.)

2. For each of the $N$ symbol intervals, choose a single term containing two LTVs, such as $ab$, $a/b$, $1/(ab)$, $aa = a^2$, or $1/(aa) = 1/a^2$. Call this term the interval's *contribution* to the decision. The two LTVs may be the same, such as $a$ and $a$. The contribution may also be chosen to be $1 = a/a = b/b = \cdots$ (In the current notation, the LTV names are re-used in successive intervals, with subscripts to distinguish the interval.)

3. For each interval, starting with the first one and doing them in sequential

order, choose a pair of place-holders with one place-holder from the numerator and one from the denominator. (For example, in the second interval one may pair $c_1$ with $d_1$ in (6.11).) In the third and later intervals, place-holders may only be paired if they share the same paired place-holders in the previous interval. (For example, if $c_1$ above was paired with $d_1$, then $e_1$ may be paired with any $g_k$, but not with any $h_k$.)

4. For each two place-holders so paired, split their associated interval's contribution between them. Each place-holder gets one of the contribution's LTVs or its reciprocal so that the ratio of LTVs equals the contribution. (For example if the second interval's contribution is $ab$ let $c_1 = a$ and $d_1 = 1/b$ so that $c_1/d_1 = a/(1/b) = ab$). Do this for all $\prod_{k=1}^{n} M_k$ pairs of place-holders in each interval after the first, $n = 1, 2, \ldots, N-1$. The first interval has only a single pair of place-holders.

5. The simplified likelihood ratio is the product of the $N$ intervals' contributions.

For future reference, I define a *contribution* as follows:

**Definition 6** *If a likelihood ratio simplifies, the* contribution *of a symbol interval in the observation period is the common ratio of the LTVs from that symbol interval which appears in the simplified likelihood ratio.*

An example, illustrated in Figure 6.1, clarifies the procedure:

1. Choose an observation period of $N = 3$ symbol intervals and an input alphabet size of $M = 4$ (any other choices would do as well).

2. For the first interval, choose $a_0/b_0$ as its contribution; for the second interval, choose $a_1 b_1$; and for the third choose $1 = a_2/a_2 = b_2/b_2$. The choices of LTVs are arbitrary since no signal set has been assigned as yet: any signal may be associated with any LTV.

3. Decide to pair each place-holder in the numerator with the one directly below it in the denominator.

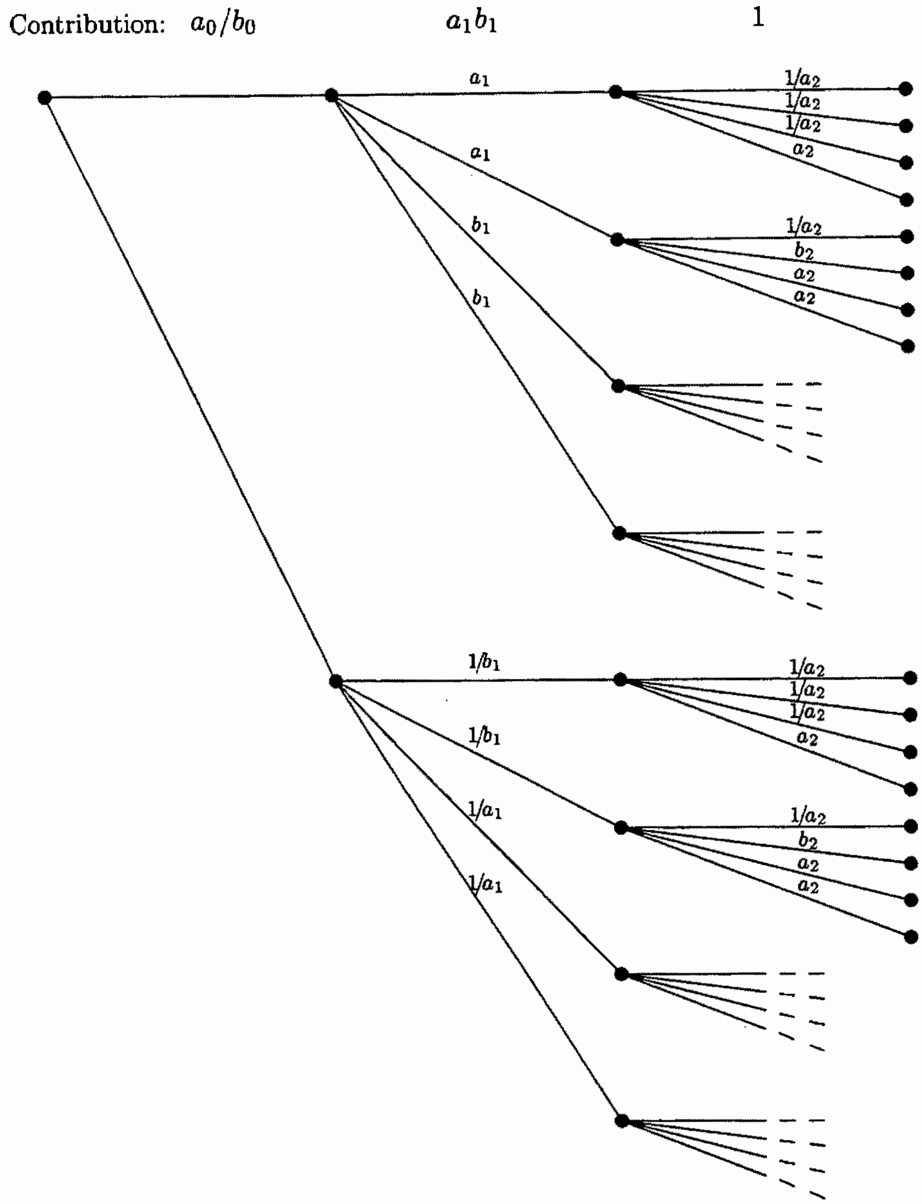



Figure 6.1: Tree illustrating the construction procedure.

4. Start the construction in the first interval by splitting its contribution $a_0/b_0$:

$$\Lambda = \frac{a_0[\cdots]}{b_0[\cdots]} \quad \left(\text{or } \frac{\frac{1}{b_0}[\cdots]}{\frac{1}{a_0}[\cdots]}\right) \tag{6.12}$$

In the second interval, split the contribution $a_1 b_1$ in $M$ (not necessarily different) ways:

$$\Lambda = \frac{a_0\left[a_1(\cdots) + a_1(\cdots) + b_1(\cdots) + b_1(\cdots)\right]}{b_0\left[\frac{1}{b_1}(\cdots) + \frac{1}{b_1}(\cdots) + \frac{1}{a_1}(\cdots) + \frac{1}{a_1}(\cdots)\right]} \tag{6.13}$$

Continue in the same way in the third interval:

$$\Lambda = \frac{a_0 \left[a_1(\frac{1}{a_2} + \frac{1}{a_2} + \frac{1}{a_2} + a_2) + a_1(\frac{1}{a_2} + b_2 + a_2 + a_2) + \cdots\right]}{b_0 \left[\frac{1}{b_1}(\frac{1}{a_2} + \frac{1}{a_2} + \frac{1}{a_2} + a_2) + \frac{1}{b_1}(\frac{1}{a_2} + b_2 + a_2 + a_2) + \cdots\right]} \tag{6.14}$$

5. The simplified likelihood ratio is:

$$\Lambda = \frac{a_0}{b_0} \cdot a_1 b_1 \cdot 1 = \frac{a_0 a_1 b_1}{b_0} \tag{6.15}$$

A quick check shows that all the identities listed in Appendix C can be generated using this procedure.

### 6.1.2 Properties of the construction procedure

Any likelihood ratio which can be constructed with this procedure simplifies, and no simplifying likelihood ratio exists which cannot be generated with this procedure. This claim will be proved in the next section. The procedure is therefore both *necessary* and *sufficient* with respect to likelihood ratio simplification.

The likelihood ratios constructed by the procedure have the following properties:

- In each interval, only a maximum of four signals (two LTVs and their reciprocals) are used. If the contribution is chosen to contain a single squared LTV (such as $a^2$ or $1/a^2$), only two signals are possible in that interval.

- If they differ, both LTVs from each interval are present in the simplified likelihood ratio (and its corresponding simplified receiver). This implies that only a bi-orthogonal signal set (or an antipodal signal set if the LTVs in that interval are the same) can deliver optimal error performance in a simplified receiver.

- If an interval's contribution is chosen to be 1, the signals in that interval do not contribute to the decoding decision.

## 6.2 Bounds on the Performance of Simplified Receivers

### 6.2.1 Simplification of likelihood ratios

**Theorem 1** *The likelihood transform of a trellis-based likelihood ratio simplifies if and only if, within each symbol interval, LTVs from the numerator can be paired with LTVs from the denominator such that their ratios are equal, provided that in the third and later intervals, LTVs may not be paired which do not share common factors from the prior intervals. A simplified likelihood ratio is equal to the product of the common ratios in each interval.*

This is essentially a restatement of the construction procedure of the previous section. For example, the likelihood ratio based on the identity (4.45) or (6.3) found in section 4.3.3,

$$\frac{a_0\left[a_1\left(a_2+a_2\right)+b_1\left(a_2+b_2\right)\right]}{b_0\left[\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{a_2}\right)\right]} = \frac{a_0a_1b_1a_2b_2}{b_0} \tag{6.16}$$

simplifies because in the first interval, any $a_0$ and $b_0$ can form a single ratio which becomes the first interval's contribution to the decoding decision; in the second interval, the variables from the numerator and denominator can be formed into ratios which are equal:

$$\frac{a_1}{1/b_1} = \frac{b_1}{1/a_1} = a_1b_1$$

and in the third interval,

$$\frac{a_2}{1/b_2} = \frac{a_2}{1/b_2} = \frac{a_2}{1/b_2} = \frac{b_2}{1/a_2} = a_2b_2$$

On the other hand, the very similar likelihood ratio

$$\frac{a_0\left[a_1\left(a_2+a_2\right)+b_1\left(a_2+b_2\right)\right]}{b_0\left[\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{a_2}\right)+\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right]} \tag{6.17}$$

does *not* simplify because the paired LTVs in the last interval do not have equal ratios. Note that it is not allowed in (6.17) to pair for example $b_2$ (numerator) with $1/a_2$ (denominator) in an attempt to obtain equal ratios: they do not have common prior factors.

To make the proof of Theorem 1 easier, it is necessary first to provide two lemmas.

**Lemma 1** *In AWGN, LTVs from different symbol intervals cannot be related directly through a functional relationship even if they are associated with the same signal. In particular,*

$$a_0 a_1 = b_0 b_1 \quad \Leftrightarrow \quad a_0 = b_0 \text{ and } a_1 = b_1$$

LTVs are samples of statistically independent random variables. The lemma follows directly from the properties of random variables.

**Lemma 2** *If $n_i$ and $d_i$, $i = 1, 2, \ldots, M$, are transcendental numbers, then expressions of the form*

$$\frac{\sum_i n_i}{\sum_i d_i}$$

*simplify if and only if all the numbers from the numerator can be paired with numbers from the denominator in such a way that their ratios are equal, that is, if a permutation of $j \in \{1, 2, \ldots, M\}$ can be found such that all $n_i/d_j$ are equal. The common ratio equals the simplified expression.*

For example, the expression

$$\frac{a + a + \frac{1}{b}}{b + b + \frac{1}{a}} = \frac{a}{b} \tag{6.18}$$

simplifies because each term in the numerator can be paired with a term in the denominator (in this case the one directly below it) so that their ratios are equal. From left to right the ratios are:

$$\frac{a}{b} = \frac{a}{b} = \frac{\frac{1}{b}}{\frac{1}{a}} \tag{6.19}$$

The simplified expression is equal to the common ratio. On the other hand

$$\frac{a+a+\frac{1}{b}}{\frac{1}{a}+\frac{1}{a}+b}$$

does not simplify because Lemma 2 cannot be satisfied.

**Proof of Lemma 2:** Let

$$\Lambda = \frac{\sum_i n_i}{\sum_i d_i} = \frac{n_1+n_2+\cdots+n_M}{d_1+d_2+\cdots+d_M} \tag{6.20}$$

Assume that the expression simplifies, so that it can be written as

$$\Lambda = f(n_1,\ldots,n_M,d_1,\ldots,d_M) \tag{6.21}$$

where $f(\ldots)$ is a single term. It follows that

$$n_1+\ldots+n_M = f(n_1,\ldots,n_M,d_1,\ldots,d_M)(d_1+\ldots+d_M) \tag{6.22}$$

Equality is only possible between two sums of transcendental numbers when some permutation of their individual terms are pairwise equal. Therefore, the single-term factor $f(\ldots)$, when multiplied by each number (term) in the denominator, makes it equal to a corresponding number in the numerator:

$$n_i = f(n_1,\ldots,n_M,d_1,\ldots,d_M)d_j \tag{6.23}$$

or

$$\frac{n_i}{d_j} = f(\ldots) \tag{6.24}$$

for all $i = 1,2,\ldots,M$ and some mapping of $i$ on to $j \in \{1,2,\ldots,M\}$.

To prove the inverse, assume (6.24) to be true, and therefore

$$n_1+n_2+\cdots+n_M = f(n_1,\ldots,n_M,d_1,\ldots,d_M)(d_1+d_2+\cdots+d_M) \tag{6.25}$$

or

$$\frac{n_1+n_2+\cdots+n_M}{d_1+d_2+\cdots+d_M} = f(n_1,\ldots,n_M,d_1,\ldots,d_M) \tag{6.26}$$

which simplifies because $f(\ldots)$ is a single term. *End of proof.*

If $M = 2$ and $a, b, c$ and $d$ are positive real numbers such that $a/c = b/d$, then

$$\frac{a+b}{c+d} = \frac{a}{c} \tag{6.27}$$

simplifies. Specifically, the MSK identity (4.1)

$$\Lambda = \frac{a+b}{\frac{1}{b}+\frac{1}{a}} = ab, \quad a, b > 0 \tag{6.28}$$

simplifies because

$$\frac{a}{1/b} = \frac{b}{1/a} \tag{6.29}$$

In Lemma 2, variables in the numerator have to be paired with variables in the denominator in the right permutation to achieve simplification of the expression (in that their ratios had to be equal). Assuming that an expression does simplify, it is unnecessary to complicate the proofs that follow by explicitly finding the right permutation each time. It will be assumed from now on that the variables of simplifying expressions are paired in a correct permutation, which is simply a matter of reorganizing the expression without changing its value. This is most easily done in practice by writing the expression so that paired variables in the numerator and denominator are directly above one another, and pairing them from left to right. For example, I will conventionally write

$$\frac{a+b}{\frac{1}{b}+\frac{1}{a}} \quad \text{not} \quad \frac{a+b}{\frac{1}{a}+\frac{1}{b}} \quad \text{because} \quad \frac{a}{1/a} \neq \frac{b}{1/b}$$

The first three lemmas dealt with individual LTVs and with simple expressions comprising a quotient of sums of LTVs. They are now used to prove the general case of Theorem 1:

**Proof of Theorem 1:** Rewrite the general form of a likelihood ratio (6.11) as

$$\frac{b}{a}\Lambda = \frac{c_1(e_1(\cdots)+e_2(\cdots)+\cdots)+c_2(f_1(\cdots)+f_2(\cdots)+\cdots)+\cdots}{d_1(g_1(\cdots)+g_2(\cdots)+\cdots)+d_2(h_1(\cdots)+h_2(\cdots)+\cdots)+\cdots} \tag{6.30}$$

Assume that the right hand side of (6.30) simplifies. Using Lemma 2, terms from the numerator are paired with terms from the denominator into ratios:

$$\frac{c_1(e_1(\cdots)+e_2(\cdots)+\cdots)}{d_1(g_1(\cdots)+g_2(\cdots)+\cdots)} = \frac{c_2(f_1(\cdots)+f_2(\cdots))}{d_2(h_1(\cdots)+h_2(\cdots))} = \cdots \tag{6.31}$$

(where I have used the assumption that terms are paired in the right permutation for simplification). Lemma 1 permits (6.31) to be decomposed into

$$\frac{c_1}{d_1} = \frac{c_2}{d_2} = \cdots \tag{6.32}$$

and

$$\frac{e_1(\cdots)+e_2(\cdots)+\cdots}{g_1(\cdots)+g_2(\cdots)+\cdots} = \frac{f_1(\cdots)+f_2(\cdots)}{h_1(\cdots)+h_2(\cdots)} = \cdots \tag{6.33}$$

Taking each of the quotients in (6.33) in turn, the process of decomposition can be continued recursively until equalities of ratios of single LTVs are obtained. The decomposition process produces the series of equalities

$$\begin{aligned}\frac{c_1}{d_1} &= \frac{c_2}{d_2} = \cdots \\ \frac{e_1}{g_1} &= \frac{e_2}{g_2} = \cdots = \frac{f_1}{h_1} = \frac{f_2}{h_2} = \cdots \\ &\vdots\end{aligned} \tag{6.34}$$

which proves the theorem in the forward direction.

In the inverse direction, assume that (6.34) is given. Starting with the last interval, assemble (6.34) into expressions of the form (6.33):

$$\begin{gathered}\frac{e_1(\cdots)+e_2(\cdots)+\cdots}{g_1(\cdots)+g_2(\cdots)+\cdots} \\ \frac{f_1(\cdots)+f_2(\cdots)+\cdots}{h_1(\cdots)+h_2(\cdots)+\cdots} \\ \vdots\end{gathered} \tag{6.35}$$

Lemma 2 guarantees that all these expressions simplify and that they are equal. Moving back in time by one symbol interval, form the products

$$\frac{c_1}{d_1} \cdot \frac{e_1(\cdots)+e_2(\cdots)+\cdots}{g_1(\cdots)+g_2(\cdots)+\cdots}$$

$$\frac{c_2}{d_2} \cdot \frac{f_1(\cdots) + f_2(\cdots)}{h_1(\cdots) + h_2(\cdots)} \tag{6.36}$$

$$\vdots$$

The product of a single term and a simplifying expression still simplifies into a single term: therefore all these expressions simplify and, by (6.34), they are equal. This construction process can be continued until the original likelihood ratio (6.30) or (6.11) is obtained, and is guaranteed to simplify. *End of proof.*

Theorem 1 justifies the construction procedure of the previous section.

### 6.2.2 Performance bounds

**Theorem 2** *Simplified receivers are optimal (in the maximum likelihood sense) only for modulation schemes which use no more than two pairs of antipodal signals.*

The following lemma will be useful in the proof of Theorem 2:

**Lemma 3** *In equal-energy signalling in AWGN, the only two functional relationships $f$ that can relate two LTVs $b = f(a)$ in the same symbol interval are functional identity: $b = a$; and one that establishes multiplicative reciprocity between them: $b = 1/a$.*

This is the same result as the one stated before in simpler terms on page 104.
**Proof of Lemma 3:** If a functional relationship exists between two LTVs, they would have to be in the same symbol interval, according to Lemma 1. Let the LTV $a$ be the result of a signal correlation $\lambda(S_x)$ in AWGN:

$$a = e^{\frac{2}{N_0}\lambda(S_x)} = e^{\frac{2}{N_0}\int_0^T r(\tau+nT)s_x(\tau)\,d\tau}, \quad 0 \leq \tau < T, \ \ n = 0, 1, 2, \ldots \tag{6.37}$$

where $r(t)$ is the received signal. Similarly let

$$b = e^{\frac{2}{N_0}\lambda(S_y)} \tag{6.38}$$

Since the noise is by definition additive and statistically independent of the signal, the only possible relationship between correlations $\lambda(S_x)$ and $\lambda(S_y)$ is linear scaling:

$$s_y(\tau) = Cs_x(\tau) \;\Leftrightarrow\; \lambda(S_y) = C\lambda(S_x), \quad C \in \mathbb{R} \tag{6.39}$$

For equal-energy signalling, we require that

$$\|S_x\|^2 = \|S_y\|^2 = E \tag{6.40}$$

The only solutions to (6.39) occur at $C = \pm 1$:

$$\begin{aligned} s_y(\tau) = +s_x(\tau) \quad &\Leftrightarrow \quad \lambda(S_y) = +\lambda(S_y) \quad \text{and} \\ s_y(\tau) = -s_x(\tau) \quad &\Leftrightarrow \quad \lambda(S_y) = -\lambda(S_y) \end{aligned} \tag{6.41}$$

Therefore, by (6.37) and (6.38) and the fact that $e$ is transcendental, it follows that the only possible relationships between $a$ and $b$ are

$$\begin{aligned} b = e^{\frac{2}{N_0}\lambda(S_y)} &= e^{\frac{2}{N_0}\lambda(S_x)} = a \quad \text{or} \\ b = e^{\frac{2}{N_0}\lambda(S_y)} &= e^{-\frac{2}{N_0}\lambda(S_x)} = \frac{1}{a} \end{aligned} \tag{6.42}$$

*End of proof.*

**Proof of Theorem 2:** A simplified receiver is the maximum likelihood realization of a simplifying likelihood ratio. For every symbol interval of a simplifying likelihood ratio, Theorem 1 states that all the LTVs in the interval can be combined into a series of equalities of the form

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \cdots \tag{6.43}$$

Since we are interested in the possible number of different signals in this interval, we may without loss of generality assume that

$$a_1 \neq a_2 \neq \cdots \quad (\text{or } b_1 \neq b_2 \neq \cdots) \tag{6.44}$$

The number of ways in which the equalities in (6.43) can be realized without reusing LTVs is limited by Lemma 3. Let $a$ and $b$ be two LTVs with $a \neq b$. One of two cases applies in (6.43):

1. Assign $a_1 = a$ and $b_1 = b$: In this case the only possible further assignment which does not violate either (6.44) or Lemma 3 is $a_2 = 1/b$ and $b_2 = 1/a$. This limits the total number of signals in the interval to two pairs of antipodal signals: $s_a(\tau)$, $s_{1/a}(\tau)$, $s_b(\tau)$, and $s_{1/b}(\tau)$.

2. Assign $a_1 = a$ and $b_1 = 1/a$: In this case no further assignments can be made, which limits the number of signals in the interval to a single pair of antipodal signals.

Therefore, a simplified receiver cannot optimally receive any modulation scheme in which the total number of possible signals in any symbol interval exceeds four, or in which the signals are not antipodally paired. *End of proof.*

A consequence of the proof of Theorem 2 is the implication that if all the LTVs within each symbol interval are assumed to be distinct, then the input alphabet size cannot exceed $M = 2$, since $M$ is the number of terms in each (recursive) sum of LTVs in the likelihood ratio. If repetition of LTVs within a symbol interval of a simplifying likelihood ratio is allowed, $M$ can be made as large as desired, but there will still be only two distinct pairs of antipodal signals present in each interval. The repetition of LTVs corresponds to identically labelled branches from a node, or identical transmitted signals (perhaps with a change in transmitter state) for different inputs.

**Theorem 3** *Only modulation schemes with binary ($M = 2$) inputs can have simplified receivers which use information from all the symbols in their observation period.*

**Proof of Theorem 3:** Figure 6.2 shows three branches from a common state in a trellis. (Four or more branches could also have been considered, but using three is sufficiently general to prove the theorem.) They are labelled $x$, $y$, and $z$. Theorem 2 restricts the LTVs that may be used to label the branches with to

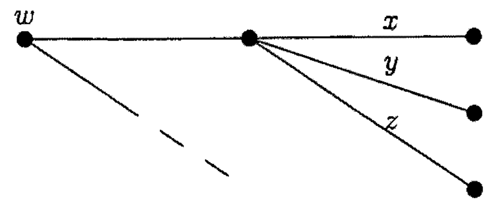



Figure 6.2: Three branches from a common state in a tree, to illustrate allowable ways of labelling them.

the set $\{a, b, 1/b, 1/a\}$. Three likelihood ratios may be formed to decide between pairs of branches.

There are only two essentially different ways of labelling the three branches: with repetition of a label; and without repeating a label. For example, one may label $(x, y, z) = (a, b, a)$ with repetition of a label; or label $(x, y, z) = (a, b, 1/b)$ without repeating a label. In the first case, the likelihood ratio which decides between the two identically labelled branches will have a contribution of 1 in its first interval. The first interval therefore does not necessarily contribute information towards the decoding decision when labels are repeated.

This contravenes the requirement that the receiver should use information from all the symbols in the observation period: repetition of labels are therefore not allowed.

The question may be asked whether one may not choose to ignore the likelihood ratio which has to decide between identically labelled branches, and concentrate instead on the branches which are differently labelled in order to come to a decoding decision. The answer is no, because of the possibility that the other likelihood ratios (two of them in the case of three branches) may contradict one another by not choosing either their common branch or the other two branches. All three likelihood ratios are needed, combined in log-linear fashion, for maximum likelihood decisions.

In the second case, labelling without repetition of labels, it is necessary to examine the likelihood ratio from the state one interval earlier, marked $w$ in Figure 6.2. In this likelihood ratio, all three (different) LTVs assigned to branches $x$, $y$ and $z$ appear together in the second interval of the numerator (or the denominator). The only possible contribution from this interval is 1, because there is no way of satisfying Theorem 2 as well as Lemma 2 with three or more different LTVs together in the numerator (or denominator) in a single interval of a simplifying likelihood ratio.

Therefore, it is not possible to label three or more branches with LTVs so that the interval in question is both part of a simplifying likelihood ratio and contributes to the decision.

Having only two branches ($M = 2$) does not pose the same problems in terms of Theorem 2 and Lemma 2, and therefore, only binary modulation can be optimally received in a simplified receiver which uses information from all the intervals in its observation period. *End of proof.*

Theorem 3 immediately removes the use of the family of MSK-like identities (4.43) or (6.1) from consideration for the simplification of likelihood ratios, except for the MSK case of $M = 2$. All the other identities found in the searches of Chapter 4 satisfy Theorems 2 and 3.

**Theorem 4** *Any coded modulation scheme which has a simplified receiver which uses information from every symbol during its observation period, has no more than two states in its trellis.*

The implication of this theorem is that the transmitter can have no more than a single binary memory element, if the receiver is a simplified receiver.

**Proof of Theorem 4:** According to Theorem 3, only binary modulations need be considered. The tree in Figure 6.3 represents a general binary likelihood ratio. For the purpose of determining the maximum number of usable states, the observation period is assumed to be infinitely long. We shall concentrate on the last symbol interval shown in Figure 6.3 to be the interval of interest.

Assume that it is known that the transmitter was in the state marked $x$ in Figure 6.3, and that a likelihood ratio $\Lambda_x$ is formed at this point to decide which of the two branches from $x$ were most likely followed. (The state $x$ can be regarded as representative of any state in the interval of interest.) Similarly, likelihood ratios $\Lambda_y$ and $\Lambda_z$ can be formed with the states marked $y$ and $z$ as

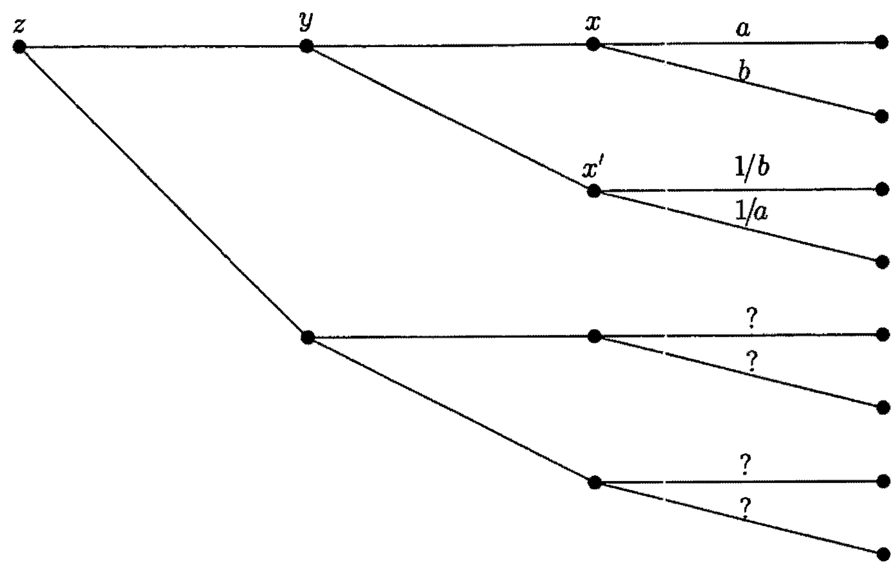



Figure 6.3: A binary tree to show possible labelling of the branches in the last interval.

their origins respectively. The LTVs labelling the branches in the interval of interest will occur in different intervals relative to the start of each likelihood ratio $\Lambda_x$, $\Lambda_y$ and $\Lambda_z$.

The branches in the interval of interest may be labelled in many ways, within the limitations imposed by Theorem 2. Depending on which likelihood ratio is considered ($\Lambda_x$, $\Lambda_y$, or $\Lambda_z$), the labels of the upper half of the branches within the scope of the likelihood ratio are assumed to be the LTVs in the numerator, and the LTVs in the lower half are in the denominator. For example, the upper two branches in the interval (both from $x$) are assumed to be in the numerator of $\Lambda_y$ and the next two branches (both from the state marked $x'$) are in the denominator.

Whichever likelihood ratio is being considered, Theorem 2 and Lemma 2 allow only four essentially different cases, as distinguished by their contributions:

1. Contribution $a/b$: Pairs of upper (numerator) and lower (denominator) branches can be labelled $a$ and $b$ respectively, or $1/b$ and $1/a$.

2. Contribution $ab$: Pairs of upper and lower branches can be labelled $a$ and $1/b$ respectively, or $b$ and $1/a$.

3. Contribution $a^2$: Pairs of upper and lower branches must be labelled $a$

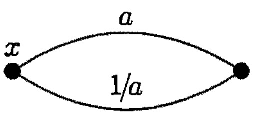



Figure 6.4: Binary uncoded antipodal modulation resulting from the choice of contribution of $a^2$.

and $1/a$ respectively.

4. Contribution 1: Any of the four LTVs $a$, $b$, $1/b$, or $1/a$ may appear in the numerator (upper branches), provided that the corresponding LTVs in the denominator (lower branches) are the same.

One may formulate more cases based on other contributions such as $b/a$ or $1/b^2$, but these differ from the cases above by the naming of LTVs only. In the last case (4) the contribution is 1 and according to Theorem 1 the symbols in the interval of interest do not contribute information to the decoding decision. Case 4 will therefore not be allowed in a receiver which uses information from all the symbols within its observation period.

All three the likelihood ratios $\Lambda_x$, $\Lambda_y$ and $\Lambda_z$ have to co-exist in the receiver, being invoked sequentially to make the next decoding decision.

Starting with case 3: If the two branches from $x$ are labelled with $a$ and $1/a$ to give a contribution of $a^2$ in $\Lambda_x$, then it is not possible to label any of the other branches in the interval of interest such that $\Lambda_y$ or $\Lambda_z$ can use information from the interval. The only case which can be made to apply to $\Lambda_y$ and $\Lambda_z$ (and all earlier likelihood ratios which include $x$ in their scope) is case 4 (contribution of 1) which is not allowed. Therefore, it is possible to label a total of only two branches in the interval, which implies uncoded binary antipodal modulation as shown in Figure 6.4.

Turning to the remaining cases 1 and 2: If the two branches from $x$ are labelled with $a$ and $b$ to give a contribution of $a/b$ in $\Lambda_x$, then the only case available to $\Lambda_y$ is case 2 which allows labelling of the next two branches from $x'$ with $1/b$ and $1/a$. This is the situation illustrated in Figure 6.3. Alternatively, $\Lambda_x$ can be formed with case 2 (branch labels $a$ and $1/b$ from $x$) and $\lambda_y$ with

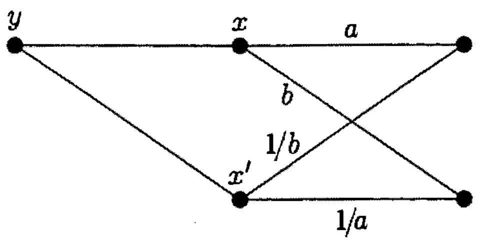



Figure 6.5: Two-state trellis resulting from a choice of contributions $a/b$ and $ab$.

case 1 (branch labels $b$ and $1/a$ from $x'$). Either way, there is no way to label the rest of the branches in the interval of interest for $\Lambda_z$ without using case 4. Therefore, it is possible to label a total of only four branches in the interval, which implies at most two states in the trellis. Figure 6.5 shows one way of turning Figure 6.3 into a trellis, in this case the trellis of MSK.

*End of proof.*

It is possible to increase the number of states in a trellis arbitrarily by using a contribution of 1 in some intervals of each likelihood ratio, but doing so cannot result in any improvement in performance. For example, the lower four branches in Figure 6.3 could be labelled with $a$, $b$, $1/b$, and $1/a$ from top to bottom (that is, copying the labels of the top four branches), which makes a contribution of 1 to the likelihood ratio $\Lambda_z$ from the state marked $z$. This would result in MSK with a superfluous binary memory element in the transmitter, as was seen previously in (5.9) on page 122.

If non-contributing intervals are inserted between two contributing intervals, potentially useful schemes such as the one shown in Figure 5.9 and (5.17) may result.

A slight variation of the proof of Theorem 4 results in:

**Corollary 1** *Simplified receivers of binary modulations cannot use information for the decoding decision from more than two symbol intervals in the observation period.*

Like Theorem 4, Corollary 1 is true irrespective of the length of the observation period.

### 6.2.3 Implications for the performance of simplified receivers

Theorems 2 and 4 and their corollaries obviously impose significant limitations on the performance that may be expected from simplified receivers.

The assumption of half-rate coding made in section 4.3.3 is now supported by Theorem 2 and Theorem 3: only binary modulation schemes can have simplified receivers which use all available information, while only a choice of four signals is available. Note that the construction procedure also implicitly assumes half-rate coding.

Theorem 2 prescribes a narrowly defined signal set of at most two pairs of antipodal signals. Optimal performance is attained when the signals are bi-orthogonal, as in the case of MSK, with a free squared Euclidean distance of $d^2_{\text{free}} = 2$.

Theorem 4 and Corollary 1 make it impossible to improve on this error performance with a simplified receiver by adding states or lengthening the observation period. This is the reason for the failure of the trellis searches of section 4.2 to find useful modulation schemes which exceed MSK in error performance.

The MSK signal set is perhaps the most natural for Theorem 2's prescription, although coded QPSK or offset QPSK are also accommodated. These results contribute substantially toward MSK's status as a unique modulation scheme.

# Chapter 7

# Conclusion

This thesis has conclusively identified the conditions under which simplified receivers may exist, and has clearly bounded the performance that can be expected of simplified receivers of equal-energy modulations.

A simplified receiver is derived from the set of likelihood ratios from every state in the modulation scheme's trellis, expressed over a sufficiently long but finite observation period. The likelihood ratios each have to simplify into a single exponential term, in which case a simple maximum likelihood receiver can be constructed from a bank of correlators (or matched filters), some delay elements and some adders.

The first objective of the thesis was to determine whether a useful simplified receiver exists for a continuous-phase frequency shift keying (CPFSK) scheme. It can now be stated with certainty that none exists which can improve on the performance of MSK. This includes four-level CPFSK with modulation index $h = 1/4$, and the principal continuous phase modulation (CPM) considered in this work, quadrature minimum shift keying (QMSK). Any uncertainty which might have existed due to the incomplete searches of the very large search space, has been dispelled by the derivation of exact bounds on the performance of general simplified receivers.

Except for MSK and other modulation schemes with similar likelihood ratios, the class of coded equal-energy modulation schemes contains no simplified receivers as defined in this thesis. The Viterbi algorithm (or a sequential or feedback decoding algorithm) remains the practical solution to maximum likelihood sequence estimation. It has the important advantage of generality, at the cost of higher complexity.

A lot of work has been reported on 'simplified' CPM receivers: reduced-complexity Viterbi detectors [79], limited search receivers [5], average matched filter receivers [56, 32], and MSK-type receivers [77, 78][3, p.295]. In contrast, no major new development has been reported in maximum likelihood receivers for CPM since the Viterbi algorithm [84, 24]. In this respect this work explains the lack of simple CPM receivers reported in the literature. It is perhaps not surprising that the simplest types of 'simplified' CPM receivers [77, 78] mentioned above are nothing more than MSK receivers slightly modified for detecting related signal waveforms.

It is interesting that all the most promising simplified receivers for QMSK were catastrophic. It was recognized that, in general, catastrophic trellises do not necessarily perform as badly as suggested in sections 3.5 and 4.2.4. It requires specific data sequences to track a given zero-distance pair of paths exactly, and the probability that such sequences will occur becomes increasingly small as the length of the sequence grows. Bursts of errors may occur, but the receiver may be able to resynchronize with the transmitter and resume correct demodulation. There is some evidence for this hypothesis: Rimoldi [68] has calculated that the probability of a catastrophic event in CPM transmissions is vanishingly small. His definition of catastrophic CPM schemes was slightly different to mine, because he considered partial response CPM without coding.

Accordingly, a class of schemes called limited catastrophic schemes was identified in section 5.1.1, with potential for practical implementation. However, none of the catastrophic QMSK schemes which had better error performance than MSK were limited, a result which was later confirmed by the bounds set on simplified receiver performance. A re-examination of Rimoldi's catastrophic partial response schemes from the point of view of likelihood ratios may provide a new perspective.

The second objective of the thesis was to explore the space of mathematical identities which may aid simplification of likelihood ratios of digital phase modulation in general. The identity which causes the simplification of the MSK likelihood ratio remained prominent, in that there does not exist any other identity for simplification of likelihood ratios over two symbol intervals. Over an observation period of three symbol intervals, only the case of binary inputs could be exhaustively searched, and the three types of identity discussed in section 5.2.3 were found. These identities, too, appear to be elaborations of the MSK identity, and the fact that no others were found was pivotal in suggesting that an analytical approach may be tractable.

This does not mean that simplified receivers do not exist at all. It was shown in section 4.1 that if the continuous-phase requirement is dropped, simplified receivers can be found by construction from simplified likelihood ratios. The two-symbol receiver from the end of Chapter 5 is an example of what may be practically achieved. This scheme and others like it will however be limited in error performance to that of MSK, and may furthermore have poorer spectral efficiency and have a more complex receiver than MSK. Its usefulness will have to depend on other properties, such as the provision of time diversity in a fading channel.

The third objective of the thesis was to attempt a formal analysis of simplifying likelihood ratios. Accordingly, a construction procedure was formulated in section 6.1 to generate all the found identities automatically, and it was verified that it was not possible to create any identities by this procedure which were not already in the list of found identities. By examining the structure of the construction procedure, it was possible to proceed with a formal proof in Theorem 1, which specified how a likelihood ratio simplifies and hence what the limits of possibility are on its construction.

With the simplification process well understood, it became possible to show what characteristics the modulation scheme and receiver had to have in order for the likelihood ratio to simplify. These characteristics can be divided and discussed as two cases:

1. the case when the receiver uses information from every symbol in its observation period; and

2. when some symbol intervals may not contribute to the decoding decision.

The first case is the only one with pretensions to good error performance: it is not possible to design a receiver with large free Euclidean distance for a fixed number of states without using information from every symbol interval. Failing to do so forces an increase in the number of states required to keep paths from remerging prematurely.

In this first case, it was found that simplified receivers which use information from all the symbols in their observation period are optimal only

1. for antipodal bi-orthogonal signal sets with at most four signals (Theorem 2);

2. for binary modulations (Theorem 3); and

3. for observation periods of no more than two symbol intervals (Theorem 4).

This is a severely restrictive set of conditions, which seems to be almost a restatement of MSK's characteristics. Apart from MSK, there are only a few modulations which could conform to it: half-rate coded 4-PSK and offset-QPSK are two examples.

In the second case, it is not required that information from all the symbols in the observation period is used, so that the last of the conditions above (3) may be relaxed. The above-mentioned scheme of section 5.2.4 is a typical example. Relaxing this condition adversely affects the scheme's potential performance, because one would normally expect to be able to design coded modulation schemes with improved error performance, given a longer observation period.

It is possible to extend the observation period indefinitely while maintaining simplification of the likelihood ratio, but at most two of the symbol intervals will contribute to the decoding decision (Corollary 1). Since each of those two intervals will contribute two likelihood transform variables (or signals) to the decoding decision, best performance is obtained when the two signals are orthogonal. The conclusion for all cases is inevitably that the best free squared Euclidean distance achievable is $d^2_{\text{free}} = 2$, the same as MSK.

In the course of this work, a powerful and intuitively appealing notation was developed for analysing the likelihood ratios of coded modulation schemes. The likelihood transform greatly facilitates

- the derivation of the likelihood ratio for a maximum likelihood receiver directly from the encoder's trellis diagram;
- comparisons between superficially unrelated modulation schemes;
- immediate verification of whether a simplified receiver exists; and
- the concise formulation of the likelihood parameter receiver of Osborne and Luntz [56]

for any digital phase modulation.

Some oversights in the published literature have been uncovered using the likelihood transform, such as the erroneous proposal of a simplified receiver for 4-CPFSK by Kritzinger [40], and good performance claimed for a degenerate scheme by Anderson and de Buda [4].

More generally, the likelihood transform may be useful outside the present context. Using the transform, it is possible to discern at a glance whether two apparently unrelated modulation schemes which use different signal sets are related at the level of the Bayesian likelihood ratio test, and hence to draw conclusions about the bounds on their performance. The likelihood transform is independent of specific choices of signal set.

It is not necessary for likelihood ratios to simplify in order to use the likelihood transform. It may for example be possible to explore the relationship between likelihood ratios formulated in different encoder states and the design rules for rotationally invariant trellises [86, 87, 81]. The likelihood transform may provide a useful tool in justifying some heuristic design rules.

This dissertation has only considered equal-energy signalling. Likelihood ratios can be written for any modulation scheme, and may have different properties for more general signalling. The likelihood transform can still be applied in the same way: $a$ and $a^2$ are the likelihood transform variables of $s_a(t)$ and $2s_a(t)$ respectively, for example.

The main contribution of this thesis lies in advancing the state of the art of likelihood ratio analysis in equal-energy coded modulation, specifically with reference to bounds that were set on the structure of simplifying likelihood ratios. These bounds have implications for receiver design in the entire class of coded digital phase modulation.

# Appendix A

# The Number of QMSK Trellis Codes

This appendix gives a description of the enumeration of the possible trellis codes of QMSK, as needed in section 4.2.4.

Only half of the QMSK encoder states are used in any one interval. In so-called even intervals (see the description of QMSK in section 4.2.2), the choice of used states is fixed, but in odd intervals, the trellis may use one of many combinations of states.

Assuming for the moment that a particular choice has been made on which states to use, and that we only consider a time-invariant CE ($J = 1$) over a single symbol interval, the problem reduces to finding out how many trellises may be constructed with $n_s = ZP/2$ states, where $Z$ and $P$ are the number of CE and CPE states respectively.

I shall assume throughout that parallel transitions are not allowed, that is, branches from the same state which merge again immediately in a common next state. A parallel transition cannot be differentiated at the receiver because there is a Euclidean distance of zero between its branches. I call the number of trellises with no parallel transitions $T(n_s)$.

The general solution is not trivial. A naïve combinatorial approach soon bogs down in the fact that, for four or more states, the trellis need not necessarily be fully interconnected when it is viewed over one interval as an isolated

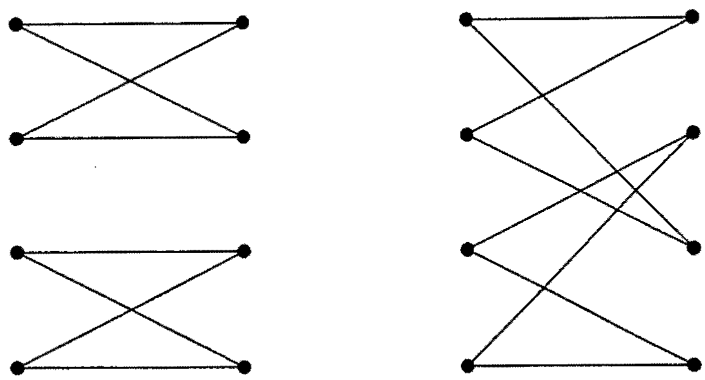


Figure A.1: Two trellises, both of which contain separate sub-graphs. The first is not practically usable, but the second merges its trajectories when cascaded in time with itself.

graph. Figure A.1 shows two four-state examples, both of which contain two separate 'orbits' when viewed as topological graphs rather than trellises. There may, and do, exist mutually exclusive sub-trellises, which do not meet at any state. This severely complicates their enumeration, because one has to count also the number of ways in which each trellis is partitioned. Some of them are not of practical interest (like the first trellis in Figure A.1), because we do not generally want to have mutually exclusive sets of signal trajectories, but in most cases cascading the one-interval trellis with itself merges the trajectories (like the second trellis in Figure A.1).

A method of enumerating the number of trellises $T(n_s)$ as a function of the number of states $n_s$ has been found only recently [57]. It is based on Pólya's Method of Enumeration, when the problem is cast in terms of non-isomorphic two-regular edge and vertex labelled graphs. I counted the numbers of trellises for two to eight states with the help of a computer program, and the results are listed in Table A.1. The combinatorial explosion is evident. Using the new method of enumeration, one may calculate the number of trellises for sixteen states to be $T(16) = 365747519384917483413600000$ [57].

Clearly, with the computing resources available presently, one cannot hope to examine all the trellises for more than about eight states. This places a practical limit of $Z = 4$ on the number of CE states.

So far we have assumed that the choice of used states is known. In even

| $n_s$ | $T(n_s)$ |
|---|---|
| 2 | 1 |
| 3 | 6 |
| 4 | 90 |
| 5 | 2040 |
| 6 | 67950 |
| 7 | 3110940 |
| 8 | 187530840 |

Table A.1: Number of possible trellises as a function of the number of states.

symbol intervals, the choice is fixed: only the states which fall within the phase states corresponding to 0 and $\pi$ radians are allowed, and all of these are used. In odd intervals, however, the choice is free. The number of ways in which $n_s$ states may be chosen from $2n_s$, is

$$\binom{2n_s}{n_s} = \frac{(2n_s)!}{(n_s!)^2}$$

If there are $n_s = ZP/2 = 8$ used states, the expression above gives 12870 combinations of used states in every odd symbol interval. Over a CE cycle length of $J = 4$, the number of trellises is

$$\binom{2n_s}{n_s}^{\frac{J}{2}} T^J(n_s) \approx 2 \times 10^{41}$$

for $Z = 4$.

Not all of these combinations are meaningfully different, because as far as the simplification of likelihood ratios is concerned, all CE states within one phase (CPE) state are equivalent. Only the phase states and the signal trajectories between them constitute the signal, and the receiver is and should be oblivious to the actual CE states used to ensure that the QMSK signal definition is met. If two CE's produce the same output sequence in response to an arbitrary input sequence, then any difference in their sequences of internal states is irrelevant. (If this kind of thinking seems unfamiliar, it is probably because we are not restricting ourselves to time-invariant linear systems. There may be many mappings possible from input to state to output, because the CE is merely

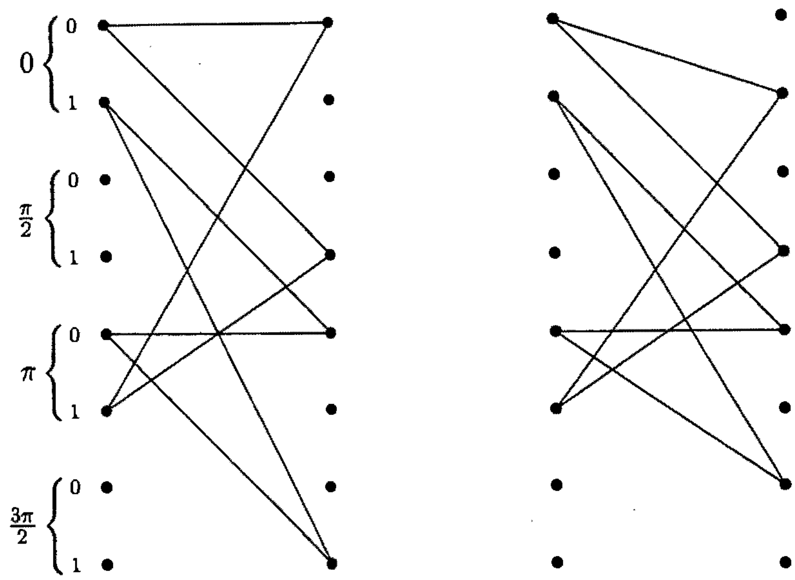



Figure A.2: Trellises with identical signal sets.

a look-up table.)

In terms of generating trellises, the deciding factor therefore is how *many* CE states to be used are chosen from each phase state, and not *which* particular ones are chosen. Henceforth two trellises will be considered identical if they differ only in a choice of CE states within one phase (CPE) state. The branches of these trellises can be labelled with the CE output in an identical way. Figure A.2 shows an example of two trellises ($Z = 2$) which will have identical likelihood ratios from all the starting states shown, yet have different choices of used states in the second interval. Note that only odd symbol intervals have a choice of which states are in use. In even intervals the used states are already fixed as all those belonging in the even phase states 0 and 2.

It is not necessary to calculate the exact number of combinations so generated, because the counting process is immediately complicated again by further restrictions that can be placed on the choice of used states in an effort to cut down on their numbers. In the end, the only practical way of finding out how many combinations there are, when all these *ad hoc* pruning rules are applied, is to count them.

The kinds of restrictions that may be placed on the choice of used states arise from the definition of the QMSK signal. We may specify that each of the four QMSK phases must be used at some time, otherwise we can hardly call the modulation QMSK. If we are desperate to find a simplified receiver, we may compromise by saying that at least one 'non-MSK' phase (1 or 3) must be used. (For if we allow a trellis which does not use any of the 'non-MSK' phases 1 and 3, then we have pure MSK with coding, not QMSK.)

Furthermore, these combinations have to be formulated over the whole of the time-varying cycle time $J$ of the CE, because different sets of phases may be allowed by the CE in different symbol intervals. The only immutable rule is the QMSK signal definition which says in effect that in even intervals, the signal may only use (start from) phase states 0 and 2, and in odd intervals phase states 0, 1, 2 or 3 may be used. The number of combinations go up combinatorially with the number of CE states $Z$ and the CE cycle length $J$.

For example, if no restrictions are placed on the choice of used states, $Z = 2$, $J = 2$ results in nineteen combinations of used states. The process of generating and testing the $T(ZP/2)$ possible trellises therefore has to be repeated nineteen times. On the other hand, if we specify that all four phases have to be used at some time during the $J = 2$ CE cycle, only one possible combination results: it uses one CE state from each phase state.

The number of combinations of used states grows very rapidly with an increase in the number of CE states $Z$ and the length of the CE cycle $J$. Still further pruning rules can be applied: if two trellises differ only in a $\pi$ radian phase shift, they are equivalent. Similarly if two trellises are identical within a time shift of an even number of symbol intervals, they are equivalent. And so on for trellises which are time inverses of one another, or phase inverses. Their likelihood ratios will have exactly the same fate, it is only necessary to test one of them.

All these pruning rules were used in a program called USED which generated all the combinations of used states, and then applied the various shifting and mirroring transformations mentioned above to prune trellises which may be derived from one another. The result was a more manageable number of combinations, shown in Table A.2 for number of CE states $Z = 2$ and 4 and CE

| $Z$ | $J$ | Generated | Kept |
|---|---|---|---|
| 2 | 2 | 18 | 10 |
| 2 | 4 | 360 | 126 |
| 4 | 2 | 84 | 44 |
| 4 | 4 | 7224 | 2536 |

Table A.2: Number of combinations of used states generated, and the number kept after heavy pruning with the program USED.

cycle length $J = 2$ and 4. This version of USED generated only trellises which used at least one odd phase state (1 or 3) during the CE cycle length $J$, and pruned all those which could be derived from another trellis by transformation through

1. a $\pi$ radian phase shift;
2. a phase reversal;
3. a time shift by an even number of symbol intervals; and
4. a time reversal combined with a time shift of an arbitrary number of symbol intervals.

USED also sorted the combinations according to a heuristic rule, so that the ones most likely to contain simplifying likelihood ratios may be searched first. The trellises which are most MSK-like are deemed to have more potential, that is, the ones which tend to cluster their used states in non-adjacent phases.

USED produced a file of candidate used state combinations, which was fed into a program called ALL for historical reasons. (ALL is an abbreviation of ALLINONE, which got its name because it combined all the various trellis tests in one program.) ALL generated all possible trellises with the combination of used states from USED and tested each trellis as described in section 4.2.4.

It was important for the efficiency of the search that ALL should generate as few trellises as possible for evaluation. It was not so easy to decide which trellises may be safely pruned without discarding a potential simplified receiver. Of course parallel transitions were not generated from the outset. One quick test was applied though: during the generation of the trellis, as each branch was connected, a check was performed to see if the new connection resulted in a merging of paths which parted only two intervals ago. Any such merging immediately limits the free Euclidean distance to two at best, the same as that of MSK. Since there is not much point in discovering receivers for coded modulations which perform as well as MSK, such trellises were not even generated.

# Appendix B

# The Search Algorithm for Identities

This appendix describes in detail the algorithm used for generating and evaluating all possible likelihood ratios derived from trellises or trees with given input alphabet size $M$ and observation period $N$. It is intended for readers who may want to duplicate my results, and may be skipped without compromising understanding of the rest of the thesis. It is recommended that section 4.3.3 be read first, which describes the general background to the search and gives an overview of the algorithm. Appendix C gives the complete results of the search.

It is impossible to describe every detail of the implementation of the algorithm here. Those readers who need more information may request a complete program listing in C from me.

## B.1 Simplification

The search is done by generating all possible mathematically distinct expressions for the abbreviated likelihood ratio for a given input symbol alphabet size $M$ and observation period $N$. For each expression, the numerator and denominator are each stripped of all common factors. For convenience in the description of this algorithm, the removal and discarding of common factors (including any common denominators) from either the numerator of the denominator separately, is called *simplification.* This local definition of simplification is not to be confused with simplification of likelihood ratios as defined before. If the resulting numerator and denominator are identical after such simplification, they cancel, proving that an identity of the wanted form does exist. The removed common factors then constitute the right hand side of the identity.

For example, if $M = 2$ and $N = 3$, one possible expression is (4.39) encountered before. Multiplying out and collecting common factors, the expression becomes

$$\begin{aligned} \Lambda &= \frac{a_1(a_2+b_2)+b_1(a_2+b_2)}{\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)} \\ &= \frac{a_1a_2+a_1b_2+b_1a_2+b_1b_2}{\frac{1}{a_1}\frac{1}{a_2}+\frac{1}{a_1}\frac{1}{b_2}+\frac{1}{b_1}\frac{1}{a_2}+\frac{1}{b_1}\frac{1}{b_2}} \\ &= \frac{a_1a_2+a_1b_2+b_1a_2+b_1b_2}{\frac{b_1b_2+b_1a_2+a_1b_2+a_1a_2}{a_1b_1a_2b_2}} \end{aligned} \tag{B.1}$$

The numerator has no common factors, but the denominator has a common denominator which is now discarded. After simplification, the new numerator and denominator are now equal $(a_1a_2 + a_1b_2 + b_1a_2 + b_1b_2)$. They cancel, and an identity has been found. This one is simply the MSK identity applied to both interval 1 and 2: one could factorize the likelihood ratio above as

$$\Lambda = \frac{a_1+b_1}{\frac{1}{a_1}+\frac{1}{b_1}}\frac{a_2+b_2}{\frac{1}{a_2}+\frac{1}{b_2}} \tag{B.2}$$

The algorithm generates expressions in the form of the second line of (B.1), except that each variable is symbolically replaced by an integer so that we are not limited to the letters of the alphabet. The numerator and denominator are each represented by an array of integers. Reciprocals such as $1/a$ are represented by negative integers. The two arrays are simplified separately and then compared for identity before the next expression is generated.

## B.2 Mathematically Distinct Expressions

For the best efficiency possible, the algorithm should ideally only generate mathematically distinct expressions. By this is meant that one expression should not be obtainable from another by simple substitution of variables. The problem that arises is that it is not easy to fill the place-holder variables in the likelihood ratio with signal set variables in such a way that all generated expressions are mathematically distinct. This is because the operations of addition and multiplication in the likelihood ratio are commutative. It therefore seems appropriate to create a multi-set[1] of variables with which to fill each term in the likelihood ratio so that the order of multiplication of the variables is not significant, and to collect them in a multi-set of multi-sets so that the order of addition is not significant. This is closely analogous to partitioning an ordinary set into cells of equal size.

The algorithm attempts this by recursively creating $N-1$ multi-sets of $2m^n$ variables, $n = 1, 2, \ldots, N-1$. Multi-set number $n$ contains the variables for interval number $n$. For example, if $M = 4$, the first interval has a multi-set of variables with 4 elements, the second interval has 16, and so on. The number of branches in the code tree in interval $n$ is $M^n$, and the multi-set contains variables for both the numerator and denominator of the likelihood ratio, hence the factor of 2.

In each interval, the multi-set of variables is partitioned into smaller multi-sets called cells, of equal size $M$. The variables are copied from the cells into the place-holders of the likelihood ratio template. The order of the cells is significant, and the partition is therefore an ordered partition. For example, the two multi-sets of variables associated with (4.39) above, are $\{a_1, b_1, 1/a_1, 1/b_1\}$ and $\{a_2, a_2, b_2, b_2, 1/a_2, 1/a_2, 1/b_2, 1/b_2\}$ for intervals 1 and 2 respectively. The variables in each multi-set are chosen from the set of $m$ signal set variables. If these two multi-sets are partitioned as $\{a_1, b_1, 1/a_1, 1/b_1\}$ and $\{\{a_2, b_2\}, \{a_2, b_2\}, \{1/a_2, 1/b_2\}, \{1/a_2, 1/b_2\}\}$, the likelihood ratio of (4.39) results. Rearranging the cells in a different order, or choosing different variables from the set of signal set variables, gives a different likelihood ratio.

The number of ordered partitions of a set of $B$ elements into $B/M$ cells of equal size $M$ (assuming $B$ is a multiple of $m$), is [46]

$$\frac{B!}{(M!)^{B/M}} \tag{B.3}$$

[1] See the footnote on p.108 for the definition of a multi-set.

The algorithm for generating all the ordered partitions of a set is straightforward, but doing the same for a multi-set (which may have repeated elements) is more difficult. The approach taken here is to partition the multi-set as if it were a set, which results in many of the partitions being identical due to the repeated elements. A complete list of all the ordered partitions is therefore generated for each interval, and sorted, so that the duplicate partitions can be removed. This process is perhaps the most important one limiting the performance of the algorithm, because it is costly both in terms of speed and of memory.

To summarize the algorithm so far: For the current interval, generate all the possible multi-sets of variables, partition each of the multi-sets, sort the list of all possible partitions, and recursively do the same for the next interval. In the last interval, construct an expression from each distinct partition, simplify each expression, and report any identity that is found.

The rest of this subsection describes in more detail how the multi-sets of variables are generated and partitioned and how the resulting expressions are simplified.

## B.3 Generating Multi-Sets of Variables

The signal set is partitioned into two cells of size $M$, so that pairs of inverse signals are in different cells. The variables from one of the cells are now numbered with indices from 1 to $M$. The current index is called 'variable'. The multi-set of variables that will be used in the current interval $n$ is then represented as an array of integers, each of which is the number of times that the signal set variable with that index appears in the multi-set. The array is called 'numberof' and it it indexed by 'variable', as in 'numberof[variable]'. This representation makes the description of the multi-set inherently insensitive to the order of the variables. The sum of the elements of 'numberof[variable]' is therefore the total number of variables in the multi-set, or $2m^n$. As soon as a multi-set is so defined, it is passed on to the GenerateAllReciprocals function which turns all possible combinations of the variables in the multi-set into their reciprocals.

Using these definitions, the algorithm may be summarized as follows:

```
Start with the first of M variables as the current variable.
Fill the empty places left in the multi-set with
        'numberof[variable]' copies of the current variable, where
        'numberof[variable]' ranges from the multi-set size divided
        by M up to all the places, all the time making sure that the
        current variable does not outnumber the previous variable.
    For each value of 'numberof[variable]', if the multi-set is
            still not full, repeat the process recursively for the
            next variable, as long as there are variables left.
    If the multi-set is full, carry on to GenerateAllReciprocals.
```

This process recursively generates all possible multi-sets of $2m^n$ variables chosen from a set of $M$ variables, and passes each multi-set on the next stage, GenerateAllReciprocals.

## B.4 Generating All Combinations of Reciprocals

For each multi-set of variables generated in the current interval, the GenerateAllReciprocals function generates all the possible combinations of reciprocals of each variable. This is done be subtracting one at a time from 'numberof[variable]' and adding it to an auxiliary array called 'numrecips[variable]' which records the number of reciprocals of each variable index in the multi-set. Care is taken that turning a variable into its reciprocal does not cause it to become outnumbered by the next variable, which is not good because then the multi-set is mathematically equivalent to another multi-set through substitution of variables. Each mathematically distinct multi-set so generated is passed on to the function PartitionTheSet, which partitions it in preparation of generating the likelihood ratio expression.

## B.5 Partitioning the Multi-Set

The multi-set of variables is now explicitly created from its associated arrays of numbers of variables of each type ('numberof') and its arrays of numbers of reciprocals ('numrecips'). Care is taken to ensure that no variable can occur in more than one interval, by numbering the variables of each interval from a different range of integers.

All the possible ordered partitions of the multi-set are then generated and stored in a list, which is sorted to detect any duplicates. Duplicates are possible because the multi-set often contains more than one of each variable. For each unique partition, we start all over again recursively with the next interval, or if we are already in the last interval, the actual expression is generated from the partition list and tested.

## B.6 Generating All Ordered Partitions of a Set

As explained before, it is easier to generated all possible ordered partitions of a set than it is to do the same for a multi-set. It is assumed that the size of the input set is an integer multiple of the partition size. The algorithm follows:

```
Make the first variable in the set the current variable.
For each cell, put the current variable into the first available
        place.
    Each time check if the current variable is the last one.
    If it is not, then recursively repeat the process with the next
            variable in the set becoming the current variable.
    If it is, then the partition is complete and it is copied into
            the partition list.
```

## B.7 Generating Likelihood Ratio Expressions

The current partition is used to generate the numerator and denominator. One form of trivial identity is avoided by skipping numerator-denominator pairs which are identical even before simplification. The numerator is expected to change less often than the denominator, so unnecessary simplification of the numerator is further avoided by checking if the numerator is the same as last time. Finally, if either the numerator or the denominator simplifies to a sum of ones only, the rest of the test is skipped, since no useful identity can follow from such a condition.

The actual numerator and denominator are created as two integer arrays from the partition lists. The non-reciprocal variables are represented by integers such as $4, 5, 6, \ldots$ and reciprocals such as $1/a$ by negative integers such as -4. The integer numbering starts from a small integer such as 4. The integers below that are reserved: 0 for end-of-expression, 1 for '1', 2 for '+' and 3 for '/'. The current partition is simply copied element by element into the two arrays. The first array is assumed to be the numerator and the second is the denominator. Tokens for the addition operator (+) are inserted into the arrays as necessary to create a well-formed likelihood ratio. (Adjacent variables are assumed to be multiplied together.) The resulting expression is already 'multiplied out' and ready for simplification.

## B.8 Simplification of the expression

The numerator and denominator are simplified separately, one at a time. The simplification routine does not distinguish between them. The common denominator and all other common factors are dropped, and only the resulting sum of terms is returned. The variables within each term and the terms themselves are sorted to ensure that algebraically identical results are returned as identical expressions.

The input to the simplification routine is in the form of an integer array, for example

$$[5, -6, 2, 5, -7, 0] \leftrightarrow b\frac{1}{c} + b\frac{1}{d}$$

and the output is returned in the same array as for example

$$[6, 2, 7, 0] \leftrightarrow c + d$$

The first step is to compile a table of powers from the input array, containing the number of times each variable appears in each term (its power or exponent). This representation makes it easy to normalize the expression by simultaneously bringing all the terms over a common denominator and extracting all the common factors, because multiplying and dividing by a variable is equivalent to incrementing and decrementing its exponent (power) respectively.

The power table comprises a list of all the variables encountered so far (rows of the table), with each list entry containing an array of the exponent of that variable in each term (columns of the table). The input array is scanned and the exponent of the current variable is incremented or decremented in each term in which the variable was found, depending on whether the variable is a reciprocal (negative) or not (positive). If the given variable has not been seen before, a new entry is created for it in the variable list.

The next step is to remove all common factors and denominators, which are dropped, and sort each term of the remaining sum. The representation of each term is then unique because the order of appearance of the variables in it has been fixed. Getting all the terms in the expression over the same common denominator and removing all common factors including the common denominator, is easily and simultaneously done by normalizing each variable so that its smallest power in any term is 0. Normalizing is done by scanning the terms and finding the smallest power of each variable, which is then subtracted from each term.

Finally, the output array is formed from the modified and sorted power table. Before the final simplified expression is created, it is necessary to create a list of terms and sorting it, so that the order of appearance of the terms in the expression is also fixed. This guarantees that when two simplified expressions are subsequently compared, their corresponding integer arrays will be identical if and only if the expressions are mathematically identical.

# Appendix C

# Results of the Search for Likelihood Ratio Identities

The complete results, with the exception of some trivial cases, of the search for likelihood ratio identities are listed here. Section 4.3.3 describes the parameters of the search and summarizes the results. Appendix B gives the details of the search algorithm.

Results are listed first for input alphabet size $M = 2$ to 8 with observation period length $N = 2$, and then for $M = 2$ and $N = 3$. In all the cases where $N = 2$, the subscripts have been left off the variables for the sake of brevity, because they all come from interval $n = 1$.

The search on the case $M = 2, N = 3$ produced a total of 520 identities. Only 84 of them are listed here, the rest were either of the form

$$[a_1(\cdots) + a_1(\cdots)] \,/\, [b_1(\cdots) + b_1(\cdots)] = \cdots$$

in which common factors ($a_1$ and $b_1$) were already present in the expression, or of the form

$$\left[a_1(\cdots) + \frac{1}{b_1}(\cdots)\right] \,/\, \left[\frac{1}{a_1}(\cdots) + b_1(\cdots)\right] = \cdots$$

which could be obtained from the results listed below by simple substitution (of $\frac{1}{b}$ by $b$ and *vice versa*). The search algorithm was therefore not perfect with regard to its ability to generate only mathematically distinct expressions. The remaining identities listed below still contain many trivial cases, but they are all listed for the sake of completeness.

| $M$ | $N$ | Identities |
|---|---|---|
| 2 | 2 | $\left(a+\frac{1}{b}\right) / \left(\frac{1}{a}+b\right) = \frac{a}{b}$ |
| | | $(a+b) / \left(\frac{1}{a}+\frac{1}{b}\right) = ab$ |
| 3 | 2 | $\left(2a+\frac{1}{b}\right) / \left(\frac{1}{a}+2b\right) = \frac{a}{b}$ |
| 4 | 2 | $\left(3a+\frac{1}{b}\right) / \left(\frac{1}{a}+3b\right) = \frac{a}{b}$ |
| | | $\left(2a+2\frac{1}{b}\right) / \left(2\frac{1}{a}+2b\right) = \frac{a}{b}$ |
| | | $(2a+2b) / \left(2\frac{1}{a}+2\frac{1}{b}\right) = ab$ |
| 5 | 2 | $\left(4a+\frac{1}{b}\right) / \left(\frac{1}{a}+4b\right) = \frac{a}{b}$ |
| | | $\left(3a+2\frac{1}{b}\right) / \left(2\frac{1}{a}+3b\right) = \frac{a}{b}$ |
| 6 | 2 | $\left(5a+\frac{1}{b}\right) / \left(\frac{1}{a}+5b\right) = \frac{a}{b}$ |
| | | $\left(4a+2\frac{1}{b}\right) / \left(2\frac{1}{a}+4b\right) = \frac{a}{b}$ |
| | | $\left(3a+3\frac{1}{b}\right) / \left(3\frac{1}{a}+3b\right) = \frac{a}{b}$ |
| | | $(3a+3b) / \left(3\frac{1}{a}+3\frac{1}{b}\right) = ab$ |
| 7 | 2 | $\left(6a+\frac{1}{b}\right) / \left(\frac{1}{a}+6b\right) = \frac{a}{b}$ |
| | | $\left(5a+2\frac{1}{b}\right) / \left(2\frac{1}{a}+5b\right) = \frac{a}{b}$ |
| | | $\left(4a+3\frac{1}{b}\right) / \left(3\frac{1}{a}+4b\right) = \frac{a}{b}$ |
| 8 | 2 | $\left(7a+\frac{1}{b}\right) / \left(\frac{1}{a}+7b\right) = \frac{a}{b}$ |
| | | $\left(6a+2\frac{1}{b}\right) / \left(2\frac{1}{a}+6b\right) = \frac{a}{b}$ |
| | | $\left(5a+3\frac{1}{b}\right) / \left(3\frac{1}{a}+5b\right) = \frac{a}{b}$ |
| | | $\left(4a+4\frac{1}{b}\right) / \left(4\frac{1}{a}+4b\right) = \frac{a}{b}$ |
| | | $(4a+4b) / \left(4\frac{1}{a}+4\frac{1}{b}\right) = ab$ |

| $M$ | $N$ | Identities (continued) | | |
|---|---|---|---|---|
| 2 | 3 | $\left[a_1(a_2+a_2)+b_1(a_2+a_2)\right]/\left[\frac{1}{a_1}(b_2+b_2)+\frac{1}{b_1}(b_2+b_2)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1(a_2+a_2)+b_1(a_2+a_2)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]$ | $=$ | $a_1b_1a_2^2$ |
| | | $\left[a_1(a_2+a_2)+b_1\left(a_2+\frac{1}{b_2}\right)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)+\frac{1}{b_1}(b_2+b_2)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1(a_2+a_2)+b_1(b_2+b_2)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right]$ | $=$ | $a_1b_1a_2b_2$ |
| | | $\left[a_1(a_2+a_2)+b_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+\frac{1}{b_1}(b_2+b_2)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1(a_2+b_2)+b_1(a_2+b_2)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right]$ | $=$ | $a_1b_1a_2b_2$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1(a_2+a_2)\right]/\left[\frac{1}{a_1}(b_2+b_2)+\frac{1}{b_1}\left(\frac{1}{a_2}+b_2\right)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1\left(a_2+\frac{1}{b_2}\right)\right]/\left[\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+b_2\right)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1(b_2+b_2)+b_1(a_2+a_2)\right]/\left[\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]$ | $=$ | $a_1b_1a_2b_2$ |
| | | $\left[a_1(b_2+b_2)+b_1(b_2+b_2)\right]/\left[\frac{1}{a_1}(a_2+a_2)+\frac{1}{b_1}(a_2+a_2)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1(b_2+b_2)+b_1\left(\frac{1}{a_2}+b_2\right)\right]/\left[\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)+\frac{1}{b_1}(a_2+a_2)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1(b_2+b_2)+b_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]/\left[\frac{1}{a_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+\frac{1}{b_1}(a_2+a_2)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1(b_2+b_2)\right]/\left[\frac{1}{a_1}(a_2+a_2)+\frac{1}{b_1}\left(a_2+\frac{1}{b_2}\right)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1\left(\frac{1}{a_2}+b_2\right)\right]/\left[\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)+\frac{1}{b_1}\left(a_2+\frac{1}{b_2}\right)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+b_1(b_2+b_2)\right]/\left[\frac{1}{a_1}(a_2+a_2)+\frac{1}{b_1}\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right]$ | $=$ | $\frac{a_1b_1b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]/\left[\frac{1}{a_1}(a_2+a_2)+\frac{1}{b_1}(a_2+a_2)\right]$ | $=$ | $\frac{a_1b_1}{a_2^2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+b_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right]/\left[\frac{1}{a_1}(a_2+a_2)+\frac{1}{b_1}(b_2+b_2)\right]$ | $=$ | $\frac{a_1b_1}{a_2b_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right]/\left[\frac{1}{a_1}(a_2+b_2)+\frac{1}{b_1}(a_2+b_2)\right]$ | $=$ | $\frac{a_1b_1}{a_2b_2}$ |
| | | $\left[a_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+b_1(a_2+a_2)\right]/\left[\frac{1}{a_1}(b_2+b_2)+\frac{1}{b_1}\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]$ | $=$ | $\frac{a_1b_1a_2}{b_2}$ |
| | | $\left[a_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right]/\left[\frac{1}{a_1}(b_2+b_2)+\frac{1}{b_1}(a_2+a_2)\right]$ | $=$ | $\frac{a_1b_1}{a_2b_2}$ |

| $M$ | $N$ | Identities (continued) |
|---|---|---|
| 2 | 3 | $\left[a_1 (a_2 + a_2) + b_1 (a_2 + b_2)\right] / \left[\frac{1}{a_1} (a_2 + b_2) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + a_2) + b_1 \left(a_2 + \frac{1}{a_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{a_2}\right) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + a_2) + b_1 (b_2 + b_2)\right] / \left[\frac{1}{a_1} (b_2 + b_2) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + a_2) + b_1 \left(b_2 + \frac{1}{b_2}\right)\right] / \left[\frac{1}{a_1} \left(b_2 + \frac{1}{b_2}\right) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + a_2) + b_1 \left(\frac{1}{a_2} + b_2\right)\right] / \left[\frac{1}{a_1} \left(\frac{1}{a_2} + b_2\right) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + a_2) + b_1 \left(\frac{1}{a_2} + \frac{1}{a_2}\right)\right] / \left[\frac{1}{a_1} \left(\frac{1}{a_2} + \frac{1}{a_2}\right) + \frac{1}{b_1} (a_2 + a_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + b_2) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} (a_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + b_2) + b_1 (a_2 + b_2)\right] / \left[\frac{1}{a_1} (a_2 + b_2) + \frac{1}{b_1} (a_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + b_2) + b_1 \left(a_2 + \frac{1}{a_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{a_2}\right) + \frac{1}{b_1} (a_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + b_2) + b_1 \left(a_2 + \frac{1}{b_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{b_2}\right) + \frac{1}{b_1} (a_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 (a_2 + b_2) + b_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right)\right] / \left[\frac{1}{a_1} \left(\frac{1}{a_2} + \frac{1}{b_2}\right) + \frac{1}{b_1} (a_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{a_2}\right) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} \left(a_2 + \frac{1}{a_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{a_2}\right) + b_1 (a_2 + b_2)\right] / \left[\frac{1}{a_1} (a_2 + b_2) + \frac{1}{b_1} \left(a_2 + \frac{1}{a_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{a_2}\right) + b_1 \left(a_2 + \frac{1}{a_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{a_2}\right) + \frac{1}{b_1} \left(a_2 + \frac{1}{a_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{a_2}\right) + b_1 \left(b_2 + \frac{1}{b_2}\right)\right] / \left[\frac{1}{a_1} \left(b_2 + \frac{1}{b_2}\right) + \frac{1}{b_1} \left(a_2 + \frac{1}{a_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{b_2}\right) + b_1 (a_2 + b_2)\right] / \left[\frac{1}{a_1} (a_2 + b_2) + \frac{1}{b_1} \left(a_2 + \frac{1}{b_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(a_2 + \frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2} + b_2\right)\right] / \left[\frac{1}{a_1} \left(\frac{1}{a_2} + b_2\right) + \frac{1}{b_1} \left(a_2 + \frac{1}{b_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 (b_2 + b_2) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} (b_2 + b_2)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(b_2 + \frac{1}{b_2}\right) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} \left(b_2 + \frac{1}{b_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(b_2 + \frac{1}{b_2}\right) + b_1 \left(a_2 + \frac{1}{a_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{a_2}\right) + \frac{1}{b_1} \left(b_2 + \frac{1}{b_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(\frac{1}{a_2} + b_2\right) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} \left(\frac{1}{a_2} + b_2\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(\frac{1}{a_2} + b_2\right) + b_1 \left(a_2 + \frac{1}{b_2}\right)\right] / \left[\frac{1}{a_1} \left(a_2 + \frac{1}{b_2}\right) + \frac{1}{b_1} \left(\frac{1}{a_2} + b_2\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(\frac{1}{a_2} + \frac{1}{a_2}\right) + b_1 (a_2 + a_2)\right] / \left[\frac{1}{a_1} (a_2 + a_2) + \frac{1}{b_1} \left(\frac{1}{a_2} + \frac{1}{a_2}\right)\right] = a_1 b_1$ |
| | | $\left[a_1 \left(\frac{1}{a_2} + \frac{1}{b_2}\right) + b_1 (a_2 + b_2)\right] / \left[\frac{1}{a_1} (a_2 + b_2) + \frac{1}{b_1} \left(\frac{1}{a_2} + \frac{1}{b_2}\right)\right] = a_1 b_1$ |

| $M$ | $N$ | Identities (continued) |
|---|---|---|
| 2 | 3 | $[a_1(a_2+a_2)+b_1(a_2+a_2)]\,/\,[a_1(b_2+b_2)+b_1(b_2+b_2)] = \frac{a_2}{b_2}$ |
| | | $\left[a_1(a_2+a_2)+b_1\left(a_2+\frac{1}{b_2}\right)\right]/\left[a_1(b_2+b_2)+b_1\left(\frac{1}{a_2}+b_2\right)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1(a_2+a_2)+\frac{1}{a_1}(a_2+a_2)\right]/\left[a_1(b_2+b_2)+\frac{1}{a_1}(b_2+b_2)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1(a_2+a_2)+\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)\right]/\left[a_1(b_2+b_2)+\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1(a_2+b_2)+\frac{1}{a_1}(a_2+b_2)\right]/\left[a_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right] = a_2b_2$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1(a_2+a_2)\right]/\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1(b_2+b_2)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1\left(a_2+\frac{1}{b_2}\right)\right]/\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1\left(\frac{1}{a_2}+b_2\right)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+\frac{1}{a_1}(a_2+a_2)\right]/\left[a_1\left(\frac{1}{a_2}+b_2\right)+\frac{1}{a_1}(b_2+b_2)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1\left(a_2+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)\right]/\left[a_1\left(\frac{1}{a_2}+b_2\right)+\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)\right] = \frac{a_2}{b_2}$ |
| | | $[a_1(b_2+b_2)+b_1(a_2+a_2)]\,/\left[a_1\left(\frac{1}{a_2}+\frac{1}{a_2}\right)+b_1\left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] = a_2b_2$ |
| | | $[a_1(b_2+b_2)+b_1(b_2+b_2)]\,/\,[a_1(a_2+a_2)+b_1(a_2+a_2)] = \frac{b_2}{a_2}$ |
| | | $\left[a_1(b_2+b_2)+b_1\left(\frac{1}{a_2}+b_2\right)\right]/\left[a_1(a_2+a_2)+b_1\left(a_2+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1(b_2+b_2)+\frac{1}{a_1}(b_2+b_2)\right]/\left[a_1(a_2+a_2)+\frac{1}{a_1}(a_2+a_2)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1(b_2+b_2)+\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)\right]/\left[a_1(a_2+a_2)+\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1(b_2+b_2)\right]/\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1(a_2+a_2)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+b_1\left(\frac{1}{a_2}+b_2\right)\right]/\left[a_1\left(a_2+\frac{1}{b_2}\right)+b_1\left(a_2+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+\frac{1}{a_1}(b_2+b_2)\right]/\left[a_1\left(a_2+\frac{1}{b_2}\right)+\frac{1}{a_1}(a_2+a_2)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+b_2\right)+\frac{1}{a_1}\left(\frac{1}{a_2}+b_2\right)\right]/\left[a_1\left(a_2+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(a_2+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+b_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right]/\,[a_1(a_2+b_2)+b_1(a_2+b_2)] = \frac{1}{a_2b_2}$ |
| | | $\left[a_1\left(\frac{1}{a_2}+\frac{1}{b_2}\right)+\frac{1}{a_1}\left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right]/\left[a_1(a_2+b_2)+\frac{1}{a_1}(a_2+b_2)\right] = \frac{1}{a_2b_2}$ |

| $M$ | $N$ | Identities (continued) |
|---|---|---|
| 2 | 3 | $\left[a_1 (a_2+a_2) + b_1 (a_2+a_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = a_2^2$ |
| | | $\left[a_1 (a_2+a_2) + b_1 (b_2+b_2)\right] / \left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = a_2 b_2$ |
| | | $\left[a_1 (a_2+a_2) + b_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] / \left[a_1 (b_2+b_2) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1 (a_2+a_2) + \frac{1}{a_1} (a_2+a_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = a_2^2$ |
| | | $\left[a_1 (a_2+a_2) + \frac{1}{a_1} (b_2+b_2)\right] / \left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = a_2 b_2$ |
| | | $\left[a_1 (a_2+a_2) + \frac{1}{a_1} \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] / \left[a_1 (b_2+b_2) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1 (a_2+b_2) + b_1 (a_2+b_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2}+\frac{1}{b_2}\right)\right] = a_2 b_2$ |
| | | $\left[a_1 (b_2+b_2) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + b_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1 (b_2+b_2) + \frac{1}{a_1} (a_2+a_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] = a_2 b_2$ |
| | | $\left[a_1 (b_2+b_2) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + \frac{1}{a_1} \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + b_1 (b_2+b_2)\right] / \left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + b_1 (a_2+a_2)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + b_1 (a_2+a_2)\right] = \frac{1}{a_2^2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + b_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] / \left[a_1 (b_2+b_2) + b_1 (a_2+a_2)\right] = \frac{1}{a_2 b_2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} (b_2+b_2)\right] / \left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + \frac{1}{a_1} (a_2+a_2)\right] = \frac{b_2}{a_2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + \frac{1}{a_1} (a_2+a_2)\right] = \frac{1}{a_2^2}$ |
| | | $\left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} \left(\frac{1}{b_2}+\frac{1}{b_2}\right)\right] / \left[a_1 (b_2+b_2) + \frac{1}{a_1} (a_2+a_2)\right] = \frac{1}{a_2 b_2}$ |
| | | $\left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + b_1 (a_2+a_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + b_1 (b_2+b_2)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + b_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + b_1 (b_2+b_2)\right] = \frac{1}{a_2 b_2}$ |
| | | $\left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + \frac{1}{a_1} (a_2+a_2)\right] / \left[a_1 \left(\frac{1}{a_2}+\frac{1}{a_2}\right) + \frac{1}{a_1} (b_2+b_2)\right] = \frac{a_2}{b_2}$ |
| | | $\left[a_1 \left(\frac{1}{b_2}+\frac{1}{b_2}\right) + \frac{1}{a_1} \left(\frac{1}{a_2}+\frac{1}{a_2}\right)\right] / \left[a_1 (a_2+a_2) + \frac{1}{a_1} (b_2+b_2)\right] = \frac{1}{a_2 b_2}$ |

# Bibliography


[1] O. Agazzi, D.G. Messerschmitt, and D.A. Hodges. Nonlinear echo cancellation of data signals. *IEEE Transactions on Communications*, COM-30(11):2421–2433, November 1982.

[2] F. Amoroso and J.A. Kivett. Simplified MSK signaling technique. *IEEE Transactions on Communications*, COM-25(4):433–441, April 1977.

[3] J.B. Anderson, T. Aulin, and C-E. Sundberg. *Digital Phase Modulation.* Plenum Press, 1986.

[4] J.B. Anderson and R. de Buda. Better phase-modulation error performance using trellis phase codes. *Electronics Letters*, 12(22):587–588, 1976.

[5] J.B. Anderson and S. Mohan. Sequential coding algorithms: A survey and cost analysis. *IEEE Transactions on Communications*, COM-32:169–176, 1984.

[6] J.B. Anderson and C-E.W. Sundberg. Advances in constant envelope coded modulation. *IEEE Communications Magazine*, 29(12):36–45, December 1991.

[7] J.B. Anderson, C-E.W. Sundberg, T. Aulin, and N. Rydbeck. Power-bandwidth performance of smoothed phase modulation codes. *IEEE Transactions on Communications*, COM-29(3):187–195, March 1981.

[8] J.B. Anderson and D.P. Taylor. A bandwidth-efficient class of signal space codes. *IEEE Transactions on Information Theory*, IT-24:703–712, 1978.

[9] T. Aulin, N. Rydbeck, and C-E.W. Sundberg. Continuous phase modulation — part II: Partial response signaling. *IEEE Transactions on Communications*, COM-29(3):210–225, March 1981.

[10] T. Aulin and C-E.W. Sundberg. Continuous phase modulation — part I: Full response signaling. *IEEE Transactions on Communications*, COM-29(3):196–209, March 1981.

[11] A. Baker. *Transcendental Number Theory*. Cambridge University Press, 1975.

[12] S. Benedetto, R. Garello, M. Mondin, and G. Montorsi. Geometrically uniform TCM codes over groups based on $L\times$MPSK constellations. *IEEE Transactions on Information Theory*, 40(1):137–152, January 1994.

[13] S. Benedetto, M.A. Marsan, G. Albertengo, and E. Giachin. Combined coding and modulation: Theory and applications. *IEEE Transactions on Information Theory*, 34(2):223–236, March 1988.

[14] E. Biglieri, D. Divsilar, P.J. McLane, and M.K. Simon. *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, 1991.

[15] B. Böhm, J.A. Schoonees, and R.M. Braun. Data to frequency mappings in various MSK schemes. *IEE Electronics and Communication Engineering Journal*, 6(1):13–20, February 1994.

[16] A.R. Calderbank and J.E. Mazo. A new description of trellis codes. *IEEE Transactions on Information Theory*, IT-30(6):784–791, November 1984.

[17] A.R. Calderbank and N.J.A. Sloane. New trellis codes based on lattices and cosets. *IEEE Transactions on Information Theory*, IT-33(2):177–195, March 1987.

[18] A. Chouly and H. Sari. Six-dimensional trellis-coding with QAM signal sets. *IEEE Transactions on Communications*, 40(1):24–33, January 1992.

[19] W.B. Davenport, Jr. and W.L. Root. *An Introduction to the Theory of Random Signals and Noise*. McGraw-Hill, New York, 1958.

[20] R. de Buda. Coherent demodulation of frequency-shift keying with low deviation ratio. *IEEE Transactions on Communications*, COM-20:429–435, June 1972.

[21] F. de Jager and C.B. Dekker. Tamed frequency modulation: A novel method to achieve spectrum economy in digital transmission. *IEEE Transactions on Communications*, COM-26(5):534–542, May 1978.

[22] M.L. Doelz and E.H. Heald. Minimum-shift data communication system, March 1961. U.S. Patent 2977417.

[23] R.M. Fano. A heuristic discussion of probabilistic decoding. *IEEE Transactions on Information Theory*, IT-9:64–74, April 1963.

[24] G.D. Forney, Jr. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, March 1973.

[25] G.D. Forney, Jr. Coset codes — part I: Introduction and geometrical classification. *IEEE Transactions on Information Theory*, 34(5):1123–1151, September 1988.

[26] G.D. Forney, Jr. Coset codes — part II: Binary lattices and related codes. *IEEE Transactions on Information Theory*, 34(5):1152–1187, September 1988.

[27] P. Galko and S. Pasupathy. Optimal linear receiver filters for binary digital signals. In *Conference Proceedings of the 1982 International Conference on Communications ICC' 82*, pages 1H6.1–5, June 1982.

[28] P. Galko and S. Pasupathy. Linear receivers for correlatively coded MSK. *IEEE Transactions on Communications*, COM-33(4):338–347, April 1985.

[29] P.A. Galko. Optimization of linear receivers for data communication signals. *IEEE Transactions on Information Theory*, 34(1):79–92, January 1988.

[30] S.A. Gronemeyer and A.L. McBride. MSK and offset QPSK modulation. *IEEE Transactions on Communications*, COM-24(8):809–820, August 1976.

[31] S. Haykin. *Digital Communications.* John Wiley and Sons, 1988.

[32] W. Hirt and S. Pasupathy. Suboptimal reception of binary shift keying signals. *IEE Proceedings, Part F*, 128(3):125–134, September 1981.

[33] P.K.M. Ho and P.J. McLane. Spectrum, distance and receiver complexity of encoded continuous phase modulation. *IEEE Transactions on Information Theory*, 34(5):1021–1032, September 1988.

[34] M.S. Hodgart. Research into generalised MSK. Internal report, University of Cape Town, June 1992.

[35] M.S. Hodgart and J.A. Schoonees. A robust MSK demodulator through DSP. In *Proceedings of IEEE 1992 South African Symposium on Communications and Signal Processing COMSIG '92*, pages 211–216. IEEE, September 1992.

[36] W. Holubowicz and F. Morales-Moreno. Convolutional coding of binary CPM schemes with no increase in receiver complexity. *IEEE Transactions on Communications*, 43(2/3/4):1221–1224, February/March/April 1995.

[37] H. Imai and S. Hirakawa. A new multilevel coding method using error-correcting codes. *IEEE Transactions on Information Theory*, IT-23(3):371–377, May 1977.

[38] P. Kabal and S. Pasupathy. Partial response signaling. *IEEE Transactions on Communications*, COM-23(9):921–934, September 1975.

[39] E. Kreyszig. *Advanced Engineering Mathematics.* John Wiley & Sons, 1979.

[40] S.E. Kritzinger. An investigation into 4-CPFSK modulation and demodulation design. Master's thesis, University of Cape Town, August 1994.

[41] P.A. Laurent. Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP). *IEEE Transactions on Communications*, COM-34(2):150–160, February 1986.

[42] E.A. Lee and D.G. Messerschmitt. *Digital Communication.* Kluwer, 1988.

[43] L-N. Lee. On optimal soft-decision demodulation. *IEEE Transactions on Information Theory*, IT-22(4):437–444, July 1976.

[44] S. Lin and D.J. Costello, Jr. *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, 1983.

[45] G. Lindell, C.E. Sundberg, and T. Aulin. Minimum Euclidean distance for combinations of short rate 1/2 convolutional codes and CPFSK modulation. *IEEE Transactions on Information Theory*, IT-30(3):509–519, May 1984.

[46] S. Lipschutz. *Theory and Problems of Probability.* McGraw-Hill, 1974.

[47] J.L. Massey. Coding and modulation in digital communications. In *Proceedings 1974 International Zurich Seminar on Digital Communications*, pages E2.1–2.4, March 1974.

[48] J.L. Massey. A generalized formulation of minimum shift keying modulation. In *Conference Record of the International Conference on Communications ICC '80*, volume 2, pages 26.5.1–26.5.4. IEEE, June 1980.

[49] J.L. Massey. The how and why of channel coding. In *Proceedings International Zurich Seminar*, pages E1.1–1.7, March 1984.

[50] J.L. Massey and M.K. Sain. Inverses of linear sequential circuits. *IEEE Transactions on Computers*, C-17(4):330–337, April 1968.

[51] H.R. Mathwich, J.F. Balcewicz, and M.H. Hecht. The effect of tandem band and amplitude limiting on the $E_b/N_0$ performance of minimum (frequency) shift keying (MSK). *IEEE Transactions on Communications*, COM-22(10):1525–1539, October 1974.

[52] H. Miyakawa, H. Harashima, and Y. Tanaka. A new digital modulation scheme, multi-mode binary CPFSK. In *Proceedings, Third International Conference on Digital Satellite Communications, Kyoto, Japan*, pages 105–112, November 1975.

[53] F. Morales-Moreno, W. Holubowicz, and S. Pasupathy. Optimization of coded TFM via matched codes. *IEEE Transactions on Communications*, 42(2/3/4):1586–1594, February/March/April 1994.

[54] F. Morales-Moreno and S. Pasupathy. Structure, optimization, and realization of FFSK trellis codes. *IEEE Transactions on Information Theory*, 34(4):730–751, July 1988.

[55] K. Murota and K. Hirade. GMSK modulation for digital mobile radio telephony. *IEEE Transactions on Communications*, 29(7):1044–1050, July 1981.

[56] W.P. Osborne and M.B. Luntz. Coherent and noncoherent detection of CPFSK. *IEEE Transactions on Communications*, COM-22(8):1023–1026, August 1974.

[57] P. Bonnington, University of Auckland, 1996. Personal communication.

[58] S. Pasupathy. Minimum shift keying: A spectrally efficient modulation. *IEEE Communications Magazine*, 17(4):14–22, July 1979.

[59] P.Z. Peebles. *Digital Communication Systems*. Prentice-Hall, 1987.

[60] S.S. Pietrobon, R.H. Deng, A. Lafaneché, G. Ungerboeck, and D.J. Costello, Jr. Trellis-coded multidimensional phase modulation. *IEEE Transactions on Information Theory*, 36(1):63–89, January 1990.

[61] S.V. Pizzi and S.G. Wilson. Convolutional coding combined with continuous phase modulation. *IEEE Transactions on Communications*, COM-33(1):20–29, January 1985.

[62] A. Premji and D.P. Taylor. Receiver structures for multi-h signalling formats. *IEEE Transactions on Communications*, COM-35(4):439–451, April 1987.

[63] J.G. Proakis. *Digital Communications.* McGraw-Hill, 1989.

[64] B.E. Rimoldi. *Continuous-Phase Modulation and Coding for Bandwidth and Energy Efficiency.* PhD dissertation, Swiss Federal Institute of Technology Zurich, 1988.

[65] B.E. Rimoldi. A decomposition approach to CPM. *IEEE Transactions on Information Theory*, 34(2):260–270, March 1988.

[66] B.E. Rimoldi. Design of coded CPFSK modulation systems for bandwidth and energy efficiency. *IEEE Transactions on Communications*, 37(9):897–905, September 1989.

[67] B.E. Rimoldi. Five views of differential MSK: A unified approach. In R.E. Blahut, editor, *Communications and Cryptography: Two Sides of One Tapestry.* Kluwer, 1994.

[68] B.E. Rimoldi and Q. Li. Catastrophic continuous phase modulation schemes and their non-catastrophic equivalents. *IEEE Transactions on Information Theory*, 40(3):687–695, May 1994.

[69] C.R. Ryan, A.R. Hambley, and D.E. Vogt. 760 mbit/s serial MSK microwave modem. *IEEE Transactions on Communications*, COM-28(5):771–777, 1985.

[70] S.M. Ryu and C.K. Un. A simple method for MSK modulation and demodulation. *IEEE*, 73(11):1690–1691, November 1985.

[71] T.A. Schonhoff. Symbol error probabilities for $M$-ary CPFSK: Coherent and noncoherent detection. *IEEE Transactions on Communications*, COM-24:644–652, June 1976.

[72] J.A. Schoonees and R.M. Braun. A note on the orthogonality of MSK signals. In *Proceedings of IEEE 1993 South African Symposium on*

*Communications and Signal Processing COMSIG '93*, pages 123–127. IEEE, August 1993.

[73] F.G. Stremler. *Introduction to Communication Systems, third edition.* Addison-Wesley, 1990.

[74] C-E.W. Sundberg. Continuous phase modulation. *IEEE Communications Magazine*, 24(4):25–38, April 1986.

[75] E.D. Sunde. Ideal binary pulse transmission by AM and FM. *Bell Systems Technical Journal*, 38:1357–1426, November 1959.

[76] A. Svensson. Optimum linear detector at small and large noise power for a general binary composite hypothesis testing problem. *IEE Proceedings Part F*, 134(7):689–694, December 1987.

[77] A. Svensson and C.-E. Sundberg. Optimum MSK-type receivers for CPM on Gaussian and Rayleigh fading channels. *IEE Proceedings, Part F*, 131(5):480–490, 1984.

[78] A. Svensson and C.-E. Sundberg. Serial MSK-type detection of partial response continuous phase modulation. *IEEE Transactions on Communications*, COM-33(1):44–52, 1985.

[79] A. Svensson, C-E. Sundberg, and T. Aulin. A class of reduced complexity Viterbi detectors for partial response continuous phase modulation. *IEEE Transactions on Communications*, COM-32(10):1079–1087, October 1984.

[80] H.L. Van Trees. *Detection, Estimation, and Modulation Theory, Part 1.* John Wiley and Sons, 1968.

[81] M.D. Trott, S. Benedetto, R. Garello, and M. Mondin. Rotational invariance of trellis codes — part I: Encoders and precoders. *IEEE Transactions on Information Theory*, 42(3):751–765, May 1996.

[82] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, IT-28(1):55–67, January 1982.

[83] G. Ungerboeck and I. Csajka. On improving data-link performance by increasing the channel alphabet and introducing sequence coding. In *1976 International Symposium on Information Theory*, June 1976.

[84] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, IT-13:260–269, April 1967.

[85] A.J. Viterbi and J.K. Omura. *Principles of Digital Communication and Coding*. McGraw-Hill, 1967.

[86] L-F. Wei. Rotationally invariant convolutional channel coding with expanded signal space—part 1: $180^o$. *IEEE Journal on Selected Areas in Communications*, SAC-2(5):659–671, September 1984.

[87] L-F. Wei. Rotationally invariant convolutional channel coding with expanded signal space—part 2: Nonlinear codes. *IEEE Journal on Selected Areas in Communications*, SAC-2(5):672–686, September 1984.

[88] L-F. Wei. Trellis-coded modulation with multidimensional constellations. *IEEE Transactions on Information Theory*, IT-33(4):483–501, July 1987.

[89] L-F. Wei. Rotationally invariant trellis-coded modulation with multidimensional $M$-PSK. *IEEE Journal on Selected Areas in Communications*, 7:1281–1295, December 1989.

[90] J.M. Wozencraft and I.M. Jacobs. *Principles of Communication Engineering*. John Wiley and Sons, 1965.

[91] Z.C. Zhu and A.P. Clark. Rotationally invariant coded PSK signals. *IEE Proceedings Part F*, 134:43–52, February 1988.

[92] R.E. Ziemer, J.H. Stilwell, and C.R. Ryan. Conversion and matched filter approximations for serial minimum-shift keyed modulation. *IEEE Transactions on Communications*, COM-30(3):495–509, March 1982.