A

Dissertation

On

Efficiency Modeling in the multi path propagation - Time continuous Rayleigh Fading Channel

Submitted to
Department of Electronics & Communication Engineering

Under Guidance of

Dr. (Mrs.) Poonam Sinha
Head of Department EC, & IT & PG Coordinator

in partial fulfillment of the requirements for the degree of

Master of Technology
(Digital Communication)

Submitted by

Rajeev Prakash Shrivastava

Barkatullah University Institute of Technology
Barkatullah University, Bhopal - India
2013

A

Dissertation

On

Efficiency Modeling in the multi path propagation - Time continuous Rayleigh Fading Channel

Submitted to
Department of Electronics & Communication Engineering

Under Guidance of

Dr. (Mrs.) Poonam Sinha
Head of Department EC, & IT & PG Coordinator

in partial fulfillment of the requirements for the degree of

Master of Technology
(Digital Communication)

Submitted by

Rajeev Prakash Shrivastava

Barkatullah University Institute of Technology
Barkatullah University, Bhopal - India
2013

DECLARATION

I hereby declare that the entire work which is presented in this dissertation entitled, "***Efficiency Modeling in the multi path propagation – Time continuous Rayleigh Fading Channel***", is in the partial fulfillment for the award of degree of Masters of Technology in Digital Communication Session 2013. This is authentic record of my own work, carried out under the guidance of **Dr. (Mrs.) Poonam Sinha, Department of Electronics and Communication Engineering, BUIT, Barkatullah University Bhopal.**

The matter embodied in this dissertation has not been submitted by me for award of any other degree or diploma.

Rajeev Prakash Shrivastava
M.Tech (Digital Communication)

CERTIFICATE

This is to certified that the dissertation entitled , "***Efficiency Modeling in the multi path propagation – Time continuous Rayleigh Fading Channel***".

Submitted by Rajeev Prakash Shrivastava to the Department of Electronics and Communication Engineering, BUIT, Barkatullah University Bhopal is a bonafide work carried out by the candidate during the period November 2012 to May 2013 under the guidance of undersigned . It is further certified that the work reported in the dissertation fulfills the requirement of the ordinance related to the degree of **M. Tech in Digital Communication from Barkatullah University Bhopal.**

Guided by

Dr. (Mrs.) Poonam Sinha
Head of Department
EC, & IT & PG Coordinator

Forwarded By

Dr. D. C. Gupta
Director BUIT Bhopal

Director,
University Institute of Technology
Barkatullah University, Bhopal.

ACKNOWLEDGEMENT

This dissertation involves the collection and analysis of information from a wide variety of sources and efforts of many people beyond me. Thus it would not have been possible to achieve the results reported in this document without their help, support and encouragement.

I would like to express my gratitude to the following peoples for their help in the work leading to this project .

Dr. (Mrs.) Poonam Sinha, Head of Department EC, & IT & PG Coordinator : who motivated me both technically and morally for achieving the greater success to complete the dissertation and for her useful comments on the subject matter and for knowledge I gained by sharing ideas with her.

I would like to thank **Dr. (Mrs.) Sunanda Manke,** and **Dr. Neha Tiwari** for extending her continuous help during the tenure of my course.

Dr. D. C. Gupta for providing an environment to complete the dissertation well in time.

I also thank all the staff members of our college and technicians for their help in making the dissertation a success.

Finally, I take this opportunity to extend my deep appreciation to my family and friends for all they meant to me during the crucial times of completion of the dissertation.

ABSTRACT

Frequency flat, fast Rayleigh fading may be considered the most critical disturbance in a wireless communication system. In its most general form, it is modeled as a multiplicative time continuous random (zero mean complex Gaussian) distortion of the transmitted signal. In order to achieve an efficient communication here, each part of the communication link must be carefully designed based on the properties of the time continuous channel. Such an approach is distinct from most previous works, which have used strategies developed for the additive white Gaussian noise channel (AWGN) as starting points.

Crucial in this development is how to deal with the deep fades and the rapid fluctuations of the received signal power. To reduce the influence of the deep fades on the error probability, diversity techniques must be used. Coded interleaved modulation can be regarded one such strategy, where the diversity effect arises as a result of nearby encoder output symbols being subjected to statistically independent fading. Central in achieving this independence is the inter leaver, which spreads the symbols in time. A higher diversity order is obtained if the encoder output bits, instead of symbols, are interleaved. By also using codes of lower rates, the diversity is increased even further. The channel symbol constellation must be expanded accordingly. The resulting system is referred to as bit-interleaved channel symbol expansion diversity (CSED) and is the best low-complexity coded modulation scheme presented so far for the Rayleigh fading channel.

On the fast fading channel, coding is only part of the solution, however. Here, matched filter detectors originally developed for AWGN are unable to efficiently handle the rapid (compared with the signaling rate) fluctuations of the received signal power. More sophisticated solutions are needed. It is essential to make the transition from the time continuous received signal to a discrete representation in the receiver, without losing too much accuracy. Both an adequate number of discrete observables and a sufficiently long observation interval are required. When this is fulfilled, error probability curves with steeper slopes and considerably lower error floors than those of matched filter based receivers are obtained. The steeper slopes are results of the implicit diversity of the random message component in the received signal, which is captured when multiple discrete observables (per symbol interval) are used. This effect is more pronounced the faster the fading is, therefore uncoded signaling has a potentially better performance at fast fading than in slow fading.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Channel Symbol Expansion Diversity (CSED)
Inter Symbol Interference (ISI)
Additive White Gaussian noise (AWGN)
Viterbi Algorithm (VA)
Trellis Coded Modulation (TCM)
Maximum A Posteriori (MAP)

CHAPTER 1

Introduction

1.1 Introduction
1.2 Motivation
1.3 Objective
1.4 Organisation of Dissertation

CHAPTER 1

Introduction

1.1 Introduction
1.2 Motivation
1.3 Objective
1.4 Organisation of Dissertation

CHAPTER 1

1.1 Introduction

During the last decades there has been a tremendous growth in communications all over the world. Public telephone networks have evolved from mostly accommodating analogue voice services to also becoming media for efficient data communications. The modems of today use sophisticated technology, which is based on the knowledge acquired at the absolute front line of science and they operate at data rates close to what is theoretically possible. When even higher data rates are needed, most of the communication takes place over satellite links or fiber optic cables, where transmission at several Gbit/s is a reality. The efficient digital communication links are the foundation on which e.g. ATM-networks (Asynchronous Transfer Mode) are being built. A driving force in this development is multimedia services, which call for highly reliable and flexible transmission rates.

This scenario stands in sharp contrast with existing cellular systems, which are based on technology distant from recent scientific achievements. These systems provide only low-speed data communication at error rates that are far from acceptable in e.g. a wireless ATM connection. For future networks to appear transparent to the user in the sense that the same level of service will be provided regardless the medium of transmission, the mobile communications must become considerably more efficient.

To reach this point, transmission techniques must be improved and it seems likely that wireless communications will develop in the same way as data communication over voice grade telephone lines and ultimately approach the theoretical limits. Such a development provides tremendous challenges for manufacturers as well as researchers in many fields. The special problems associated with the wireless channel can be summarized in two words; multipath propagation.

1.2 Motivation

Propagation of the radio waves over multiple paths between the mobile user and the base station is commonly encountered in cellular systems. The multiple paths, which

result from reflections, are superimposed at the receiving unit. When the arrival times of the different rays are of the same order of magnitude as the duration of the transmitted symbols, successive symbols are smeared together. This effect is often referred to as inter symbol interference (ISI). For paths, where the time difference is comparable to the period of the carrier frequency another effect results. In a general scenario both the ISI and the interference pattern are time-varying due to motion of the transmitter/receiver or of the environment.

The received signal in fading is modeled by a Rayleigh (distributed envelope and a uniformly distributed phase, hence the name Rayleigh fading. It arises, not only in cellular systems, but also e.g. in terrestrial long-distance radio communication and in underwater acoustic communications. Another example is in mobile satellite systems.

The cellular systems of today, e.g. GSM, are designed more to circumvent the problems imposed by the fading than to actually solve them. To avoid too rapid a fluctuation of the received signal power, a sufficiently high signaling rate is chosen. By doing so, the received signal can be regarded approximately constant over a data symbol interval. However, the signaling rate cannot be chosen very high, since that would instead lead to severe problems with ISI. Hence the choice of transmission rate is a compromise, which probably limits the efficiency of the system.

To achieve efficient systems operating at high frequencies in the future, knowledge of how to deal with fading channels will be far more important than when e.g. the GSM system was developed. Current proposals for systems operating at these frequencies are again designed to avoid fast fading by the use of very high symbol rates. However, the lesson learned from the history of the modem development for telephone lines tells us that even the most difficult problems are worth solving in the effort towards achieving the potential performance of the channel.

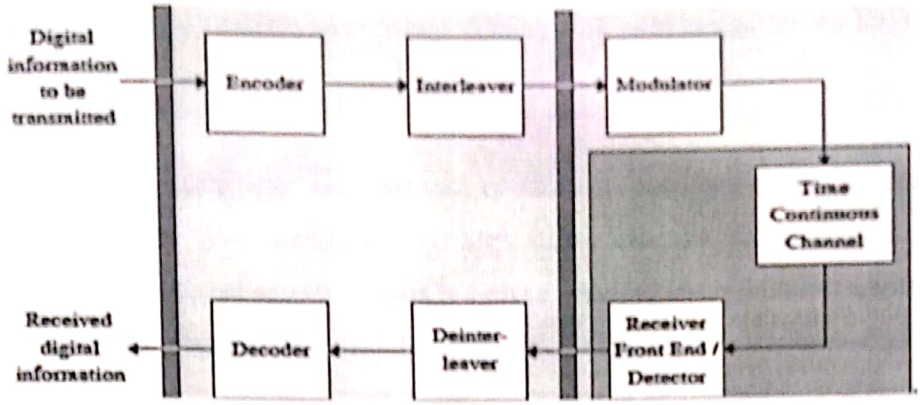
1.3 Objective

An overall aim of this dissertation can be summarized as:

Find an efficient digital communication link for the time continuous Rayleigh fading channel. Efficiency is loosely defined as good detection performance at a reasonably low complexity. To achieve this, each part of the system must be carefully designed

for use on the time continuous Rayleigh fading channel. The special properties of the channel are used as starting points in the design.

A model of a communication link for this channel is shown in figure 1.1



The dissertation describes a step-wise development of a system based on this model. The results obtained and presented there, shall be regarded equally important as those presented for the whole system. Knowledge gained and the techniques developed for those parts are tied together to a communication link for the time continuous Rayleigh channel.

1.4 Organization of Dissertation

Chapter 3 treats the problem of transmission of a single symbol. The focus is on how to deal with the time continuous channel.

Chapter 4 treats sequence detection of uncoded signaling on the Rayleigh fading channel. The optimal detector for a system based on the discretization strategies is derived. In most cases it is prohibitively complex; therefore a simplified suboptimal detector is proposed. Lower as well as (approximate) upper bounds on the error probability for the optimal detector are developed. Those are compared with extensive simulations of the performance of the simplified detector. A main result here is that the simplified detector can be designed to perform close to the optimal one. The derivation of the optimal detector resembles, but while they treat systems based on sampling a more general class of systems are dealt with here. The algorithm derived here is, after some minor modifications, used in the receiver front end.

Design of coded modulation is the subject of chapter 5. The concept of CSED (Channel Symbol Expansion Diversity) is developed. It is based on bitwise interleaving, but it significantly outperforms the scheme. The reason is that CSED enables a further increase of the effective diversity. Except for very slow fading, it provides an inaccurate representation of the fading process. The reason for using it here is that it is widely employed in papers dealing with code design for the Rayleigh channel.

In chapter 6, the results are combined into an efficient coded modulation scheme to be operated on the time continuous Rayleigh fading channel. As in chapter 5, a receiver front end producing soft outputs is derived . A comparison between a systems that is carefully designed for use in Rayleigh fading and a scheme originally developed for AWGN reveals a tremendous difference in performance. It is both the code design and the way the transition between continuous and discrete time is made that gives this difference. The overall system in chapter 6 as well as those in chapters 3 and 4 is found to be rather robust against non- perfect knowledge of channel parameters. This is a desirable property for practical systems, where estimation of those parameters must be made.

Here idealizations made in order keep the problem at a manageable level. Firstly, no distortion and disturbances except the Rayleigh fading and the additive noise are assumed. Those assumptions are unrealistic, since e.g. a perfectly linear power amplifier in the transmitter is required. Another objection is that in a communication system there are other users giving rise to co channel interference. Secondly, perfect knowledge of the first and second order statistics of the fading process is assumed, except for a few cases where robustness is considered. In a real system estimation of the channel parameters must be carried out, which inevitably leads to estimation errors. Thirdly, the common assumptions of perfect symbol synchronization and perfect knowledge of the carrier frequency in the receiver are made.

CHAPTER 2

Review of Literature survey

2.1 Historical notes

2.2 Overview Rayleigh fading channel

CHAPTER 2

REVIEW OF LITRATURE SURVEY

2.1 Historical notes

Scientific work in the area of communication in Rayleigh fading has historically very much been driven by technological demands. The first driving force was the development of military long-distance tropo scatter communication during the fifties and sixties. To reach beyond the horizon, clouds of particles in the troposphere were used as reflectors for the radio waves. The differences in path length between the scattered waves here give rise to Rayleigh fading. Not only the troposphere was used for this purpose, but also the ionosphere.. To create an artificial medium for reflection, a huge amount of very thin and short copper wires acting as electrical dipoles were placed in a belt surrounding the world.

An interesting observation is that many of the leading researchers were employed by industrial companies and that their basic researches were meant as foundations for product development. Even though very much work was performed in the area of detection, the resulting systems were not reliable enough for some computer communications applications. This was the main force behind the introduction of coding theory to signaling on the Rayleigh channel and it led to development of communication equipment.

During the 70´s there seems to have been a drop of the interest in fading channel communications, but in the mid 80´s there was a large increase of scientific activities in the area. Contributions range all the way from coding strategies to sequence detection techniques and some of the results constitute scientific milestones in the area. Successful techniques developed for the additive white Gaussian noise (AWGN) channel during the late 60´s and 70´s like the Viterbi algorithm and trellis coded modulation (TCM) were adapted for use on the Rayleigh fading channel during this period.

2.2 Overview Rayleigh fading channel

Over the last 50 years a tremendous amount of work has been presented about communication over the Rayleigh fading channel. It seems that much of the research

has used techniques developed for the additive white Gaussian noise channel as starting points. Those strategies have then been adapted to the Rayleigh channel, or rather to some model of the channel. This has in many cases put unnecessary restrictions on the system design and has led to rather poor performances. Considerably less works use the Rayleigh fading channel itself as a starting point.

The aim is to give an overview of some of the scientific work presented about communication over the Rayleigh fading channel. The focus here is on landmark papers especially in the areas of coding and detection and more detailed references will be given later in the Dissertation. As mentioned above, there are three main effects caused by the fading. Those are the deep fades of the signal power, time varying distortion and inter symbol interference. Even though the cause of them is the same, different strategies are required to deal with each of the impairments. To handle the first one, i.e. the deep fades, diversity techniques are needed. The principle is that a certain information bit shall reach the receiver via different paths. The paths can e.g. be different antennae (space diversity), different frequencies (frequency diversity), different times (time diversity) or combinations thereof. When one of the paths is in a deep fade, the other ones may not, giving a higher resistance against the fades than if only a single path was used.

A fair amount of research has been made on different diversity techniques and especially how to combine the diversity paths in the receiver. Now consider time diversity, which is of great interest. The simplest way to achieve this is to transmit the same information bit at several sufficiently spaced times, i.e. to use what is usually referred to as repetition coding. This is a rather inefficient type of code for the AWGN channel, but it gives a relatively good performance in comparison with simple codes developed for AWGN on the Rayleigh channel. While the information bits associated with a destroyed code word cannot be recovered, there is a high probability that at least one of the repetitions in the repetition code is not destroyed by a fade.

Early works considered mostly the problem of detection of a single symbol, but as a result of the revolution, sequence detection on the Rayleigh fading channel was proposed.

CHAPTER 3

Single Symbol Signaling

CHAPTER 3

Single Symbol Signaling

3.1 Fading Channel Characteristics

The degree of distortion of the received signal in a mobile communication system generally varies in time. A coarse, but technically relevant characterization of this is obtained by making a distinction between long-term and short-term effects. Long-term effects are e.g. successive attenuation of the received signal as the mobile travels from the base-station (i.e. the other end of the link) or shadowing by buildings, trees or other obstacles. The long-term variations are very slow compared with the transmission rate and can often be considered constant, when the focus is on design or analysis of coding and detection strategies. That is the case in this, Dissertation The short-term variations, Characterization of the short-term effects is the topic of this chapter.

In a mobile radio system, reflections in e.g. buildings or cars lead to a situation where several incoming paths of transmission are superimposed at the receiver. This gives in general rise to both Rayleigh fading and inter symbol interference. The Rayleigh fading is a result of superposition of incoming waves of comparable strengths, but with different carrier phases. This gives a spatial interference pattern. Sometimes there is a direct path, which is considerably stronger than the reflected rays. The received signal is now a combination of that direct component and a Rayleigh fading part. This situation is often referred to as Ricean fading.

3.2 Single Symbol Signaling

As a first basic step towards the design of efficient communication schemes for the Rayleigh fading channel, transmission of a single symbol is studied. This is important in order to gain an understanding of how to deal with the time continuous received signal. In AWGN (Additive White Gaussian Noise), a transition to a discrete representation is made easily without loss of generality by introduction of a finite dimensional signal space description of the received signal. This is useful both for purposes of design and analysis. For instance it leads straightforwardly to the optimal

detector, which can be based on matched filters. What makes the transition from the time continuous signal to a discrete model so simple in AWGN is that the message waveforms can be regarded deterministic. This is in contrast to signaling on the Rayleigh fading channel, where it is appropriate to consider the message components of the received signal being Gaussian random processes. As a result, the detection problem on the Rayleigh fading channel becomes fundamentally different from that in AWGN. It can be generalized to: Detection of continuous time Gaussian random message processes in Gaussian noise. A common assumption, made here as well, is that the second order statistics are perfectly known to the receiver.

More recently, research has been performed mostly under the slow fading assumption, where the time variation of the fading process is assumed to be negligible over a symbol interval. In such cases matched filter detectors developed for AWGN are expected to give good performances . Sometimes such detectors are for some reason even used in fast fading. One example of this can be found, where a matched filter detector not only is used but also is erroneously claimed to be optimal on the fast fading channel. Simulation results indicate a potential performance improvement to be achieved by the use of multiple discrete observables per symbol interval. Still no thorough investigations of the discretization problem have been performed and the theoretical side of the problem is barely mentioned.

As in the publications contain very limited principal performance analyses. Only a few special channels for which analytical expressions on the error probabilities can be found are treated. An attempt to analyze the performance of DPSK on the frequency flat Rayleigh fading channel is made. The system model used there is incorrect, however, since the receiver noise is band limited while the message components of the received signal have infinite bandwidth.

This makes the results in the paper unreliable. Still the approach to the performance analysis is interesting.

This chapter presents and analyzes a broad class of low complexity single symbol detectors for the frequency flat Rayleigh fading channel. The detectors are developed with the above detection problem as a starting point. Contrary to the optimal detector,

a discretization of the received signal precedes the actual detection. This leads to a loss of information on the Rayleigh channel and makes the class of detectors suboptimal. The detector class is very general and comprises most of the solutions as special cases, but also several interesting easily implemented detectors. Error probabilities are computed numerically with high precision.

The focus is on investigation of different discretization strategies and thorough explanations of effects like error floors and implicit diversity are provided. Comparisons between matched filter detectors and more sophisticated solutions are also made. It is found that detectors only slightly more complex than those based on matched filters give far better performances.

3.3 Summary

As an initial step towards an efficient digital communication system for the time continuous frequency flat Rayleigh fading channel, detection of a single symbol was treated. The problem was formulated as choosing between M times continuous Gaussian zero mean random processes, the message processes, in additive white Gaussian noise. The first and second order statistics were assumed to be known to the receiver. A broad class of receivers was introduced. They consist of a front end, which produces a finite number of discrete observables of the received signal. Given those, an ML decision is made. It was found to be useful to view the discrete observables as a vector in a receiver signal space of finite dimensionality. Here it is important that contrary to the signal space representation for the AWGN channel, this representation can only be an approximate description of the received signal.

For such receivers, several properties were investigated using numerical symbol error probability calculations. The focus was on how the discretization in the receiver affects the error probabilities. Simple discretization methods were found to give excellent performances provided the number of observables was large enough and the observation interval was sufficiently long. Two effects were noticed when increasing the number of observables or the length of the observation interval. Firstly the level of the error floor (when such existed) was lowered and secondly there was an implicit diversity. The latter is a result of the received signal being a random process and leads

to successively steeper error probability curves as the average signal to noise ratio increases. These results form the basis for most of the chapters to follow.

In many papers about detection on the fast fading channel, matched filters are used believing this is the optimal solution. In some other papers the discrete time channel model stemming from the use of matched filters in the slow fading case is uncritically employed also for fast fading. Clearly, a discrete time model does not reflect the choice of modulator waveforms and is for that reason practically useless. Also, since the matched filter receiver creates a receiver space of minimum (for the Gaussian channel) dimensionality, the use of such a receiver results in low implicit diversity and high error floor. Its performance becomes worse as the bandwidth of the fading process increases and it was found that for normalized fading bandwidths above fn=0.01, the losses compared to detectors using a larger number of observables were high.

It was also observed that signaling at large fading bandwidths gives lower error probabilities than signaling at small fn, provided a sufficiently good detector is used. Thus rather surprisingly, making T larger by using pulses with longer duration will give a better performance on the Rayleigh channel. By a moderately increased detector complexity (compared to a matched filter detector), fast fading is not destructive, rather it improves the performance. The implicit diversity is the reason for this.

The concept of a sufficient statistic is central in detection theory. It is hard to actually find a definition of that terminology, but it is used for a function of the received signal which without loss of optimality is easier to use in a receiver than the received signal itself. On the Gaussian channel it is well known that the outputs of matched filters sampled once every symbol interval is a sufficient statistic. Definition of the term sufficient statistic is given, it does not make sense to use it for anything but for the time continuous received signal.

CHAPTER 4

Sequence Detection

CHAPTER 4

Sequence Detection

Single symbol signaling on the time continuous Rayleigh fading channel was treated. How the transition from the received signal to a discrete representation in the receiver is made was found to be crucial to the performance. Both an increased number of observables and an extension of the observation interval led to large performance improvements.

In this chapter detection of a sequence of uncoded symbols, instead of a single symbol, is considered. There are several reasons for doing this. First of all, in a practical communication system sequences and not only single symbols are transmitted. Extension of the observation interval beyond the symbol interval immediately involves other symbols in the decision. Sequence detection is one way of dealing with this, symbol-bysymbol MAP (Maximum A Posteriori) detection is another . Secondly, the fading randomly changes the phase of the transmitted signal, so that it becomes unknown to the receiver. One strategy to handle this is to insert reference symbols that are known to the receiver among the channel symbols. Sequence detection can be employed to utilize the additional information supplied by the reference symbols.

The covariance of the fading process is assumed to be known to the receiver. That is in contrast to, where perfect knowledge of the fading process itself is assumed, giving a detection problem only slightly different from sequence detection on the AWGN channel.

The estimator computes an MMSE (Minimum Mean Square Error) estimate of the fading in a certain symbol interval by using previous symbol decisions. The decision rule is then adapted depending on this estimate.

The most general receiver structure for fading channels was proposed by Morley and Snyder, By using the generalized likelihood ratio formula given by Kailath, a receiver operating directly on the time continuous received signal is obtained. In its most

general form, it evaluates the (generalized) log likelihood for each possible symbol sequence. By also applying a state space description of the received process, a recursive detector combining Kalman filters with the Viterbi algorithm is obtained. The complexity of receiver is now highly dependent on the length of the correlation in the fading process. This length determines both the number of states and how many symbol intervals of the received signal is needed in the recursive computation.

The main difference is that this detector works on samples of the received process. One Kalman filter per sequence alternative is suggested in the general case (including frequency selective fading). That solution is also proposed, but only for slow fading. The main contribution is the detector structure obtained for CPM (Continuous Phase Modulation on the frequency flat Rayleigh fading channel. When the fading process is assumed to have a finitely long correlation, that detector can be implemented by a set of linear predictors combined with the Viterbi algorithm. Linear predictors require considerably less computations than Kalman filters; therefore this receiver is much less complex than the general receiver.

Several authors have expanded upon the results; instead of CPM all band limited signals are allowed. To lower the average complexity, sequential detection is used instead of the VA.

An interesting and elegant approach was taken by a linear transformation of the received discrete process; a sequence of independent Gaussian random variables is obtained. That sequence is the innovations process of the received signal, and it directly leads to a recursive detector for general fading channels. The predictor based detector can be shown to be a special case of this receiver, obtained for the frequency flat channel.

A problem encountered when signaling on the Rayleigh fading channel is the random phase introduced in the received signal by the fading process. It makes it impossible to separate the signs of the fading process from those of the symbol sequence and results in that two sequences having symbols of opposite signs in all symbol intervals cannot be distinguished from each other. The methods always lead to an increased bandwidth. The most common method of dealing with the random phase in the

received signal is differential encoding. By doing this, the phase difference between successive symbols and not the phase itself is information bearing, therefore no external reference signals are needed. This in turn implies that the bandwidth is not expanded by this strategy. The drawback is, however, an approximately doubled error probability (for MPSK). This can be a serious problem on the Rayleigh fading channel, where the error probabilities are not decreasing as fast as on the AWGN channel (with increasing SNR).

The traditional way of exploiting the reference information supplied by the pilot symbols is to estimate the fading process. The fading is measured at the known symbol intervals and interpolation is used to obtain estimates of the fading process during the other symbol intervals. These estimates are then employed in a conventional detector for the AWGN channel. This is a heuristic approach and such systems are suboptimal in terms of minimum error probability. Another strategy is to include the different a priori probabilities induced by the pilot symbols in the detection problem, i.e. to formulate it as MAP sequence detection (or MAP symbol detection). Now the detector giving minimum probability of error can be derived.

True union bounds on the error probabilities for the block-based detector are evaluated for slowly fading time-dispersive channels when differential encoding is used. Even though the block-structure implies that only a limited number of pairwise error probabilities are needed, only small blocks can be considered. The reason is that the number of pairwise error probabilities is exponentially increasing with the block-length, and hence the computational complexity soon becomes prohibitively large.

Here MAP sequence detection for the time continuous frequency flat Rayleigh fading channel is treated. A transition from continuous time to a discrete representation is made in the same general way, the treatment here may be regarded a generalization of the works, where the discretization problem is barely mentioned. Another difference is that both causal and anti-causal detectors are derived here, while the above works only treat the causal case. The reason for allowing the detectors to be anti-causal is to enable use of the forward (in time) correlation of the fading process. This could lead to further performance improvements.

The detector derived here is well suited for discretization schemes, which produce a vector of discrete observables at the end of each symbol interval instead of a continuous stream of samples . Starting with manipulation of the likelihood function a receiver structure is obtained. With the optimal detector as a starting point, efficient suboptimal detectors are devised. These detectors are based on the least constrained breadth first MAP sequence detector (MAP under complexity constraint) . This is in contrast to, where the Viterbi algorithm is used , where further constraints are imposed. Both lower and upper bounds on the error probability of the optimal detector are derived. While the derivation of the upper bound bare , the lower bound has not previously been derived. Extensive computer simulations are also presented together with the bounds to illustrate various aspects of the detectors.

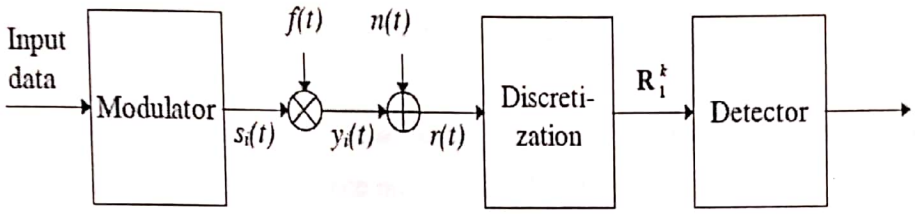


Figure 4.1 System model.

4.1 System description

The communication system in figure 4.1 conveys one of Mk sequences of k uncoded equally likely M-ary modulator symbols from time 0 to kT. Such a sequence will henceforth be referred to as the message. This is a difference compared , where a single symbol was called a message. Let the vector

$$\mathbf{s}_1^k(i) = \left[s_1(i), s_2(i), \ldots, s_k(i)\right]^T, \quad i \in \left[1, 2, \ldots, M^k\right] \qquad (4.1)$$

denote message i, where Sk(i) is the modulator symbol to be transmitted in the time interval $(k - 1)T \leq t < kT$.

The corresponding time continuous signal is Si(t). Discretization in the receiver front end yields the random vector of

observables $\mathbf{R}_1^k = (\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_k^T)^T$, where $\mathbf{r}_k = (r_{k,1}, \ldots, r_{k,N})^T$ contains the N observables obtained in the k:th symbol interval. $\mathbf{R}_1^k$ has dimension kN and can be expressed as

$$\mathbf{R}_1^k = \mathbf{Y}_1^k(i) + \mathbf{N}_1^k \tag{4.2}$$

where $\mathbf{Y}_1^k(i)$ and $\mathbf{N}_1^k$ are the discrete counterparts[5] of $y_i(t)$ and $n(t)$ over the interval $0 \le t < kT$. Conditioned a certain message $\mathbf{s}_1^k(i)$, the observables in $\mathbf{R}_1^k$ are jointly complex Gaussian with zero mean and density function

$$p_{\mathbf{R}|S}\left(\Theta_1^k \middle| \mathbf{s}_1^k(i)\right) = \frac{1}{\pi^{kN}\left|\mathbf{C}_1^k(i)\right|} e^{-(\Theta_1^k)^H \cdot (\mathbf{C}_1^k(i))^{-1} \cdot \Theta_1^k} \tag{4.3}$$

The conditional covariance matrix is

$$\mathbf{C}_1^k(i) = E\left[\mathbf{R}_1^k \left(\mathbf{R}_1^k\right)^H \middle| \mathbf{s}_1^k(i)\right] = E\left[\mathbf{Y}_1^k(i) \cdot \left(\mathbf{Y}_1^k(i)\right)^H\right] + N_0\mathbf{I} \tag{4.4}$$

where I is the identity matrix. It is in some cases possible to separate the message and fading dependence in the covariance matrix according to

$$\mathbf{C}_1^k(i) = \mathbf{S}_1^k(i) \cdot \mathbf{C}_{\mathbf{F}1}^{\ k} \cdot \left(\mathbf{S}_1^k(i)\right)^H + N_0\mathbf{I}, \tag{4.5}$$

where C_{F1}^k is the covariance associated with the fading process and S_{Fi1}^k (i) is an Nk by Nk diagonal matrix with entries Sn(i) depending only on the message i. This situation arises either when rectangular modulator waveforms are transmitted and discretization is performed or when sampling is used for arbitrary modulator waveforms. In the latter case, Sn(i) (n=1, 2, ..., kN) are just samples of si(t) and C_{F1}^k is based on samples of the fading process. The first case, however, is slightly more complicated. The l:th observable in the j:th symbol interval can be written

$$\begin{aligned} r_{j,l} &= \int_{(j-1)T}^{jT} s_i(t) \cdot f(t) \cdot \varphi_l(t-(j-1)T)dt + \int_{(j-1)T}^{jT} n(t) \cdot \varphi_l(t-(j-1)T)dt = \\ &= a_j(i) \cdot \int_{(j-1)T}^{jT} f(t) \cdot \varphi_l(t-(j-1)T)dt + \int_{(j-1)T}^{jT} n(t) \cdot \varphi_l(t-(j-1)T)dt. \end{aligned} \tag{4.6}$$

(φ)t is one of N orthonormal functions of the ON-set Φ_N used for discretization . In eq. (4.6) it is possible to move the message dependence outside the integral, since rectangular modulator waveforms are assumed, i.e. S_i (t) = a_j (i) The entry of the covariance matrix corresponding to r_{jl} and r_{mn} then becomes

$$c_{j,l;m,n} = E\left[r_{j,l} \cdot r^*_{m,n}\right] =$$

$$= a_j(i)a_m(i)^* \int_{(j-1)T}^{jT} \int_{(m-1)T}^{mT} C_f(u,v)\varphi_l(u-(j-1)T)\varphi_n(v-(m-1)T)dudv + N_0\delta(j,l;m,n). \tag{4.7}$$

δ(j,l:m,n) is equal to one only when j=m and l=n and zero otherwise. The corresponding element of the covariance matrix C_{F1}^k k , which now depends only on the fading process is

$$c_{Fj,l;m,n} = \int_{(j-1)T}^{jT} \int_{(m-1)T}^{mT} C_f(u,v)\varphi_l(u-(j-1)T)\varphi_n(v-(m-1)T)dudv . \tag{4.8}$$

The message components in eq. (4.7) are a_j(i) and a_{mj}(i) respectively, therefore S_n(i) = a_j(i) for n=jN+1, ..., (j+1)N. This property, i.e. the possibility to decompose the covariance matrix according to eq. (4.5), is important for the rest of the Dissertation. Its utility lies in that it enables a significant complexity reduction of the detectors derived of the receiver front end developed .

The above notation is given on a per symbol interval basis and is suitable when discretization by ON-sets is employed. The reason is that a vector of observables are r_k and is produced and passed to the detector at the end of each symbol interval. When sampling is used, on the other hand, the observables are presented to the detector one by one. A per observable notation is therefore also desired. Let

$$\mathbf{R}_1^{k,l} = \left[r_{1,1}, r_{1,2}, \ldots, r_{1,N}, r_{2,1}, \ldots, r_{k,1}, \ldots, r_{k,l}\right]^T \tag{4.9}$$

be the random vector containing the Nk+l "first" observables and

$$p_{\mathbf{R}|S}\left(\Theta_1^{k,l}\middle|\mathbf{s}_1^k(i)\right) = \frac{1}{\pi^{kN+l}\left|\mathbf{C}_1^{k,l}(i)\right|} e^{-\left(\Theta_1^{k,l}\right)^H \cdot \left(\mathbf{C}_1^{k,l}(i)\right)^{-1} \cdot \Theta_1^{k,l}} \tag{4.10}$$

be the conditional density function. $C_j^{k,l}$(i) is the corresponding covariance matrix. Whenever this alternative notation is used in the following, it will be explicitly indicated. The notation will be used for the arguments of the probability density functions in eq. (4.3) and (4.10) as for the random vectors themselves. Again, this is

not a mathematically strict notation, but it will be used in order to simplify the presentation.

4.2 Problem statement

Even when uncoded signaling is used on the Rayleigh fading channel, sequence detection is desired. The purpose is to achieve a good performance by exploiting the correlation in the fading process over more than one symbol interval. Another reason for using sequence detection is to optimally (in the sense of minimum probability of sequence error) deal with the additional information supplied when pilot symbol assisted modulation (PSAM) is used.

When pilot symbols are inserted in the transmitted symbol stream, the apriori symbol probabilities are no longer constant. Because of this, MAP, instead of ML sequence detection should be performed. The decision rule for the MAP sequence detector is:

Find the message maximizing the decision function,

$$P\left(\mathbf{s}_1^k(i)\middle|\mathbf{R}_1^{k+d}\right), \quad i \in \left[1,2,\ldots,M^k\right]. \tag{4.11}$$

Contrary to the approaches where only the case d=0 is dealt with, anti-causal (d>0) detectors are allowed here. The main reason for taking this new approach is that the forward (in time) correlation of the fading process can be exploited. Another reason is that the MAP symbol by symbol detector is easily developed as an extension of the MAP sequence detector obtained by this approach.

4.3 The optimal detector

The statistical properties of the received signal is in general dependent on the entire message , therefore the optimal detector must evaluate eq. (4.11) for each of the M^k possible messages. The decision function for the i:th message (eq. (4.11)) is proportional to

$$P\left(\mathbf{s}_1^k(i)\middle|\mathbf{R}_1^{k+d}\right) \propto P\left(\mathbf{s}_1^k(i)\right)p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^k(i)\right). \tag{4.12}$$

A way to recursively compute this is desired. This is achieved by dealing with the discrete time innovations process of R_l^{k+d} instead of the received vector itself. One

(conditional) innovations process for each possible message is needed, since the statistical properties of the received signal is message dependent. Those are computed by linear transformations of the received vector and can be viewed as whitening operations of the received sequence. This leads straightforwardly to a way of recursive computation of (4.12). Even though an elegant detector structure is obtained, the more straightforward will be followed and extended here. The reason for this is to present a more traditional derivation, where manipulation of the likelihood function leads to a feasible detector structure.

To achieve a recursive computation of the decision function, let first Ω_k (i) denote the logarithm of eq. (4.12). It can be expressed as

$$\Omega_k(i) = \log\left(P\left(\mathbf{s}_1^k(i)\right)p\left(\mathbf{R}_1^k\middle|\mathbf{s}_1^k(i)\right)\right) + \log\left(p\left(\mathbf{R}_{k+1}^{k+d}\middle|\mathbf{s}_1^k(i).\mathbf{R}_1^k\right)\right) = \Gamma_k(i) + \Lambda_k(i) \quad (4.13)$$

Note that when d=0, $\Omega_k(i) = \Gamma_k(i)$. Now $\Gamma_k(i)$ can be written

$$\Gamma_k(i) = \log\left(P\left(\mathbf{s}_1^{k-1}(i)\right)p\left(\mathbf{R}_1^{k-1}\middle|\mathbf{s}_1^{k-1}(i)\right)\right) + \log\left(P\left(s_k(i)\right)p\left(\mathbf{r}_k\middle|\mathbf{R}_1^{k-1}.\mathbf{s}_1^k(i)\right)\right) = \quad (4.14)$$
$$= \Gamma_{k-1}(i) + \gamma_k(i).$$

Obviously Γ_k(i) can be recursively computed from Γ_{k-1}(i) and it will therefore be referred to as the path metric of message i at time k.

The second term in eq. (4.13), . Λ_k(i) can be expanded and written

$$\Lambda_k(i) = \log\left[\sum_{\text{all } \mathbf{s}_{k+1}^{k+d}(j)} \prod_{l=1}^{d} P\left(s_{k+l}(j)\right)p\left(\mathbf{r}_{k+l}\middle|\mathbf{R}_1^{k+l-1}.\mathbf{s}_1^k(i).\mathbf{s}_{k+1}^{k+l}(j)\right)\right]. \quad (4.15)$$

In contrast to the path metric, Λ_k(i) is not directly needed in the next symbol interval and does not contribute to the recursion. Instead it is merely used in order to find the most probable symbol sequence at time k.

it is recognized that it is a Gaussian density function

$$p\left(\mathbf{r}_k\middle|\mathbf{R}_1^{k-1}.\mathbf{s}_1^k(i)\right) = \prod_{l=1}^{N} p\left(r_{k,l}\middle|\mathbf{R}_1^{k,l-1}.\mathbf{s}_1^k(i)\right) = \prod_{l=1}^{N} \frac{1}{\pi\cdot\sigma_{k,l}^2(i)} e^{-\frac{\left|r_{k,l}-\mu_{k,l}(i)\right|^2}{\sigma_{k,l}^2(i)}} \quad .(4.16)$$

(conditional) innovations process for each possible message is needed, since the statistical properties of the received signal is message dependent. Those are computed by linear transformations of the received vector and can be viewed as whitening operations of the received sequence. This leads straightforwardly to a way of recursive computation of (4.12). Even though an elegant detector structure is obtained, the more straightforward will be followed and extended here. The reason for this is to present a more traditional derivation, where manipulation of the likelihood function leads to a feasible detector structure.

To achieve a recursive computation of the decision function, let first $\Omega_k(i)$ denote the logarithm of eq. (4.12). It can be expressed as

$$\Omega_k(i) = \log\left(P\left(\mathbf{s}_1^k(i)\right)p\left(\mathbf{R}_1^k\middle|\mathbf{s}_1^k(i)\right)\right) + \log\left(p\left(\mathbf{R}_{k+1}^{k+d}\middle|\mathbf{s}_1^k(i),\mathbf{R}_1^k\right)\right) = \Gamma_k(i) + \Lambda_k(i) \quad (4.13)$$

Note that when d=0, $\Omega_k(i) = \Gamma_k(i)$. Now $\Gamma_k(i)$ can be written

$$\begin{aligned}\Gamma_k(i) &= \log\left(P\left(\mathbf{s}_1^{k-1}(i)\right)p\left(\mathbf{R}_1^{k-1}\middle|\mathbf{s}_1^{k-1}(i)\right)\right) + \log\left(P\left(s_k(i)\right)p\left(\mathbf{r}_k\middle|\mathbf{R}_1^{k-1},\mathbf{s}_1^k(i)\right)\right) = \\ &= \Gamma_{k-1}(i) + \gamma_k(i).\end{aligned} \quad (4.14)$$

Obviously $\Gamma_k(i)$ can be recursively computed from $\Gamma_{k-1}(i)$ and it will therefore be referred to as the path metric of message i at time k.

.

The second term in eq. (4.13), . $\Lambda_k(i)$ can be expanded and written

$$\Lambda_k(i) = \log\left[\sum_{\text{all } \mathbf{s}_{k+1}^{k+d}(j)} \prod_{l=1}^{d} P\left(s_{k+l}(j)\right)p\left(\mathbf{r}_{k+l}\middle|\mathbf{R}_1^{k+l-1},\mathbf{s}_1^k(i),\mathbf{s}_{k+1}^{k+l}(j)\right)\right]. \quad (4.15)$$

In contrast to the path metric, $\Lambda_k(i)$ is not directly needed in the next symbol interval and does not contribute to the recursion. Instead it is merely used in order to find the most probable symbol sequence at time k.

it is recognized that it is a Gaussian density function

$$p\left(\mathbf{r}_k\middle|\mathbf{R}_1^{k-1},\mathbf{s}_1^k(i)\right) = \prod_{l=1}^{N} p\left(r_{k,l}\middle|\mathbf{R}_1^{k,l-1},\mathbf{s}_1^k(i)\right) = \prod_{l=1}^{N} \frac{1}{\pi\cdot\sigma_{k,l}^2(i)} e^{-\frac{\left|r_{k,l}-\mu_{k,l}(i)\right|^2}{\sigma_{k,l}^2(i)}} \quad (4.16)$$

This was originally derived for a detector operating on samples of the received signal, therefore the alternative per observable notation is employed.

Instead of exploiting the above property, which suggests a detector based on a set of linear predictors for each symbol sequence, the same starting point for computation of $p(\mathbf{r}_k|\mathbf{R}_1^{k-1},\mathbf{s}_1^k(i))$ will be taken. By application of Bayes theorem , the likelihood function [5] can be written

$$p\left(\mathbf{r}_k\middle|\mathbf{R}_1^{k-1},\mathbf{s}_1^k(i)\right)=\frac{\left|\mathbf{C}_1^{k-1}(i)\right|}{\pi^N\cdot\left|\mathbf{C}_1^k(i)\right|}e^{\left(\mathbf{R}_1^{k-1}\right)^H\cdot\left(\mathbf{C}_1^{k-1}(i)\right)^{-1}\cdot\mathbf{R}_1^{k-1}-\left(\mathbf{R}_1^k\right)^H\cdot\left(\mathbf{C}_1^k(i)\right)^{-1}\cdot\mathbf{R}_1^k} \quad (4.17)$$

The derivations in principle finish with this expression and the resulting detector must compute the quadratic forms in the exponent for each of the M^k possible messages. To proceed further, note that since the conditional covariance matrices are symmetric and positive definite they can either be diagonalized or triangularized in order to simplify the computation of the exponent in eq. (4.17). Here the latter will be employed, where Cholesky factorization enables the decomposition of the condi tional covariance matrices to products of lower $(\mathbf{L}_{\mathbf{c}_1}^k(i),\mathbf{L}_{\mathbf{c}_1}^{k-1}(i))$ and upper triangular matrices according to

$$\mathbf{C}_1^k(i)=\mathbf{L}_{\mathbf{c}_1}^k(i)\cdot\left(\mathbf{L}_{\mathbf{c}_1}^k(i)\right)^H \quad (4.18)$$

and

$$\mathbf{C}_1^{k-1}(i)=\mathbf{L}_{\mathbf{c}_1}^{k-1}(i)\cdot\left(\mathbf{L}_{\mathbf{c}_1}^{k-1}(i)\right)^H. \quad (4.19)$$

The reason for choosing triangularization instead of diagonalization is that it leads to less operations on the vector of observables. In both cases the vector is pre-multiplied by some atrix ($\mathbf{L}_{\mathbf{c}_1}^k(i)$ in the triangular case). When this matrix is triangular, half the number of computations is needed compared with the case when diagonalization is used.

Now let . ξ_k(i) denote the exponent in eq. (4.17). it can be written

$$\xi_k(i) = -\left\| \mathbf{L}_{Ck}(i)^{-1} \cdot \left(\mathbf{r}_k - \mathbf{l}_{Ck}(i) \cdot \left(\mathbf{L}_{C1}^{k-1}(i) \right)^{-1} \cdot \mathbf{R}_1^{k-1} \right) \right\|^2 \qquad (4.20)$$

where $\|\cdot\|$ is the (complex vector) norm and

$$\mathbf{L}_{Ck}(i) \cdot \left(\mathbf{L}_{Ck}(i) \right)^H = \mathbf{C}_k(i) - \mathbf{c}_k(i) \cdot \left(\mathbf{C}_1^{k-1}(i) \right)^{-1} \cdot \left(\mathbf{c}_k(i) \right)^T = \Sigma_k(i). \qquad (4.21)$$

The N by $(k\text{-}1)N$ matrix $\mathbf{l}_{Ck}(i)$ is the lower left submatrix of the lower triangular matrix $\mathbf{L}_{C1}^{k}(i)$.

The N by (k-1)N matrix l_k(i) is the lower left submatrix of the lower triangular matrix , L_k(i). Note that even though Cholesky factorization is a rather complex operation to be implemented in a digital receiver, that is no large problem here. The reason is that it can be made in advance for the specific channel in use.

$$\mu_k(i) = \mathbf{l}_{Ck}(i) \cdot \left(\mathbf{L}_{C1}^{k-1}(i) \right)^{-1} \cdot \mathbf{R}_1^{k-1} \qquad (4.22)$$

After subtraction from r_k the new vector is passed through another filter bank, containing N FIR filters with tap-weights being the elements in the rows of $L_{ck}\,(i)^{-1}$, ξ_k(i) is then obtained as the norm of the N-dimensional vector at the filter bank output.

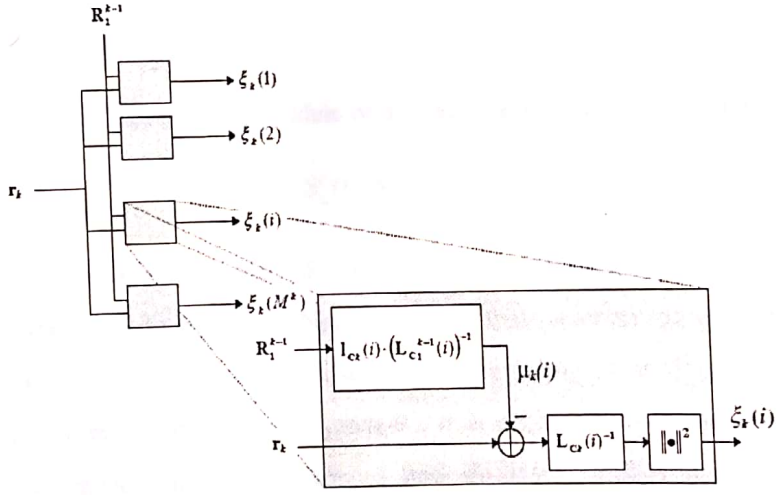


Figure 4.2 Computation of eq. (4.20).

The likelihood function in eq. (4.17) can be expressed as

$$p\left(\mathbf{r}_k \middle| \mathbf{R}_1^{k-1}, \mathbf{s}_1^k(i) \right) = \frac{1}{\pi^N \cdot \left| \Sigma_k(i) \right|} e^{-(\mathbf{r}_k - \mu_k(i))^H \cdot \Sigma_k(i)^{-1} \cdot (\mathbf{r}_k - \mu_k(i))}. \qquad (4.23)$$

This equation and eq. (4.16) look very similar, but the important difference. Here a vector of observables, r_k is treated, whereas eq. (4.16) deals with only one observable at a time. Even though this approach leads to a slightly more complex solution (eq. (4.20)), due to the multiplication by $L_{ck}(i)^{-1}$, it is still attractive. The reason is that itis well suited for discretization schemes which outputs a vector of observables instead of a sequence of single observables. When such a vector is fed to the detector at the end of the k:th symbol interval, only subtraction of μ_k(i) k ()i and multiplication by $L_{ck}(i)^{-1}$, are required to obtain the N-dimensional vector for which the norm is computed. All other computations can be made in advance during the symbol interval. Consequently, this gives a small decision delay.

Both this approach and the algorithms perform different operations for each of the M^k possible messages S_l^k(i), therefore the number of different filter banks is exponentially increasing in time. In many cases, however, this can be avoided. This is because the message dependence can be factored out of the covariance matrix

$$\mathbf{C}_1^k(i) = \mathbf{S}_1^k(i)\cdot\mathbf{C}_{\mathbf{r}_1}^{-1}\cdot\left(\mathbf{S}_1^k(i)\right)^H + N_0\mathbf{I} = \mathbf{S}_1^k(i)\cdot\left(\mathbf{C}_{\mathbf{r}_1}^{-1} + N_0\left(\left(\mathbf{S}_1^k(i)\right)^H\cdot\mathbf{S}_1^k(i)\right)^{-1}\right)\cdot\left(\mathbf{S}_1^k(i)\right)^H \cdot \tag{4.24}$$

Whenever $(\mathbf{S}_1^k(i))^H\mathbf{S}_1^k(i)$ is independent of the message $s_1^k(i)$ eq. (4.24) can be written

$$\mathbf{C}_1^k(i) = \mathbf{S}_1^k(i)\cdot\mathbf{M}_1^k\cdot\left(\mathbf{S}_1^k(i)\right)^H \tag{4.25}$$

where M_l^k(i) , now is independent of S_l^k(i) . This is recognized and exploited but only for constant envelope signals. Note, however, that factorization according to eq. (4.25) is possible in more general cases. When sampling is used (at the Nyquist frequency of the band limited noise process), it is achieved for modulator symbols with equal energies, equal pulse shapes and no inter symbol interference. For discretization by eq. (3.5), a further requirement is that rectangular modulator waveforms must be used.

Consider again eq. (4.25). Here M_l^k(i) is in fact the conditional covariance matrix of $S_l^k(i)^{-1}$ R_l^k. and is therefore positive definite, symmetric and real valued. The

computation of ξ_k(i) can then be simplified in very much the same way as above. The only difference is that the Cholesky factorizations according to equations (4.18), (4.19) and (4.21) are now applied to the matrices M_l^k, M_l^{k-1} and Σ_{mk}. (the latter matrix is defined similar ways as Σ_k(i) in eq. (4.21)). By letting L_{Ml}^k L_{Ml}^{k-1} and L_{Mk} denote the corresponding lower triangular matrices, . ξ_k(i) becomes

$$\xi_k(i) = -\left\| \mathbf{L}_{Mk}^{-1} \cdot \left((\mathbf{S}_k(i))^{-1} \cdot \mathbf{r}_k - \mathbf{l}_{Mk} \cdot (\mathbf{L}_{M1}^{k-1})^{-1} \cdot (\mathbf{S}_1^{k-1}(i))^{-1} \cdot \mathbf{R}_1^{k-1} \right) \right\|^2 . \qquad (4.26)$$

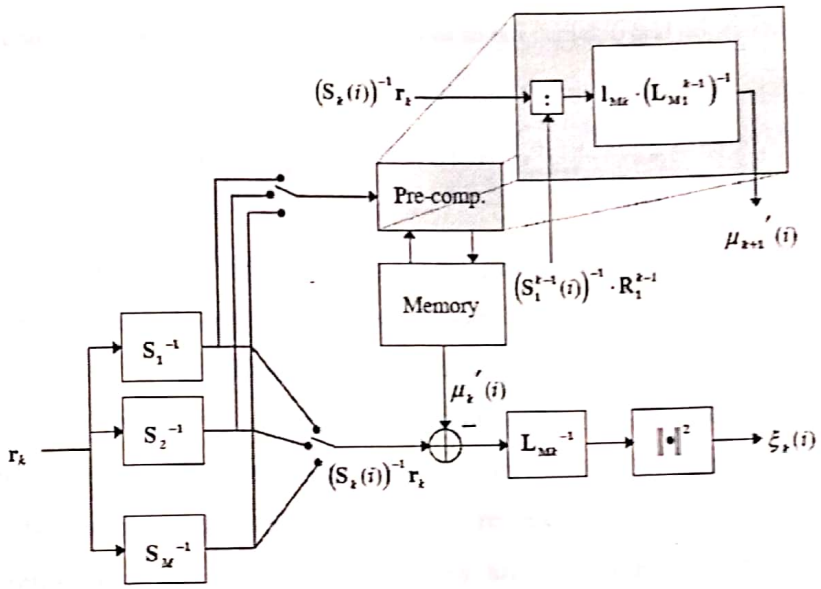


Figure 4.3 Computation of eq. (4.26). The operator ":" performs concatenation of vectors.

The determinants in eq. (4.17) can not in general be disregarded. When, however, the conditional covariance matrix can be factored according to eq. (4.25), the determinant of the conditional covariance matrix can be written

$$\left|\mathbf{C}_1^k(i)\right| = \left|\mathbf{S}_1^k(i)\right| \cdot \left|\mathbf{M}_1^k\right| \cdot \left|\left(\mathbf{S}_1^k(i)\right)^H\right| = \left|\mathbf{M}_1^k\right| \cdot \prod_{j=1}^{kN} \left|S_j(i)\right|^2 . \qquad (4.27)$$

There are probably more modulations giving this independence. The exact conditions are presently not known exactly.

4.4 Optimal Detector

The operation of the MAP sequence detector can now be summarized. At time k, the MAP detector performs an exhaustive search among the M^k candidate sequences to find the symbol sequence maximizing

$$\Omega_k(i) = \Gamma_k(i) + \Lambda_k(i). \tag{4.28}$$

The path metric $\Gamma_k(i)$ is computed recursively according to

$$\Gamma_k(i) = \Gamma_{k-1}(i) + \gamma_k(i) \tag{4.29}$$

and shall therefore be stored until the next symbol interval. The metric increment at time k is

$$\gamma_k(i) = \log\left(P(s_k(i))\right) + \log\left(\frac{\left|\mathbf{C}_1^{k-1}(i)\right|}{\left|\mathbf{C}_1^{k}(i)\right|}\right) + \xi_k(i). \tag{4.30}$$

The second term in eq. (4.28) only contributes to the decision and not to the recursion. It can be written

$$\Lambda_k(i) = \log\left[\sum_{\text{all } s_1^{k+d}(j)} \prod_{l=1}^{d} P(s_{k+l}(j)) \cdot \frac{\left|\mathbf{C}_1^{k+l-1}(j)\right|}{\left|\mathbf{C}_1^{k+l}(j)\right|} \cdot e^{\xi_{k+l}(j)}\right] \tag{4.31}$$

where the sequences S_i^{k+d}(j) all have the messages S_i^k(i) in common . The term ξ_k(i) plays a central role in the equations above. It can be computed according to eq. (4.20) and be realized by linear filtering of the received vector (figure 4.2). Since a unique set of filters is required for each possible message, the number of filters grows exponentially in time. In many cases, however, only one set of filters is required and . . ξ_k(i) can be computed from eq. (4.26) and be realized as in figure 4.3. In those cases the determinants of the covariance matrices are also independent of the symbol sequence and eq. (4.30) and (4.31) simplifies accordingly. Even when the metric computation can be performed according to eq. (4.26) the complexity is still prohibitive. There are three main reasons for this:

i. The number of sequences that must be evaluated at time k is $M^k$and is therefore exponentially increasing in time.

ii. The sizes of the matrices in eq. (4.20) and (4.26) are proportional to the time index k.

iii. The number of sequences required in the evaluation of eq. (4.31) is M^d i.e. exponentially increasing in the delay d.

Thus, except for very short messages, some form of complexity reduction is required in order to obtain realizable detectors. The price for this is a suboptimal detector, having a deteriorated performance.

4.5 Suboptimal detectors

All of the complexity problems mentioned above have to be dealt with. Starting with the second one (ii), note that the correlation of the fading decreases with time. This was obvious, where it was found that extension of the observation interval in the receiver beyond a certain length gave negligible performance gains. Because of this it seems reasonable to truncate the time dependence in the detector. The result of such truncation is that only the b+1 (b is the truncation length) most recent symbol intervals are taken into account in the computations. Truncation implies approximations in the previously derived equations; therefore some performance loss shall be anticipated. This loss, however, should become smaller when b is increased. After the truncation, equations (4.28) and (4.29) remain the same, while (4.30) and (4.31) become

$$\gamma_k(i) = \log\left(P(s_k(i))\right) + \log\left(\frac{\left|\mathbf{C}_{k-b}^{k-1}(i)\right|}{\left|\mathbf{C}_{k-b}^{k}(i)\right|}\right) + \xi_k(i) \qquad (4.32)$$

and

$$\Lambda_k(i) = \log\left[\sum_{\text{all } s_1^{k+d}(j)} \prod_{l=1}^{d} P(s_{k+l}(j)) \cdot \frac{\left|\mathbf{C}_{k+l-b}^{k+l-1}(j)\right|}{\left|\mathbf{C}_{k+l-b}^{k+l}(j)\right|} \cdot e^{\xi_{k+l}(j)}\right] \qquad (4.33)$$

All covariance matrices are replaced by their truncated counterparts and those in eq. (4.32) and (4.33) now have the dimensions Nb by Nb and N(b+1) by N(b+1) respectively.The expressions for $\xi_k(i)$ in eq. (4.20) and (4.26) are also changed to

$$\xi_k(i) = -\left\|\mathbf{L}_{ck}(i)^{-1} \cdot \left(\mathbf{r}_k - \mathbf{l}_{ck}(i) \cdot \left(\mathbf{L}_{ck-b}^{k-1}(i)\right)^{-1} \cdot \mathbf{R}_{k-b}^{k-1}\right)\right\|^2 \qquad (4.34)$$

and

$$\xi_k(i) = -\left\|\mathbf{L}_{Mk}{}^{-1} \cdot \left(\left(\mathbf{S}_k(i)\right)^{-1} \cdot \mathbf{r}_k - \mathbf{l}_{Mk} \cdot \left(\mathbf{L}_{Mk-b}^{k-1}\right)^{-1} \cdot \left(\mathbf{S}_{k-b}^{k-1}(i)\right)^{-1} \cdot \mathbf{R}_{k-b}^{k-1}\right)\right\|^2 \qquad (4.35)$$

As above, these equations lead to filter realizations (figures 4.2 and 4.3). The main difference is now that the matrices $l_{ck}(i)$, $L_{Ck-b}^{k-1}(i)^{-1}$ and l_{Mk}. $((L_{m\,k-b}^{k-1}))^{-1}$ both have dimensions N by Nb instead of N by N(k-1) and thus filter banks with N FIR-filters of length Nb are needed. The number of filter banks required for computation

of eq. (4.34) is also lowered, since there are only M^b i, instead of M^k i possible messages over the truncated interval. If the detector is robust against non-perfect knowledge of the covariance matrices, the same set of filters can be used over a range of real channels. This reduces the number of filter banks needed to be stored in the receiver or the rate at which the filter coefficients must be recomputed as the channel varies.

The truncation of the time dependence in the detector practically becomes equivalent to the assumption of finite correlation in the fading process, which is made in all derivations of realizable sequence detectors .

The first complexity problem, i.e. the exponential increase of the number of sequences to be evaluated has been dealt with in several related ways.

It is shown that complexity constrained breadth first MAP sequence detection is achieved by at every symbol interval choosing the B symbol sequences having the largest a posteriori probabilities. That is exactly what is done in the SA(B)-algorithm (M-algorithm). Even though e.g. RSSD schemes also achieve complexity constrained MAP sequence detection, this is under further structural constraints. The result is that at a given complexity in terms of number of survivors, RSSD (or any other reduced complexity strategy) cannot give a lower error probability than the SA(B)-algorithm. This assumes breadth first search. When the channel memory is truncated as described above, the a posteriori probabilities can only be approximated. If, however, the truncation length b is chosen sufficiently large, accurate approximate expressions will be obtained. For that reason the SA(B)-algorithm is expected to give a good performance also when the channel is truncated.

As pointed out above (iii), there is also an exponential increase in the number of terms required to evaluate $\wedge_k$ (i). Thus some complexity reduction may be needed when d is large. Again this is achieved by application of the SA(B)-algorithm to select the ß sequences to be used in the computation of eq. (4.33).

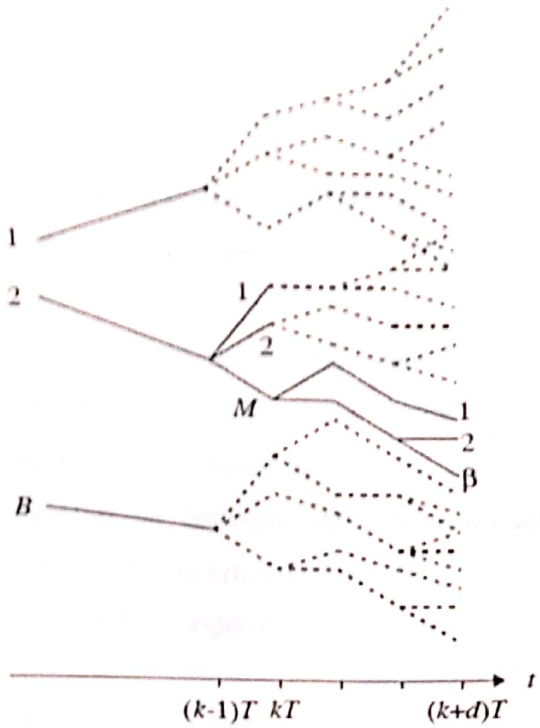


Figure 4.4 Operation of DA(B, β)

This algorithm, which uses the SA(B) twice will be referred to as the DA(B, ß), where DA means Detection Algorithm. Its operation can be illustrated as in figure 4.4. At time k only BM, of a total of M^k candidate sequences (possible messages) are evaluated. Those candidates are obtained by extending the B available sequences, i.e. the survivors, at time k-1 by the M possibly transmitted symbols in interval k. If d=0, B survivors are chosen based on the path metrics. If, however, d>0 a tree with M^d branches grows from each candidate up to time k+d. In each such tree, the same algorithm (i.e. the SA(B)) is employed to select the ß best equences to be used in the computation of . $\wedge_k$ (i). The selection of those sequences is based on the path metrics Γ_{k+l} (j) where l=1, ..., d and the j:th symbol sequence coincides with message i up to time k.

Asymptotically, at high $\frac{}{SNR}$, the expression for $\wedge_k$ (i). (i.e. eq. (4.33)), will be strongly dominated by only one sequence. This gives the simple approximate equation

$$\wedge_k(i) \approx \max_{s_1^{k+d}(j)} \left[\sum_{l=1}^{d} \log\left(P(s_{k+l}(j))\right) + \log \frac{\left|C_{k+l-b}^{k+l-1}(j)\right|}{\left|C_{k+l-b}^{k+l}(j)\right|} + \xi_{k+l}(j) \right]. \qquad (4.36)$$

which is even further simplified when the determinants are independent of j (see eq. (4.27))

$$\Lambda_k'(i) = \max_{s_1^{k+d}(j)} \left[\sum_{l=1}^{d} \log\left(P(s_{k+l}(j))\right) + \xi_{k+l}(j) \right]. \qquad (4.37)$$

The main utility of this expression is that no exponentials need to be computed. This makes practical realization considerably easier.

A problem occurs for tree-based sequence detection algorithms (like DA(B, ß)) on the Rayleigh fading channel. The metric increments during a deep fade are almost identical for all sequences. This is because the message component of the received signal is very small; therefore the corresponding observables contain mostly noise. Since the metric increments over one symbol interval in general is much smaller than that distance, only the best of those sequences will survive .

Here, the solution to this problem is not to allow survivors to trace the same symbol sequence. Instead they are forced to diverge over the most recent symbol intervals.

4.6 Complexity of the DA(B, ß)-algorithm

Here the complexity in terms of the number of multiplications per symbol interval is addressed. The complexity in the initialization (i.e. Cholesky factorizations) of the filter banks used in eq. (4.34) or (4.36) are not accounted for. The reason for this is that those are performed either in advance or only when there is a change of channel statistics and are thus not important for the delay of the algorithm.

The different parameters of the DA(B, ß)-algorithm affect the complexity in several ways. Consider first computation of . ξ_k(i) according to eq. (4.35). Initially, 4MN multiplications are needed to compute $S_k\ (i)^{-l}\ .r_k$ for the M channel symbols. After this, eq. (4.35) can be divided in calculations that must be performed after reception of r_k and those that can be made in advance. The multiplication by L_{Mk}^{-l} -1 and the computation of the squared norm belong to the first category and are essential for the decision delay of the algorithm. Those operations require $2N^2$+3N multiplications. The other operations given by eq. (4.35) can be computed prior to the reception of r_k. There 4bN^2multiplications are needed.

For DA(B, 0), this leads to a total number of $2BN^2$ (M+2b)+ MN(3B+4) multiplications out of which $2BMN^2$+MN(3B+4) must be performed when r_k has been received.

Notably, the truncation length b does only affect the computations that can be made in advance. Another interesting observation is the quadratic dependence on the number of observables, N. For DA(B, ß) computation of eq. (4.36) or (4.37) requires a total number of 2BßdM(M+2b) N^2 + dM (3MBß+4)N multiplications. Even though it is possible to decrease this number by reuse of previously computed quantities, ß>0 gives a significantly larger complexity than ß=0.

When MPSK signaling is used there is little need for storage, since only one set of filters is required (provided eq. (4.25) holds). For modulation formats having more than one energy level in the constellation, on the other hand, M_e^bsets of filters are required. For instance for 16QAM, this gives 3b 3^b sets, which soon becomes too many for reasonable implementation cost.

4.7 The problem of the random phase induced by the fading

The fading process randomly changes the phase of the transmitted signal. The change is time continuous and its rate of variation is determined by the fading bandwidth. It makes it impossible to separate the signs of the fading process from those of the message. This results in that messages having opposite signs, but equal magnitudes in all channel symbol intervals will give the same metric values. One way to circumvent this problem is of course to avoid antipodal signaling and use orthogonal signals instead. That will however lead to a waste of power, at least on the AWGN channel. Instead there are essentially three other solutions available: reference information in the form of pilot tones or pilot symbols can be transmitted or differential encoding can be employed. The word pilot here means that a reference signal, which is known to the receiver is transmitted together with the channel symbols. The reference signals can either be regularly spaced symbols (pilot symbols) or a single tone (pilot tone) . In both cases a larger bandwidth is required to convey a certain amount of information. For differential encoding, on the other hand, no external reference is needed and thus

spectral expansion is avoided. Here the focus is on PSAM but differential encoding is also dealt with.

4.8 Pilot symbols

When pilot symbols are used, known symbols are periodically (period P, i.e. P-1 channel symbols are both preceded and succeeded by a pilot symbol) inserted in the data stream. Often those symbols are chosen from a pseudo random sequence in order to avoid changing the spectral properties of the modulation scheme. The bandwidth required to transmit at a certain information rate is, however, increased by a factor P/(P1).

Most often, the pilot symbols have been used together with an interpolation scheme in order to estimate the fading process between the pilot symbols. By such a strategy the noisy values of the fading process at the pilot symbol intervals are first extracted. For modulation without memory this is straightforward, since the transmitted waveform in a pilot symbol interval only corresponds to the pilot symbol itself. In e.g. CPM-schemes on the other hand, the transmitted signals are dependent upon previously transmitted symbols. In those cases the pilot symbols are used to periodically force the modulator to a predetermined phase state , where an estimate of the fading can be obtained.

Based on several such fading values, an interpolation is performed, done by Wiener filters, but suboptimal strategies like low-pass filtering or simple interpolation schemes are often employed. The detectors then use these estimates to compensate for the fading distortion and they can be viewed as detectors for a time varying AWGN channel. The measured values of the fading process are regarded as samples. Hence the Nyquist frequency puts a limit on the pilot symbol spacing. It is demonstrated that optimal interpolation gives a drastically deteriorated performance when that limit is exceeded, i.e. when $2P>1/f_n$.

Here, the pilot symbols are not separately taken care of. Instead they are a part of the decision function for the MAP sequence detector in eq. Since the pilot symbols are not used for interpolation, the Nyquist theorem does no longer put a limit on the pilot symbol spacing.

4.9 Differential encoding

When differential encoding is used, the phase difference instead of the phase itself is the information bearing part of the signal. For instance for binary differential PSK the data bit 1 may correspond to a phase shift of 180 degrees, while the phase remains unchanged when bit value 0 occurs. By this strategy no explicit phase reference is needed, therefore the required bandwidth is not altered. The main disadvantage is that for MPSK signals the bit error rate is approximately doubled (with coherent detection) when differential encoding is introduced. The reason for this is that an incorrectly decoded channel symbol usually gives two bit errors. This difference in error probability is not very important on the AWGN-channel since it corresponds to a small loss in power.

Differential detection, where the phase difference between consecutive matched filter outputs is extracted, is one way to detect differentially encoded MPSK signals . Better performance, however, can be achieved by so called multiple symbol differential detection or more generally ML sequence detection . The DA(B, ß) can be used here.

4.10 Performance evaluation

To evaluate the error probabilities of the detectors introduced above, both simulations and numerical bounds are desired. Here lower and upper bounds on the bit error probability of MAP sequence detection is presented. Observe that these are bounds for the optimal detector and not for the suboptimal detectors presented above.

4.11 Lower bound

An approach to the analysis of tree based detectors is given , the optimal first symbol detector is analyzed. That detector makes an optimal decision of which symbol S_1 in a block of symbols was sent given that all previously transmitted symbols are correctly detected. It finds the symbol value d j δ_j (j=1, 2,..., M) that maximizes

$$P\left(s_1=\delta_j\middle|\mathbf{R}_1^k\right)=\frac{\sum_{\text{all } \mathbf{s}_1^k(i)} P\left(s_1=\delta_j,\mathbf{s}_1^k(i)\right)p\left(\mathbf{R}_1^k\middle|s_1=\delta_j,\mathbf{s}_1^k(i)\right)}{\sum_{\text{all } \mathbf{s}_1^k(i)} P\left(\mathbf{s}_1^k(i)\right)p\left(\mathbf{R}_1^k\middle|\mathbf{s}_1^k(i)\right)}. \qquad (4.38)$$

Since this detector is optimal in the sense of symbol error probability, it must give a lower or equal symbol error probability than the optimal sequence detector under the same conditions. Its performance can therefore be used as a lower bound to the error probability P(e) of the optimal MAP sequence detector.

The first symbol error probability $P_1(e)$ can be written

$$P(e) \geq P_1(e) = \sum_{\text{all } s_1^k(i)} P\left(s_1^k(i)\right) P\left(e \middle| s_1^k(i)\right) \tag{4.39}$$

where the summation is over all messages and $P(e|s_1^k(i))$ denotes the probability of a first symbol error when message i is transmitted. In general, it is very difficult to evaluate this expression and because of that, lower and upper bounds are considered.

The lower bound is obtained by application of a genie aided approach. For each message in the summation in eq. (4.39), the genie supplies the receiver with some side information. Provided an optimal detector exploiting that information is used, its error probability must be lower or equal to the error probability of the real detector for that specific transmitted sequence. The side information used here, is that there is only one (instead of M^k- 1) specified competing symbol sequence S_j^k(j) for each message. The conditional error probability can now be lower bounded by

$$P\left(e \middle| s_1^k(i)\right) \geq P_{\text{genie}}\left(s_1^k(i) \rightarrow s_1^k(j)\right) \tag{4.40}$$

where Pgenie() denotes the first symbol error probability of the optimal genie aided detector. The competing sequence can be chosen to be any symbol sequence, but the aim is to use a sequence that gives as high an error probability for the genie aided detector as possible. Let

$$P_{\text{genie}}^{\max}\left(s_1^k(i)\right) = \max_j \left\{ P_{\text{genie}}\left(s_1^k(i) \rightarrow s_1^k(j)\right) \right\} \tag{4.41}$$

be the maximal error probability given s1 S_l^k(i) is transmitted. Note that this probability in general is dependent on the transmitted sequence. The symbol error probability of the sequence detector can now be lower bounded by

$$P(e) \geq \sum_{\text{all } s_1^k(i)} P\left(s_1^k(i)\right) P_{\text{genie}}^{\max}\left(s_1^k(i)\right). \qquad (4.42)$$

Even though the use of $P_{\text{genie}}^{\max}\left(s_1^k(i)\right)$ gives the tightest possible bound it is not very use-ful in practice, since it involves finding the worst case competing sequence for every $s_1^k(i)$. Note, however, that any possible competing sequence will yield a lower bound, therefore S_l^k (j) can be chosen with the aim of simplifying eq. (4.42). A reasonable way to choose S_l^k (j) is to pick a sequence that differs from the transmitted one only in the first symbol interval. The differing symbol should also be chosen such that as large an error probability is obtained. By $P_{\text{genie}}^{1}\left(s_1^k(i) \to s_1^k(j)\right)$ denote this error probability, the lower bound becomes

$$P(e) \geq \sum_{\text{all } s_1^k(i)} P\left(s_1^k(i)\right) P_{\text{genie}}^{1}\left(s_1^k(i) \to s_1^k(j)\right). \qquad (4.43)$$

Still all transmitted sequences must be evaluated. It is therefore tempting to reduce the number of terms in the summation so that only one transmitted sequence S_l^k (j) for each of the M possible first symbols is used. This would correspond to using a fixed pairwise error probability for all transmitted sequences that share the same first symbol. Difficulties arise here in general depends on the transmitted sequence. This is due to the correlation in the fading process. In order to still have a lower bound, the smallest error probabilities for each of the M symbol alternatives must be used. Denoting these probabilities by $P_{\text{genie}}^{1,\min}(\delta_n)\ (n=1,\ldots,M)$, eq.(4.43) can be written

$$P(e) \geq \sum_{\text{all } \delta_n} P(\delta_n) P_{\text{genie}}^{1,\min}(\delta_n). \qquad (4.44)$$

For some signal constellations, e.g. MPSK, $P_{\text{genie}}^{1,\min}(\delta_n)$ is equal for all M symbols and eq. (4.44) can be simplified accordingly.

Sometimes, a lower bound on the bit-, instead of the symbol error probability is desired. That is easily obtained from eq. (4.44) by noting that a symbol error must at least correspond to one bit error and consequently

$$P(e) \geq \frac{1}{\log_2 M} \sum_{\text{all } \delta_n} P(\delta_n) P_{\text{genie}}^{1,\min}(\delta_n). \qquad (4.45)$$

Even though this equation looks very simple, the problems of how to evaluate the pair-wise error probabilities and finding $P_{\text{genie}}^{1,\min}(\delta_n)$ have not yet been treated. Starting with the first one, those probabilities can be computed by the same technique. There, error probabilities are computed for sequences of orthogonal symbols, where only the midpoint symbols in the sequences are distinct. A symmetric observation interval around those symbols is used. The sequences in eq. (4.45) have already been specified to differ only in a single (the first symbol) interval, therefore by changing the modulator waveforms from orthogonal to antipodal the technique applies. Due to the infinitely long correlation in the fading process, the observation interval should ideally be chosen from minus infinity to infinity. Fortunately, very good approximations are obtained by only using a few symbol intervals. That also simplifies the second problem, i.e. to find $P_{\text{genie}}^{1,\min}(\delta_n)$, since only a small number of transmitted sequences need to be investigated.

4.12 Upper bound

The symbol error probability will in general vary in time when pilot symbols are used. Because of this, the average error probability over all symbol intervals is desired. At time k that probability becomes

$$\begin{aligned}\overline{P(e)} &= \frac{1}{k}\sum_{l=1}^{k} P(e_l) = \frac{1}{k}\sum_{l=1}^{k}\sum_{i=1}^{M^k} P\left(\mathbf{s}_1^k(i)\right) P\left(e_l \middle| \mathbf{s}_1^k(i)\right) = \\ &= \frac{1}{k}\sum_{l=1}^{k}\sum_{i=1}^{M^k} P\left(\mathbf{s}_1^k(i)\right) P\left(\bigcup_{j=1}^{M^k} \mathbf{s}_1^k(j), e_l \middle| \mathbf{s}_1^k(i)\right)\end{aligned} \tag{4.46}$$

where e_j denotes the event of a symbol error at time l. Application of union bound gives

$$\begin{aligned}\overline{P(e)} &\le \frac{1}{k}\sum_{l=1}^{k}\sum_{i=1}^{M^k} P\left(\mathbf{s}_1^k(i)\right) \sum_{j=1}^{M^k} P\left(\mathbf{s}_1^k(i) \to \mathbf{s}_1^k(j), e_l\right) = \\ &= \frac{1}{(P-1)\cdot M^{P-1}} \sum_{l=l_0+1}^{l_0+P-1} \sum_{i=1}^{M^{P-1}} \sum_{j=1}^{M^{P-1}} P\left(\mathbf{s}_{l_0+1}^{l_0+P-1}(i) \to \mathbf{s}_{l_0+1}^{l_0+P-1}(j), e_l\right).\end{aligned} \tag{4.47}$$

$P(s_1^k(i) \to s_1^k(j), e_l)$ denotes the pairwise sequence error probability, i.e. the probability that S_l^k (j) is chosen when S_l^k (j) is transmitted. Inclusion e_j gives the further requirement that the two sequences differ at time l. The second step in the equation is obtained under the assumption of a very long transmitted sequence. The average symbol error probability can then be computed by summation over a single pilot symbol period (of length P and starting at time l_0) and the pairwise sequence error probability changes accordingly. This is similar to the bound] obtained for block-transmission.

If the bit- instead of the symbol error probability is desired eq. (4.47) can be written

$$\overline{P(e)} \le \frac{1}{(P-1)\cdot M^{P-1}\cdot \log_2 M} \sum_{l=l_0+1}^{l_0+P-1} \sum_{i=1}^{M^{P-1}} \sum_{j=1}^{M^{P-1}} w_{ij,l} \cdot P\left(s_{l_0+1}^{l_0+P-1}(i) \to s_{l_0+1}^{l_0+P-1}(j), e_l\right) \tag{4.48}$$

where $w_{ij,l}$ denotes the number of bit errors that occurs at time l when the j:th sequence is incorrectly chosen instead of message i. The pairwise sequence error probability still remains to be computed. Again, the correlation in the fading process causes problems. The reason for this is that the pairwise error probability is dependent on the duration of the observation interval. The optimal sequence detector, derived bases its decision on the whole received sequence (from time 1 to k), therefore a shorter observation interval must give an upper bound to the pairwise error probability.

Example 4.1

Consider pilot symbol assisted BPSK (rectangular modulator waveforms) with P=3. This implies that two data symbols, preceded and succeeded by pilot symbols, are transmitted. Let the observation interval start at time 0 and include both of the framing pilot symbols. For symmetry reasons, the error probabilities now are the same for each of the data symbols. The averaging with respect to the error position can then be removed. For BPSK, the pairwise error probability is found to be practically

independent

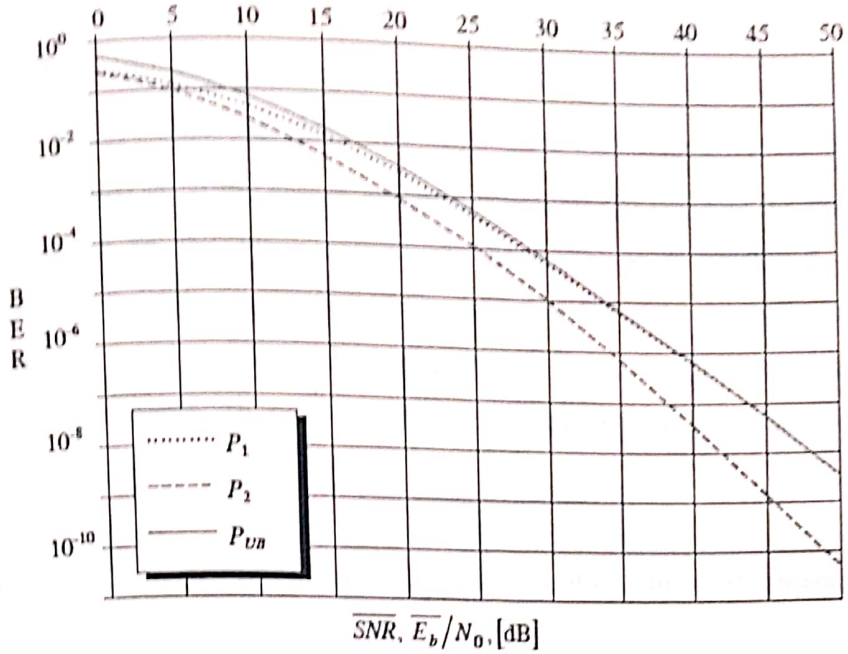


Figure 4.5 Illustration of the error probabilities in example 4.1.

(for this type of calculation) of the transmitted sequences, therefore the summation in eq. (4.48) can be made over any single message. The bit error probability can now be upper bounded by

$$\overline{P(e)} \leq \sum_{j=2}^{3} P\left(\mathbf{s}_1^2(1) \rightarrow \mathbf{s}_1^2(j)\right) = P_1 + P_2 \qquad (4.49)$$

Where message 1 is assumed to have been transmitted. Figure 4.5 shows the two pair-wise error probabilities P1 and P2 together with the union bound for N=4.

This example indicates that a good rule of thumb for obtaining an approximation to the union bound at high SNR on the fast Rayleigh fading channel may be to use only symbol sequences differing over a single symbol interval.

4.13 Simulation setup

Bit error rates (BER) as functions of average signal to noise ratios for uncoded M-ary signaling on the Rayleigh fading channel are estimated by computer simulations. Various versions of the DA(B, ß) are employed. The channel is characterized by the covariance of the fading process $C_f(u,v) = J_0(2\pi \cdot f_n(u-v))$. Now the average SNR is defined as

$$\overline{SNR} = \frac{\overline{E_b}}{N_0} = \frac{C_f(0,0)\cdot(P-1)\cdot\overline{E_s}}{\log_2(M)\cdot P\cdot N_0} \qquad (4.50)$$

where $\overline{E_b}$ is the average received energy per bit and $\overline{E_s}$ is the average transmitted symbol energy. Note that the $\overline{SNR}$ is normalized with respect to the pilot symbol spacing so that $\overline{E_b}$ is the energy per information bit. To simulate the fading process characterized by the above covariance, the technique is employed. Two white Gaussian noise processes are filtered by identical FIR filters to obtain the real and imaginary parts of samples of the fading process. To get an approximation of the time continuous channel, 16 such equally spaced samples are produced during every symbol interval.

The discretization is performed. To approximate this, numerical integration is used. The symbol interval is divided into 16 equally large sub-intervals and the integral is approximated by the midpoint values (of the sub-intervals) of the received signal. One complex sample is used to represent the fading in each sub-interval.

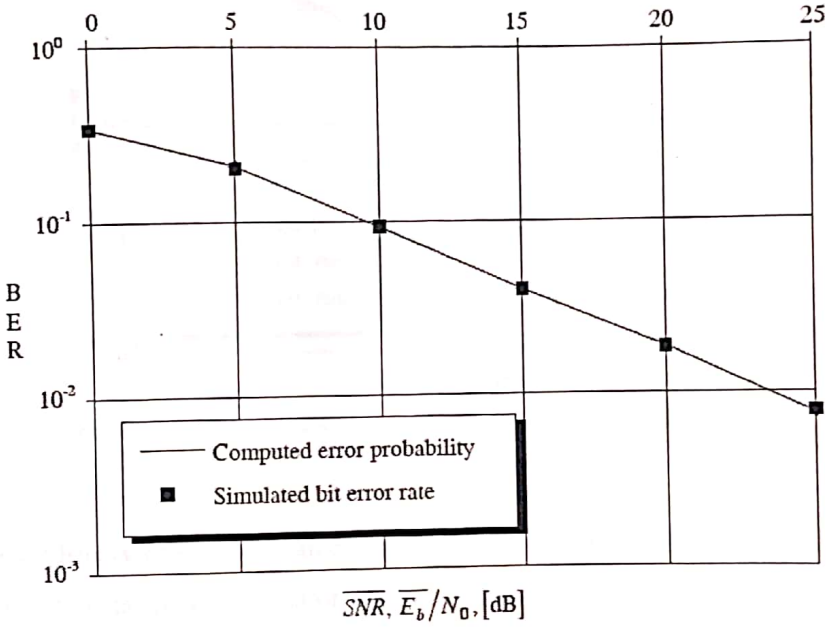


Figure 4.6 Simulation results compared with computed error probabilities for orthogonal modulator waveforms and discretization by set 1.

To study the influences of the above approximation, simulation results for the detector based on the decision rule in eq. (3.13) are compared to the computed error probabilities from chapter 3.4.1. The orthogonal modulator waveforms in figure 3.2

are used and the observation interval coincides with the symbol interval. The number of observables is N=4 and the normalized fading bandwidth is fn=0.1. The simulation results are very close to the computed error probabilities (figure 4.6) indicating that the simulation setup is an accurate approximation of the time continuous channel and the discretization .

4.14 Simulation results

The aim is both to investigate various parameters of the DA(B, ß) algorithm and to illustrate the performance that can be achieved by using efficient detection algorithms on the fast fading channel. The focus is on BPSK signaling, but results for QPSK, 8PSK and 16QAM are also given. Rectangular modulator waveforms are used and the normalized fading rate is fn=0.1, which can be characterized as being fast).

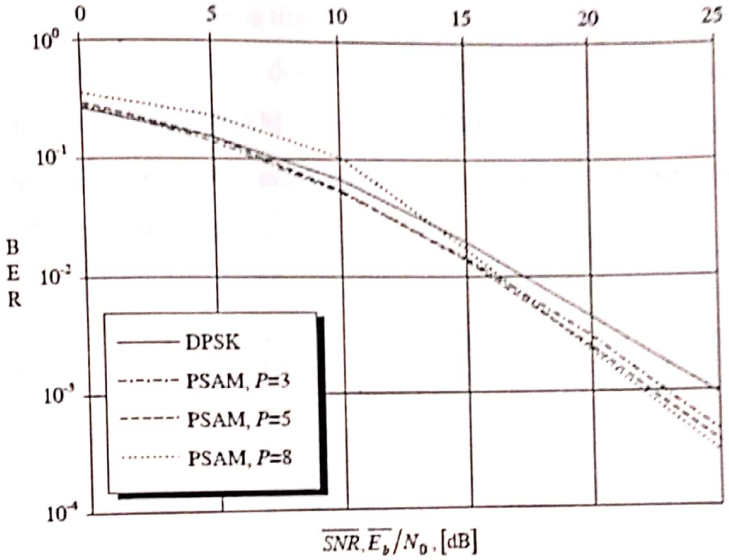


Figure 4.7 Binary PSAM compared with DPSK. N=4, B=12, β=0, b=2.

Figure 4.7 Binary PSAM compared with DPSK. N=4, B=12, ß=0, b=2.

The use of rectangular modulator waveforms and discretization by set 1 makes it possible to factorize the conditional covariance matrices according to eq. (4.25). Hence the simplified computation of ξ_k.(i) (eq. (4.35)) can be employed and only one set of filters is needed for the MPSK schemes, whereas for 16QAM 3^b sets are required. The determinants in eq. (4.32) can be omitted without loss of generality

when MPSK is used. Note that this is not the case for 16QAM, still the determinants will be omitted here as well.

One system parameter at a time is investigated, while the others are held constant. The reason for this is to limit the number of investigations and diagrams. As a complement, results for systems with completely different combinations of parameters (and consequently different complexities) are also presented. Initially only DA(B, 0) is studied, but in the end of the chapter DA(B, ß) is briefly investigated.

Pilot symbol spacing, differential signaling

The error probabilities for binary pilot symbol assisted modulation and differential encoding is displayed in figure 4.7 (N=4, B=12, ß=0 and b=2). DPSK (solid curve) here gives significantly worse performance than PSAM (Pilot Symbol Assisted Modulation). The reason for this is that most symbol errors result in two bit errors for DPSK, giving an approximately doubled bit error rate. Note, however, that the bandwidth requirements for PSAM is a factor P/(P-1) larger than for differential encoding. Despite this, the lower error probabilities obtained for PSAM motivate its use in the rest of the Dissertation.

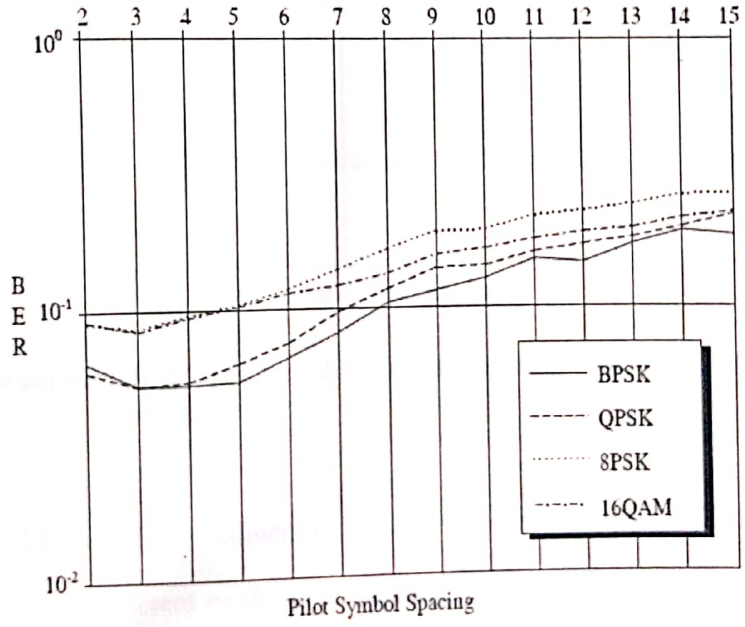


Figure 4.8 Influence of pilot symbol spacing, P at $\overline{SNR}$=10 dB. N=4, B=12, β=0, b=2.

An interesting observation is that the performance at low $\overline{SNR}$ is worse for P=8, than for P=3 and 5, whereas it gives a lower BER at high $\overline{SNR}$. The reason is probably that the number of good candidate sequences becomes larger when the pilot symbol spacing is increased (at least the number possible sequences between two pilot symbols becomes larger). These candidates may be hard to distinguish at low $\overline{SNR}$, which gives a high error probability. Asymptotically at high $\overline{SNR}$, on the other hand, approximately the same bit error rate (as a function of transmitted energy per symbol interval over N_0) is achieved regardless the value of P. In the diagram, the BER is plotted as a function of energy per information bit (over N_0)) and since less energy is spent on transmission of pilot symbols when P is large, those systems perform slightly better.

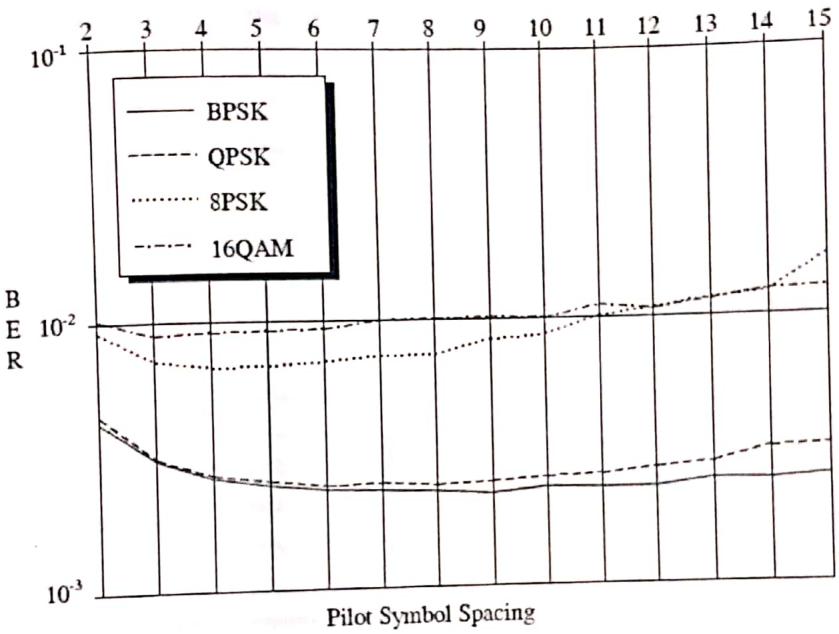


Figure 4.9 Influence of the pilot symbol spacing at $\overline{SNR}$=20 dB. N=4, B=12, β=0, b=2.

In figures 4.8 and 4.9, the influence of the pilot symbol spacing on the bit error rates is displayed for different modulation formats. At low $\overline{SNR}$ (10 dB), increasing the spacing gives a deteriorated performance, while it is essentially unaffected at 20 dB and P<15. For larger values of P there is a gradual deterioration of performance. This

is in contrast to PSAM when interpolation is employed. Those systems exhibit a severe performance loss when the pilot symbol spacing exceeds 1/(2 fn) , which here corresponds to P=5. The different signaling schemes seem to be affected in very much the same way by the distance between the pilot symbols and an overall good choice of pilot symbol spacing at fn=0.1 seems to be P=5. Note that this will result in 0.8 transmitted bits per channel symbol.

Dependence on the truncation length, b

The bit error rates for BPSK and PSAM at various truncation lengths are displayed together with the lower bound in figure 4.10 (N=4, P=5, B=12, ß=0). Very good performances are achieved for the truncation lengths b=4 and b=6, indicating that the error introduced by the truncation soon becomes negligible when b is increased. Note that the case b=6, even though the detector is suboptimal, practically gives the same BER as the lower bound for the optimal detector.

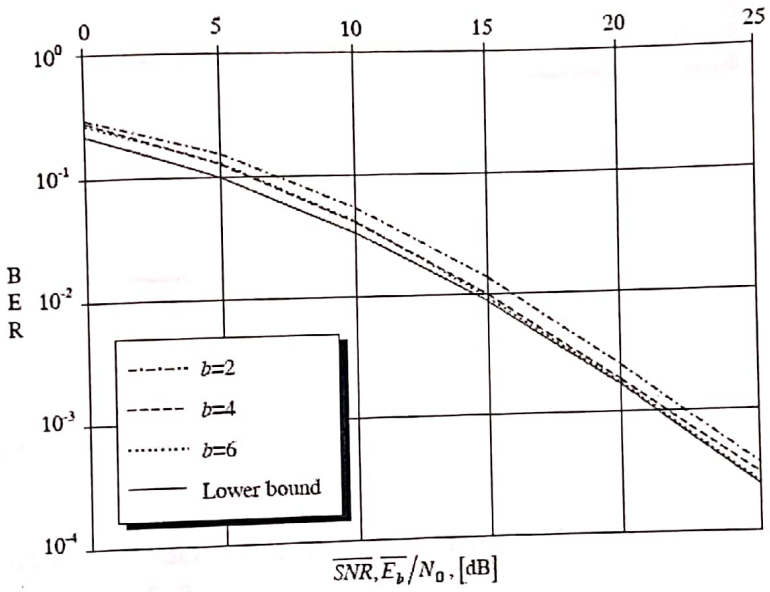


Figure 4.10 Dependence on the truncation length for BPSK (PSAM). N=4, P=5, B=12, β=0.

This property, i.e. ability to achieve close to optimal performance by increasing the complexity is of course desirable for a suboptimal algorithm. These results also support the assumption that the SA(B) (which forms the basis for DA(B, ß)) here gives a good performance compared with other reduced complexity breadth first sequence detection algorithms. The reason for this is that the correctness of this

assumption relies on the accuracy in the approximation obtained through the truncation of the time dependence.

Figure 4.11 illustrates the dependence on b for various modulation formats at $\overline{SNR}$ =25 dB (N=4, P=5, B=12, ß=0). The other schemes (QPSK, 8PSK and 16QAM) are slightly less affected by increased truncation lengths than BPSK. For 16QAM only b=2 and b=4 are simulated, since the number of different filter banks is too large (3^6) when b=6. For MPSK on the other hand only one filter bank is needed and the truncation length merely affects the complexity of the calculations that can be made in advance). Thus, increasing b does not give a larger delay in the receiver, which certainly is a desirable property.

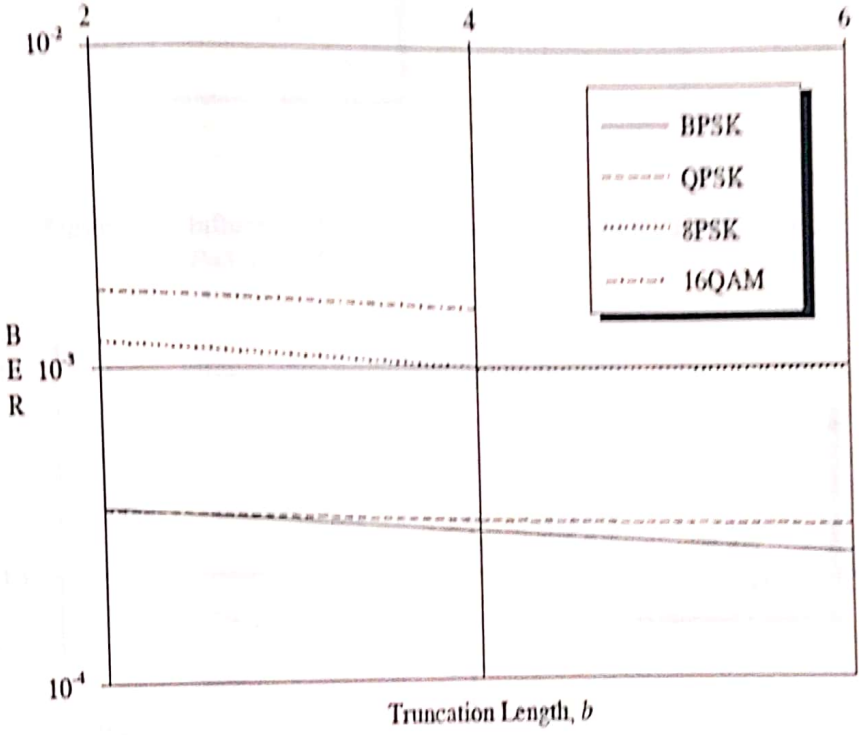


Figure 4.11 Dependence on the truncation length at $\overline{SNR}$=25 dB. *N*=4, *P*=5, *B*=12, β=0.

Dependence on the number of observables per symbol interval, N

Figure 4.12 shows the BER for BPSK at various numbers of observables (b=2, P=5, B=12 and ß=0). N=2 and N=4 give good performances, while discretization by a matched filter (N=1) leads to significantly higher bit error rates. Note that N=1

(instead of N=2) now corresponds to discretization by a matched filter, since rectangular modulator waveforms of duration T are employed.

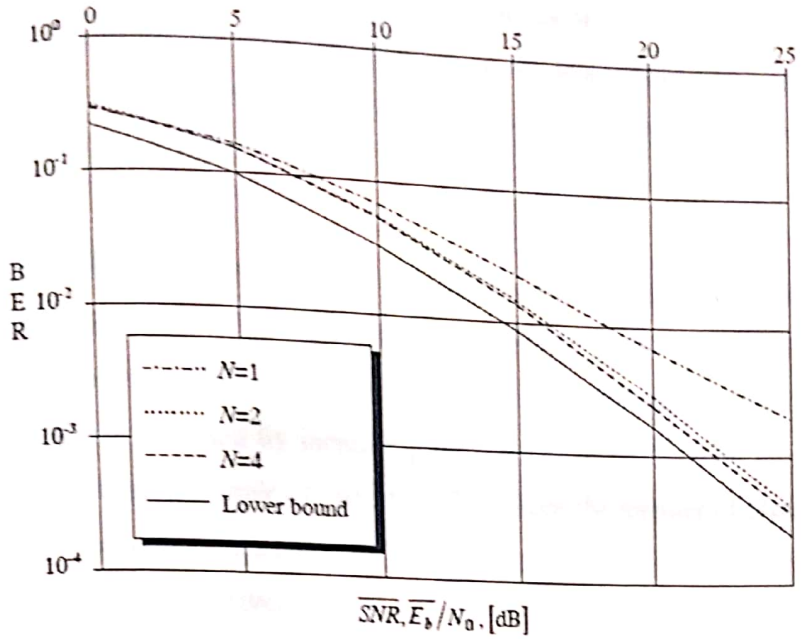


Figure 4.12 Influence of the number of observables for BPSK. B=12, P=5, b=2, β=0.

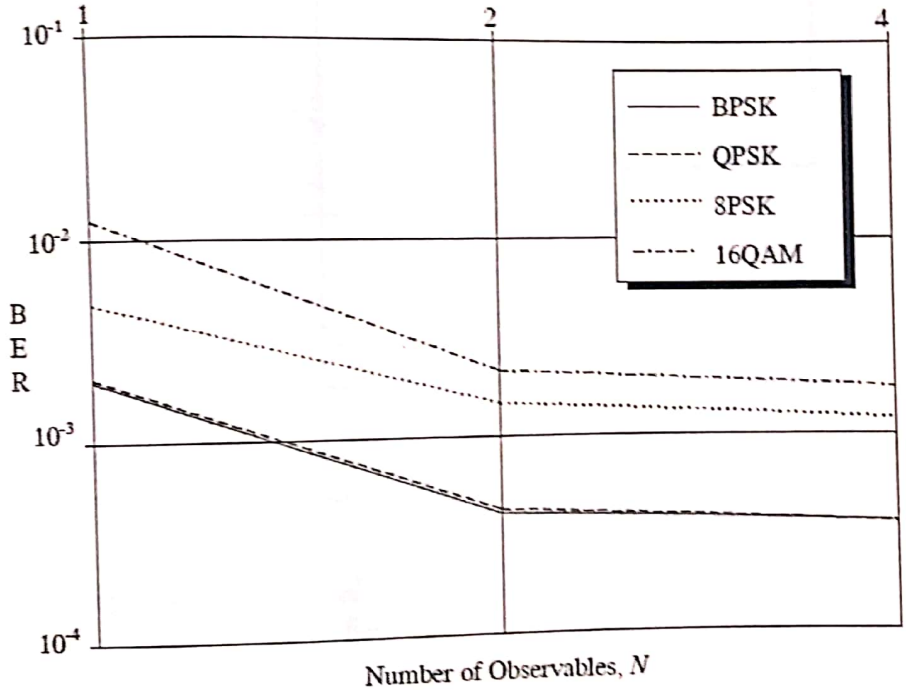


Figure 4.13 Dependence on the number of observables at $\overline{SNR}$=25 dB. P=5. B=12. b=2. β=0.

The result in figure 4.12 is in accordance with the observations in chapter 3.4, i.e. increasing the number of observables leads to a drastic improvement of the performance. Notice also the successively steeper slopes of the curves, which indicate an implicit diversity here as well. The distance to the lower bound is larger than in figure 4.10. The reason for this is that the truncation length used here (b=2) is not sufficiently long.

In figure 4.13 bit error rates at $\overline{SNR}$ =25 dB are shown for various modulation formats. Discretization by a matched filter (N=1) gives significantly worse performance than when more observables are used. Note, however, that only a small performance gain is obtained by increasing the number of observables beyond N=2. That is attractive from a complexity point of view, since the number of observables is the single most important parameter for both the decision delay and the total number of multiplications in the receiver.

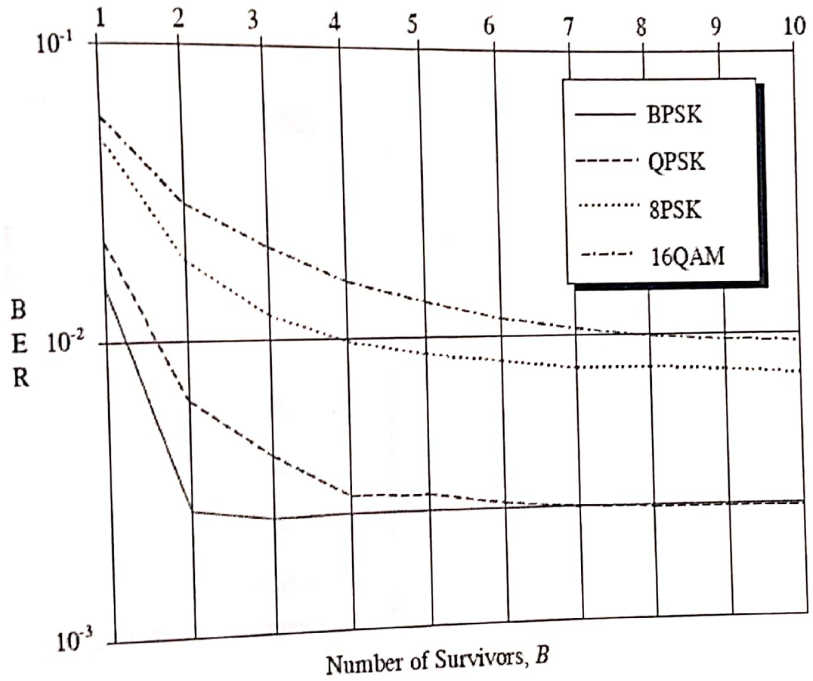


Figure 4.14 Influence of the number of survivors at $\overline{SNR}$=20 dB. N=4, P=5, b=2, β=0.

Dependence on the number of survivors, B

Figure 4.14 displays the bit error rates as functions of the number of survivors at $\overline{SNR}$ =20 dB (N=4, P=5, b=2 and ß=0). For all modulation schemes, only a few

survivors are needed in order to obtain a good performance. Good choices are for: BPSK B=3, QPSK B=4, 8PSK B=7 and 16QAM B=7 or 8. The property that only a small number of survivors is needed is similar to what is found for the SA(B)-algorithm. For TCM on the AWGN channel it is demonstrated that in many cases only a fraction of the number of survivors maintained by the Viterbi algorithm is required to obtain MLSD-performance.

These results also indicate that it may be a waste of complexity to use the Viterbi algorithm here. If e.g. a channel memory (due to the correlation) of three symbol intervals is used, 162=256 states (and survivors) are needed for detection of 16QAM. This obviously leads to a detector, which is drastically more complex than DA(8, 0), even though their performances should be almost equally good.

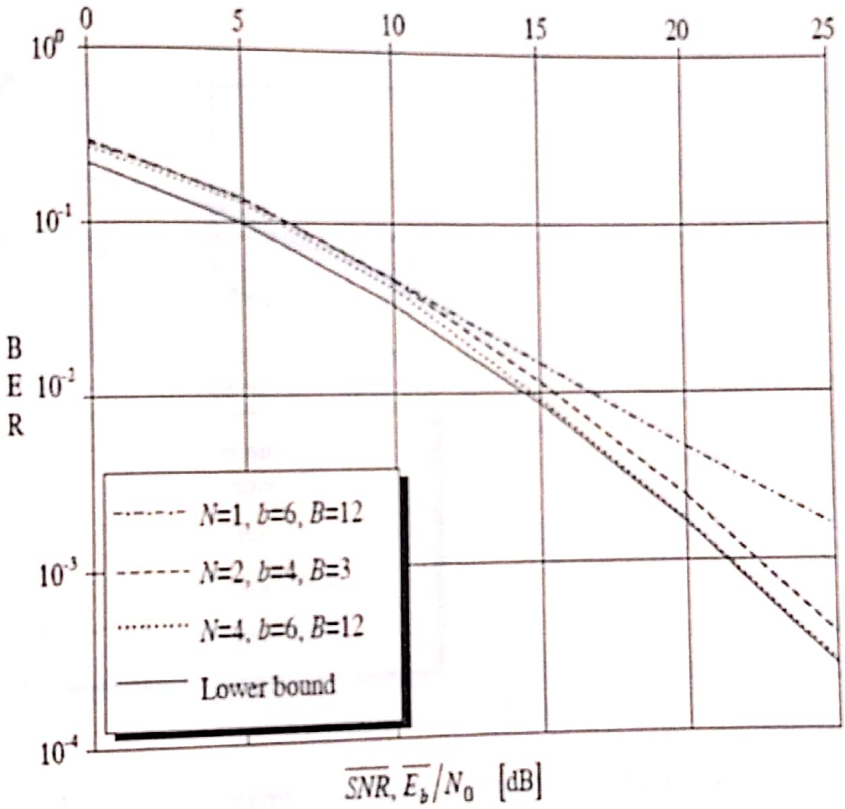


Figure 4.15 Comparison of different detector configurations for BPSK. β=0, P=5.

System examples using DA(B, 0)

So far, the influences of each parameter have been studied separately. The aim now is to find low complexity systems with good performances. The above results are used

as guidelines in the choices of parameters. Through this sub-chapter, the pilot symbol spacing is held constant at P=5. Starting with binary signaling DA(3, 0), N=2 and b=4 gives a performance within 1 dB of the lower bound and it seems to give a good trade off between performance and complexity. The BER of this system is shown in diagram 4.15. In the same diagram, the BER of a matched filter detector (N=1) with DA(12, 0) and b=6 is displayed. Even though a much larger number of survivors (B=12) is used, the performance of the matched filter detector is poor. This performance is far from what is obtained if sufficiently many observables and a long truncation length is used. Then a performance very close to the lower bound is achieved, which is illustrated by using N=4, b=6 and B=12. This suboptimal detector thus seem to give almost the same performance as the MAP sequence detector, even though the complexity of the latter is substantially larger.

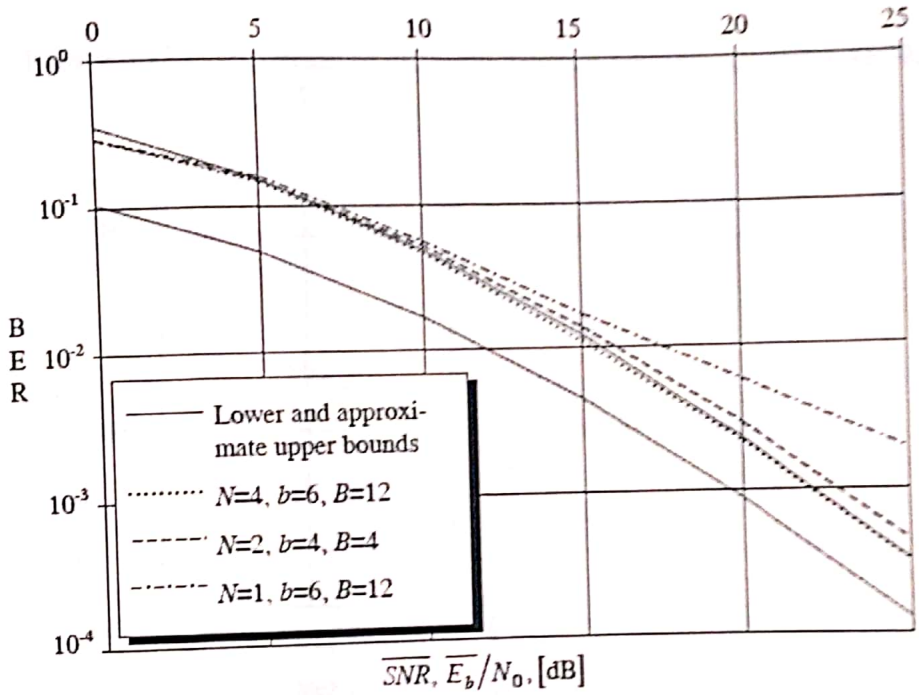


Figure 4.16 Comparison of different detector configurations for QPSK. β=0, P=5.

For QPSK, 8PSK and 16QAM (figures 4.16, 4.17, 4.18), the behavior is very much the same. Detectors based on matched filters give poor performances, but with only a little increase of complexity much lower bit error rates are obtained. Both lower and approximate upper bounds are displayed for QPSK and 8PSK .

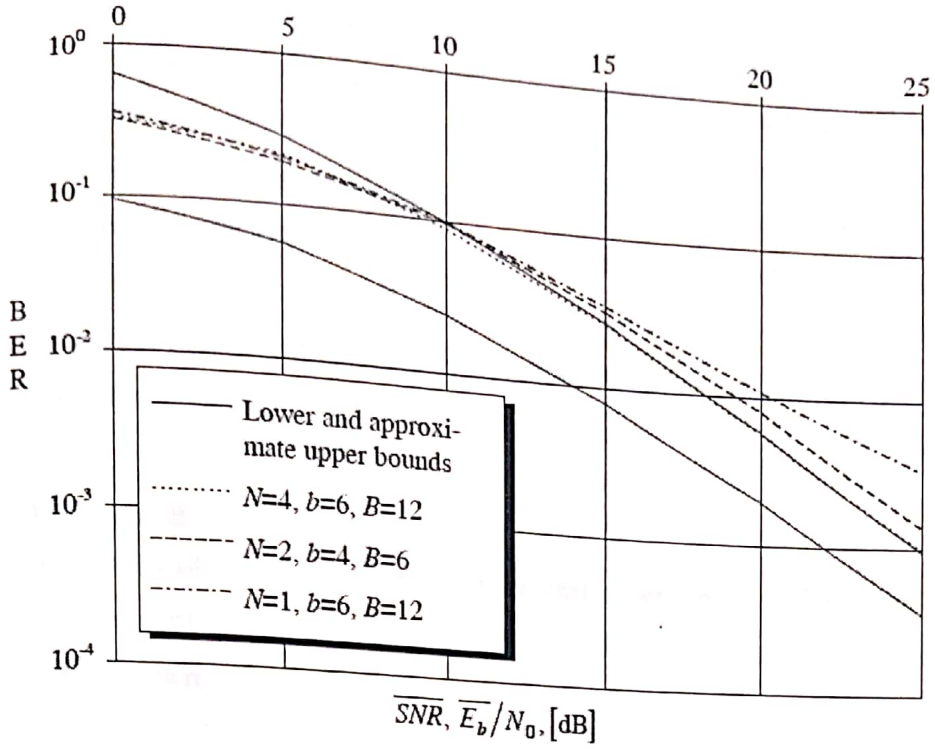


Figure 4.17 Comparison of different detector configurations for 8PSK. β=0, P=5.

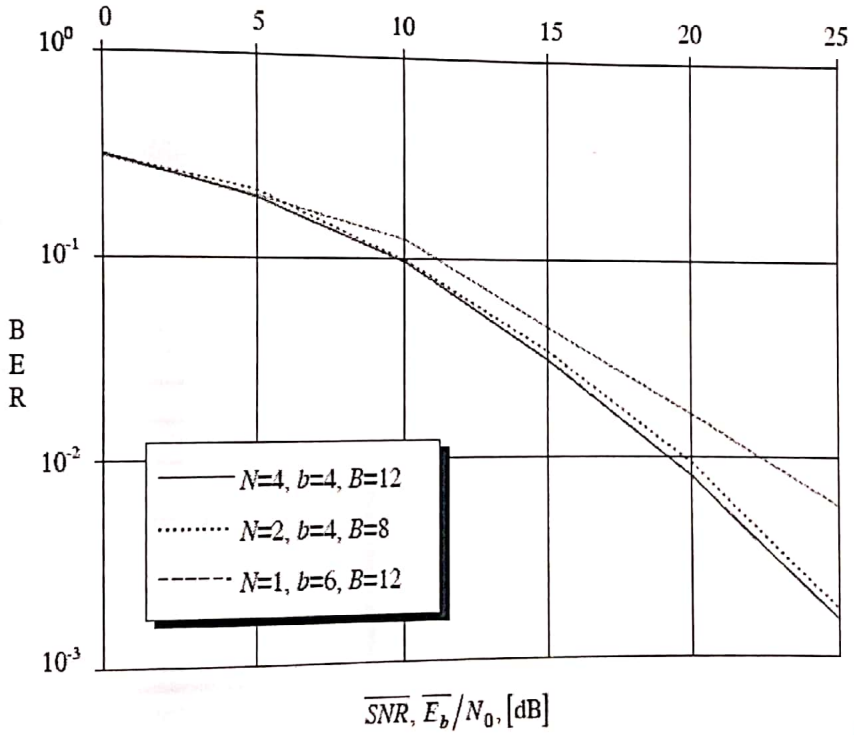


Figure 4.18 Comparison of different detector configurations for 16QAM. β=0, P=5.

In both cases the bit error rates are far from the lower bounds, whereas the approximate upper bounds seem to be very tight. Note that this is an approximate upper bound on the error probability for MAP sequence detection and not necessarily for the suboptimal algorithm considered here. For that reason some of the detector configurations give higher bit error rates than the bound. When, however, the parameters (N, b, B) are sufficiently large the simulation results are upper bounded by the approximate bound. There it seems reasonable to assume that a performance close to that of the MAP sequence detector is obtained.

Dependence on ß

The simulations of DA(B, 0) above indicate that a very good performance can be achieved with just a causal detector. For that reason it is doubtful whether an anti-causal detector is really useful or not.

In figure 4.19 simulation results for DA(2, ß) for QPSK are given (P=5, N=4 and b=2). Equation (4.36) is used for the computation of Λ_k (i) (the determinants are

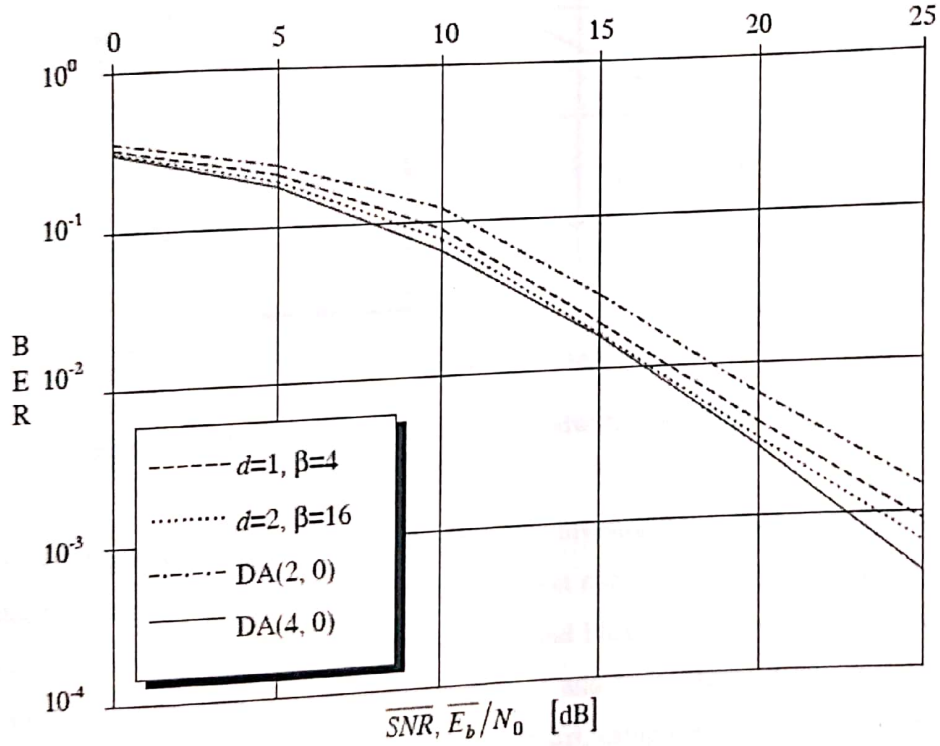


Figure 4.19 Comparison of causal (DA(B, 0)) and anti-causal (DA(B,β)) for QPSK. *P*=5, *N*=4 *b*=2.

For comparison, bit error rates for DA(2, 0) and DA(4, 0) are also presented. Clearly, the performance of DA(2, 0) is improved by increasing ß(and d). Note, however, that even better results are obtained for DA(4, 0) at a significantly lower complexity (approximately 40 times less multiplications), therefore an anti-causal detector seems to be a bad choice in this case.

Dependence on the fading bandwidth

A quite remarkable property for single symbol signaling was observed. Provided the number of observables was large enough and a sufficiently long observation interval was used, the error probabilities became lower as the fading bandwidth was increased.

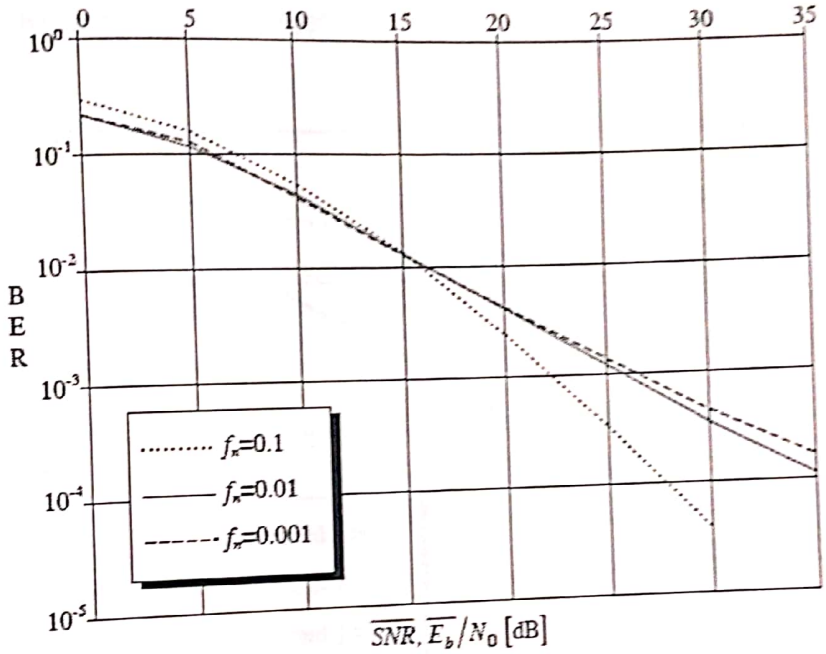


Figure 4.20 Influence of the fading bandwidth for BPSK. DA(12, 0) with N=4 and b=2. P=5.

Here the dependence on the fading bandwidth is investigated for sequence detection, instead of single symbol detection . Three different normalized fading bandwidths are used, fn=0.1, 0.01 and 0.001. Binary signaling and PSAM with P=5 are employed in all cases. The detector is DA(12, 0) with N=4 and b=2. Figure 4.20 displays the simulation results. Below an average SNR of 15 dB, using the high fading bandwidth is inferior to signaling at the other fading rates. The reason for this may be the large pilot symbol spacing relative the fading bandwidth when fn=0.1. When, however, the

SNR is increased signaling at the high fading bandwidth gives a large performance improvement. A similar effect becomes visible when comparing the curves for fn=0.01 and fn=0.001 at SNR >25 dB.

Robustness

The issue of robustness of the single symbol detector was briefly dealt with. The detector was found to be very robust against mismatches in both the shape of the fading spectrum and in knowledge of the noise variance.

Results are obtained here by simulations of sequence detection of BPSK with PSAM (P=5), where the receiver is based on DA(12, 0). The number of observables is N=4 and the truncation length is b=2. Mismatch occurs since the receiver is designed for use on a channel with rectangular fading spectrum and fn=0.1. Two channels with fading autocorrelation function and fn=0.05 and 0.15 are used. In figure 4.21 very small losses are observed for the mismatched detector compared with a detector matched to the channel when fn=0.05.

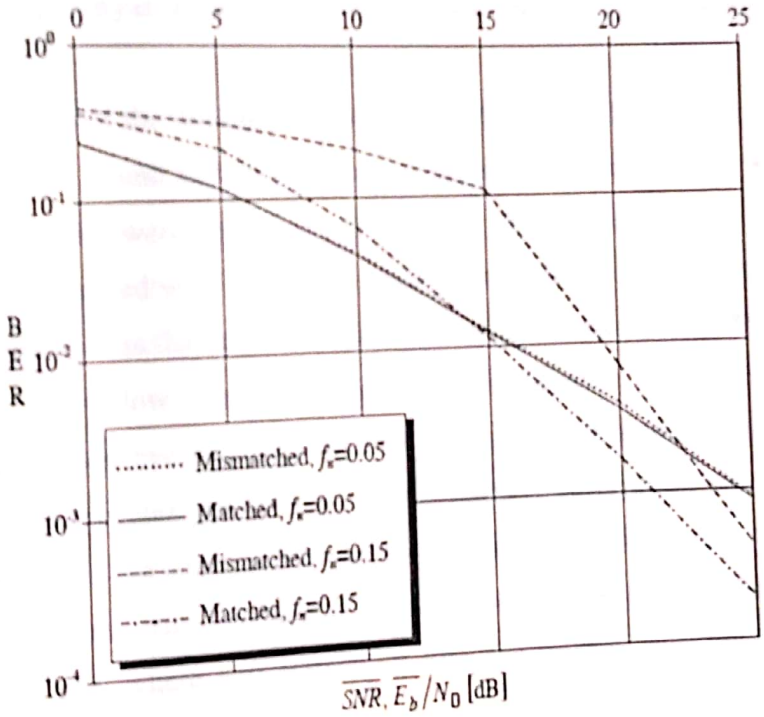


Figure 4.21 Influence of mismatch in the detector.

4.15 Summary

In this chapter sequence detection on the time continuous frequency flat Rayleigh fading channel has been treated. Contrary to anti-causal MAP-, instead of causal

ML-sequence detection was considered. By this pilot symbols are accounted for in a natural way and the correlation with future values of the fading process can be exploited. The derivation of the optimal detector followed the traditional (from AWGN) approach of manipulation of the likelihood function and the focus was on receivers, where discretization is performed. The resulting detector operates on vectors of observables instead of on single samples. This gives a low delay for that specific type of discretization. It was also found that realization of the metric computation by a single set of filters can be made for more general modulations than just constant envelope schemes.

Here as well as in most of the above papers, the optimal detector was found to be too complex, therefore a suboptimal detector (DA(B, ß)) was proposed. In contrast to the suboptimal detectors it is based on the SA(B)- (or M-) algorithm, which is known to be the best complexity constrained breadth first ML-sequence detection algorithm .

The performances of the optimal detector as well as for DA(B, ß) were investigated. Both a new lower bound and an approximate upper bound on the error probability of the optimal detector were presented and used for different modulation formats. The bounds were compared with simulation results for DA(B, ß) at fast Rayleigh fading. Among the results was that a performance close to that of the optimal detector can be achieved at fairly low complexity, provided sufficiently many (N=4) discrete observables are used every symbol interval. If on the other hand matched filtering is used (N=1), the receiver performs far worse. Another observation was that a causal detector could give the same performance as an anti-causal detector at a much lower complexity. The anti-causal version of DA(B, ß) will, however, form the basis for the receiver front end, where soft instead of hard (i.e. symbol decisions) outputs are desired.

As for the single symbol detector, the sequence detector was found to be rather robust against imperfect knowledge of the statistical properties of the channel. This is a result of the implicit diversity in the fading process.

CHAPTER 5

Coded Modulation

CHAPTER 5

Coded Modulation

Coding combined with interleaving was presented as an efficient method for achieving time diversity for digital signaling on the Rayleigh fading channel. For transmission of more than one information bit per symbol, there are essentially two coding strategies available:

1. Trellis coded modulation or
2. Multilevel coding .

The focus is on trellis coded modulation (TCM) and variations thereof. Contrary to, where a time continuous channel model were employed, the discrete time model will be used. Even though, this model was found to be appropriate for use only on slowly fading channels it is employed. Perfect knowledge of the fading process in the receiver is also assumed. The reason for using such a restrictive model is that it is commonly employed, when treating coded modulation on the Rayleigh fading channel. To enable fair comparisons, the same model should be used. Note, however, that generalization into continuous time is the subject .

The first step towards a systematic design of good codes for the Rayleigh fading channel was taken, where an upper bound on the error probability for TCM was derived. Based on this bound, the main design parameter for coded interleaved modulation on that channel was recognized as the effective diversity, which was found to be the minimum number of distinct channel symbols along any error event. Sometimes the effective diversity is also referred to as the minimum symbol Hamming distance. The product of the non-zero Euclidean distances along such an error event is found to be another, but less important design parameter . These parameters are quite different from what determines the performance of TCM on the AWGN channel. There the performance at high signal to noise ratios is mainly dependent on the minimum Euclidean distance of the scheme. This indicates that the codes developed for AWGN may not be the ones best suited for signaling on the Rayleigh fading channel.

In order to obtain a higher effective diversity than in TCM for AWGN, use of Multiple TCM (MTCM) is suggested. Here, the number of bits jointly input to the encoder is increased and the signal set is expanded accordingly by introduction of multidimensional signals. By this strategy, the effective diversity for trellises with parallel transitions becomes larger. However, for trellises having no parallel transitions only a minor improvement is achieved at the price of an increased complexity.

No TCM schemes, developed by use of either the effective diversity order or the product Euclidean distance as design criteria, were proposed. Instead, the first systematic design using these, , By considering both the effective diversity order and the product Euclidean distance, good codes were obtained. The improvements are achieved mainly for trellises with 32 states or more. Other design criteria such as Euclidean distance and error event multiplicity were considered also, giving some further improvements. Especially for small trellises, however, no significant differences compared to the codes devised for AWGN. The reason for this is that those codes attain the maximal diversity orders that are achievable for symbol interleaved TCM. For instance, for transmission of two bits per symbol encoded by an 8 state code the maximum effective diversity is only two, which is already achieved by the code .

A significant contribution to the development of TCM for the Rayleigh fading channel by generalizing the concept such that the diversity order could be further increased given a certain convolutional encoder and channel symbol set (signal constellation). The increased diversity order is here achieved by bit- instead of symbol-wise interleaving at the encoder output. In the original TCM approach, two channel symbols which are distinct over a symbol interval increase the diversity order by only one, even though the coded bits associated with the coded symbols can differ in more than one position. By bit-interleaved coded modulation (BICM) the diversity order can instead increase to the number of coded bit positions and the effective diversity order now equals the minimum Hamming distance of the code. For instance for an 8 state code with 8PSK mapping, bit-interleaving gives an effective diversity order 4 instead of 2 without bit-interleaving. The prices to be paid by the introduction of bit-interleaving are a smaller Euclidean distance and a possibly larger delay in the

transmission. Bit-interleaving decreases the Euclidean distance because the structured mapping imposed by the set partitioning in the original TCM system is destroyed. The increased delay is due to that the encoder output bits, instead of symbols, shall experience approximately independent fading. Since there is more than one encoder output bit per symbol, a larger inter leaver is needed in the bit-interleaved system to achieve the same level of independence. This leads to an increased delay.

Two other approaches have been taken in order to improve the performance of TCM on the Rayleigh fading channel. The in-phase and the quadrature-phase of a 16QAM constellation are independently encoded using two 8 state rate 1/2 convolutional encoders. This is compared to a TCM scheme employing a single 16 state code, which output is mapped onto an 8PSK constellation. It is found that the 16QAM-system has a larger diversity order, which gives a better performance. This approach was refined and more systematically studied .It was combined with bit-interleaving, which was found to give further performance improvements.

Coordinate-interleaving, whereby the signal space coordinates of the channel symbols are interleaved before transmission, was used for TCM with QAM constellations. The same strategy was repeated for MPSK modulation. This is in general less efficient e.g. a 2-D signal space, the diversity order can increase at most by a factor of 2 for each pair of distinct channel symbols. Using 8PSK, the order can be increased by a factor 3. One advantage, however, is that the Euclidean distance of the system is preserved under coordinate-interleaving, giving a better performance in AWGN .

By using a larger number of encoder output bits than in combined with bit-interleaving found that the effective diversity can be further increased .To avoid spectral expansion, the number of channel symbols in the constellation is extended accordingly. Here, this approach, referred to as Channel Symbol Expansion Diversity (CSED) is reviewed and compared to other coded modulation schemes for the Rayleigh fading channel. The maximum likelihood (ML) detector (certain restrictions applied) and a suboptimal detector for such a system are derived. The detection performance on the Rayleigh channel is illustrated using computer simulations. In order to evaluate the performance in AWGN, an equivalent Euclidean distance is derived.

5.1 Channel Symbol Expansion Diversity (CSED)

For BICM, the diversity order is the smallest number of distinct coded bits along any error event. It is possible to increase this number , by also increasing the number of positions where these bits can differ. By increasing the size of the channel symbol constellation such that channel symbols are transmitted at the same rate as before, spectral expansion is avoided. For instance instead of using a rate 2/3 convolutional encoder with 8PSK one can instead use a rate 2/4 convolutional encoder with 16PSK or 16QAM. This will leave the power spectrum the same (approximately for the latter) while increasing the diversity order.

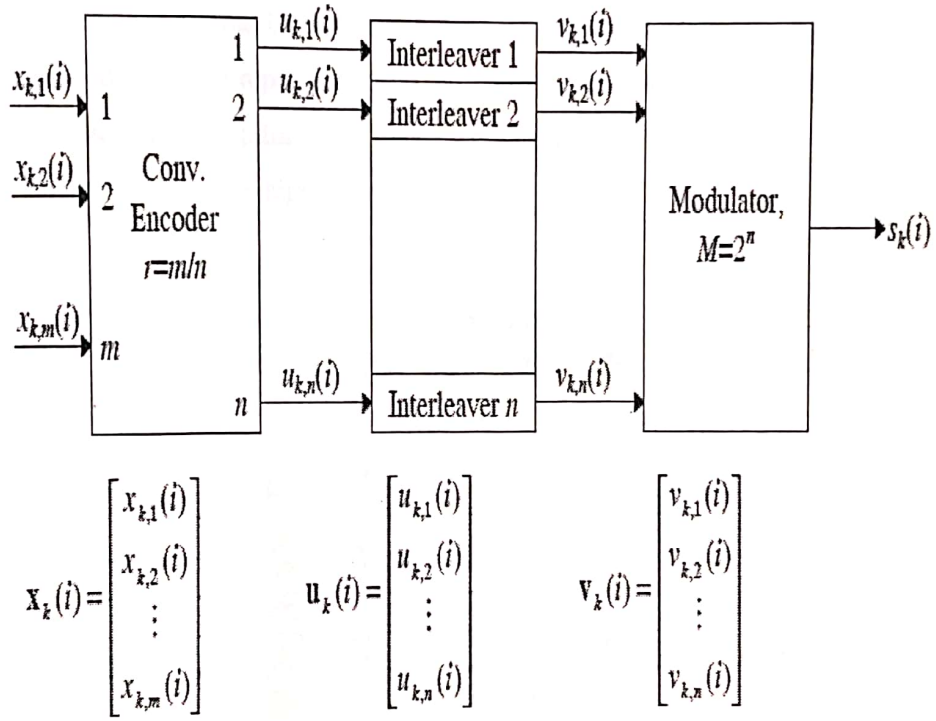


Figure 5.1 Transmitter block-diagram in a bit-interleaved CSED system.

A canonical form of the transmitter in a general bit-interleaved CSED system appears in figure 5.1. Note that it is also possible for the encoder to have only one state (no delay elements, memory less) such that an (n, m) block code is used.

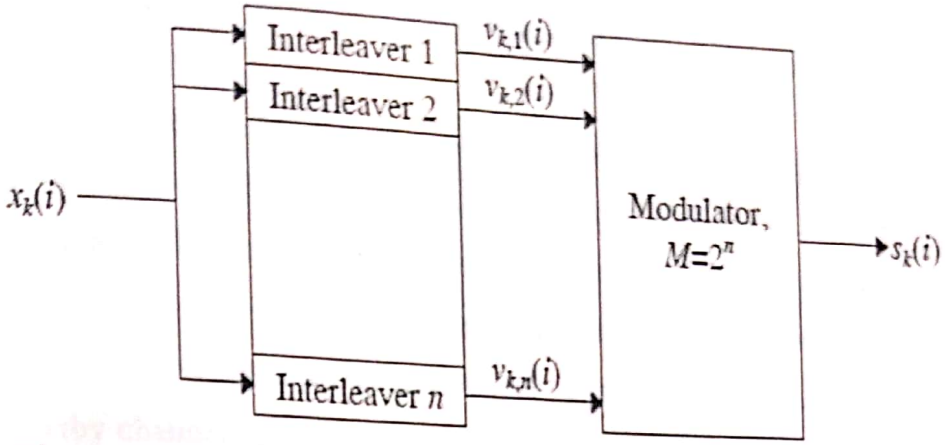


Figure 5.2 Transmitter in a repetition coded CSED system.

An interesting special case is obtained when the convolutional encoder is simply removed. Then the same information bit is mapped onto n different channel symbols. Thus the code is just a repetition code, where spectral expansion is avoided by the use of larger signal constellations. This low-complexity solution for improving the performance of uncoded signaling on

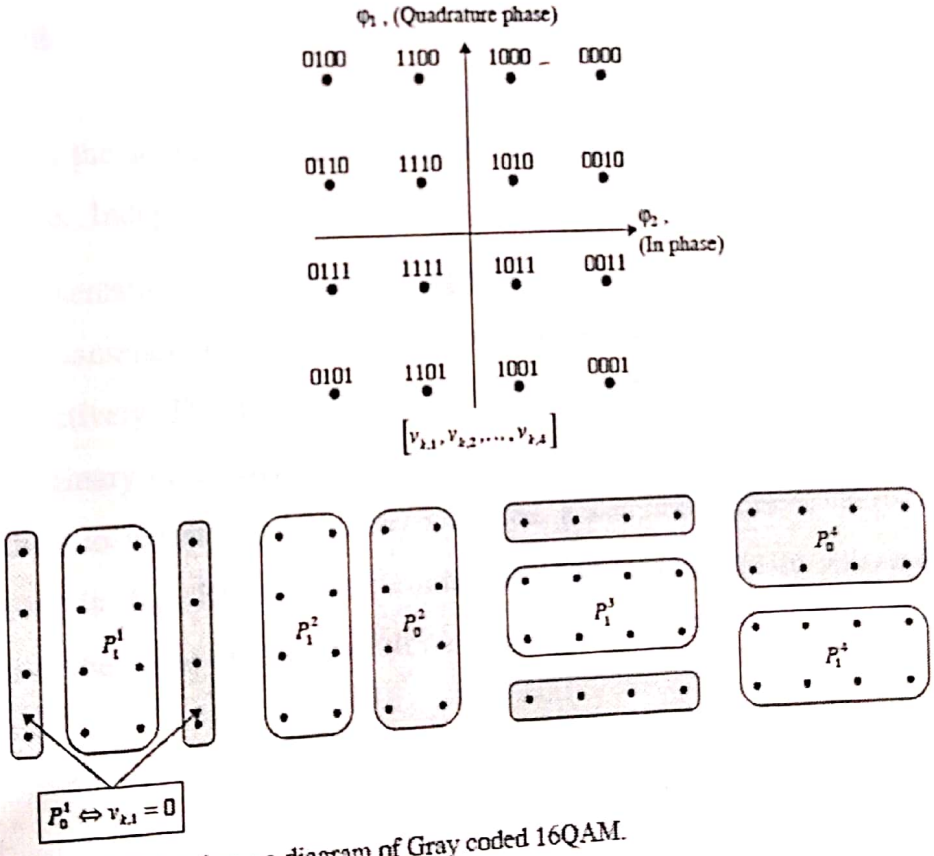


Figure 5.3 Signal space diagram of Gray coded 16QAM.

the Rayleigh channel is henceforth referred to as repetition coded CSED. The transmitter part of such a system is displayed in figure 5.2.

The mapping in the modulator is made, considering the output bits from the interleaver at each symbol interval a bit pattern ([vk,1, vk,2,...,vk,n]), which corresponds to a labeling on a point in the signal constellation. It seems reasonable to choose the mapping such that neighboring channel symbols correspond to bit labels that differ in as few bit positions as possible. By Gray coding, a mapping in which the labels of nearby channel symbols differ only in a single bit position is obtained. This choice of mapping is supported by random coding arguments given in [98]. An example of this type of mapping is given for a 16QAM constellation in figure 5.3. The channel symbols in the set denoted P_0^1 all correspond to $v_{k,l}$=0, i.e. the l:th bit input to the mapping being zero. Note that this leads to different types of partitions for different encoder output bits, which indicates that the encoder should be designed, taking the mapping into account. This is not done here, instead a fixed mapping is used according to figure 5.1. The result is that a certain encoder output bit always corresponds to the same bit position in the signal constellation.

5.2 Detection

The model for the slowly independently (between symbol intervals) fading channel is mployed here. Independence is obtained by a sufficiently large interleaver. The discrete representation of the received signal at time k is, $r_k = \rho_k e^{j\varphi_k} \cdot s_{ki} + n_k$, where $f_k = \rho_k e^{j\varphi_k}$ and n_k are complex Gaussian random variables representing the fading process and the noise process respectively. The noise and fading are independent and both have independent real and imaginary parts. S_{ki} is the discrete representation of the modulator symbol corresponding to message (sequence) i at time k and the variance of the complex valued noise is N_0. Perfect synchronization and channel state information i.e. knowledge of the phase φ_k. and envelope ρ_k of the

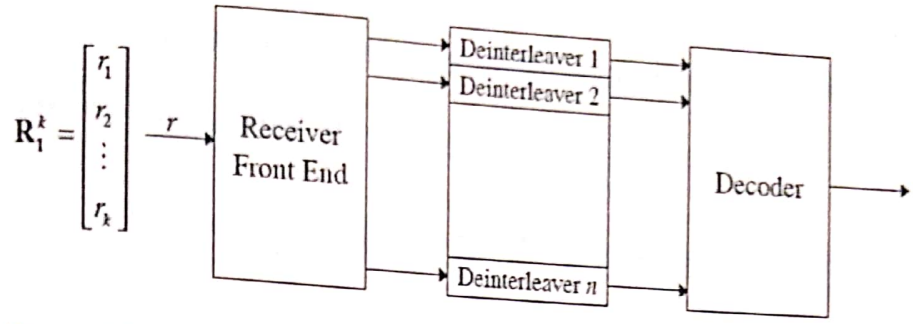


Figure 5.4 Receiver block-diagram.

fading process is assumed. Without loss of generality, the phase of the fading process in the received signal can be set to zero (φ_k.= 0) when perfect synchronization is assumed. The received signal can now be written r =.

$r_i = \rho_i s_{ki} + n_i$, where ρ_i is Rayleigh distributed. Both the channel model and the assumption of perfect knowledge of the fading process is unrealistic, especially at moderate and fast fading. However, most published results on code design are presented under these assumptions, therefore they are adopted in this chapter. In chapter 6 more realistic assumptions will be made again.

The Maximum Likelihood (ML) detector for a bit-interleaved coded modulation scheme becomes very complex. The reason is that output bits from the encoder at the same symbol interval are mapped into different channel symbols together with bits from other symbol intervals. Optimal detection in general requires (or should at least not exclude) that all information bits of a message are jointly detected, which in most cases become prohibitively complex. To obtain a detector of reasonably low complexity, a simplified suboptimal receiver structure is imposed. A block diagram is shown in figure 5.4. Instead of making a joint detection of all bits, the code and the modulation/channel are treated independently in the receiver. The receiver front end works without knowledge of the code and produces some information to be used in the decoder after de inter leaving. The decoder is just a decoder for the code and has no information about the interleaving, mapping and channel. This receiver is also applicable to repetition coded CSED. The decoder is then just a symbol detector. Under the constraints imposed by the receiver structure in figure 5.4, the maximum likelihood receiver can be derived. Input to the decoder is the a posteriori probabilities of the encoder output symbols. Thus soft, instead of hard (i.e. binary) inputs are

required in the decoder. By assuming ideal interleaving, i.e. all encoder output bits are affected by independent fading values, the optimal input at time k can be written

$$P(\mathbf{u}_k(i)|\mathbf{R},\mathbf{Z}) = P(u_{k,1}(i)|\mathbf{R},\mathbf{Z}) \cdot P(u_{k,2}(i)|\mathbf{R},\mathbf{Z}) \ldots P(u_{k,n}(i)|\mathbf{R},\mathbf{Z}) \qquad (5.1)$$

or if additive metrics are desired in the decoder (e.g. when the Viterbi algorithm is employed)

$$\log P(\mathbf{u}_k(i)|\mathbf{R},\mathbf{Z}) = \log P(u_{k,1}(i)|\mathbf{R},\mathbf{Z}) + \log P(u_{k,2}(i)|\mathbf{R},\mathbf{Z}) + \ldots + \log P(u_{k,n}(i)|\mathbf{R},\mathbf{Z}). \qquad (5.2)$$

R contains all observables of the received signal, which are available to the receiver and Z contains all side information. Here, the side information is the perfect knowledge of the envelope of the fading values (perfect synchronization is already accounted for in the channel model), i.e. $Z=(\rho 1, \rho 2, \rho 3 \ldots \ldots)^T$. The soft input needed in the decoder is thus bitwise a posteriori probabilities and the task for the receiver front end is to produce those or approximations thereof.

Now consider the receiver front end. The interleaver at the transmitter is assumed to make all transmitted symbols equally likely. Now the probability that the l:th bit at time k is zero given R and Z can be written

$$P(v_{k,l}=0|\mathbf{R},\mathbf{Z}) = \frac{\sum_{\text{all } \delta_n} P(v_{k,l}=0|s_k=\delta_n) \cdot p(\mathbf{R}|s_k=\delta_n,\mathbf{Z})}{\sum_{\text{all } \delta_n} p(\mathbf{R}|s_k=\delta_n,\mathbf{Z})} = \frac{\sum_{\delta_n \in P_0^l} p(\mathbf{R}|s_k=\delta_n,\mathbf{Z})}{\sum_{\text{all } \delta_k} p(\mathbf{R}|s_k=\delta_n,\mathbf{Z})}. \qquad (5.3)$$

where δ_n(n=1, 2,..., M) is a channel symbol in the signal constellation. The denominator is independent of the bit value, therefore it can be omitted without loss of optimality. Observe that in the second step, summation in the numerator is performed only over symbols in P_0^l , i.e. over those where v_{kl} =0. Note also that the ideal interleaving implies that only observables of the received signal from the k:th

symbol interval need to be regarded. The output from the receiver front end (i.e. the numerator of eq. (5.3)) can be simplified to

$$\sum_{\delta_n \in P_0^l} p(\mathbf{R}|s_k = \delta_n, \mathbf{Z}) = \sum_{\delta_n \in P_0^l} p(r_k|s_k = \delta_n, \rho_k). \qquad (5.4)$$

The optimal (in the sense that it enables ML decisions in the decoder) soft output is thus a sum of complex Gaussian likelihood functions of the form (again the same notation is used for the random variable and for the argument of the density function)

$$p(r_k|s_k = \delta_n, \rho_k) = \frac{1}{\pi \cdot N_0} e^{\frac{|r_k - \rho_k \cdot \delta_n|^2}{N_0}}. \qquad (5.5)$$

At high $\overline{SNR}$, the sum in eq. (5.4) becomes strongly dominated by the largest exponential, therefore an asymptotically optimal detection is obtained when at least that term is used in the soft output . When the Viterbi algorithm is employed in the decoder, it is feasible to take the logarithm of eq. (5.4) already in the receiver front end,

$$\log \sum_{\delta_n \in P_0^l} p(r_k|s_k = \delta_n, \rho_k) \overset{\text{high } \overline{SNR}}{\approx} \max_{\delta_n \in P_0^l} \left\{ -\frac{1}{N_0} |r_k - \rho_k \cdot \delta_n|^2 \right\} - \log \pi \cdot N_0 . \qquad (5.6)$$

By neglecting terms independent of $\delta_n$dn the following soft output, which leads to asymptotically (at high $\overline{SNR}$) optimal detection is obtained

$$\alpha_{k,l}(0) = \min_{\delta_n \in P_0^l} |r_k - \rho_k \cdot \delta_n|^2 . \qquad (5.7)$$

Thus the receiver front end only has to find the minimum Euclidean distances between signals from all subsets (P_0^l and P_1^l l =1, 2,..., n) adjusted by the fading envelope to the received signal at time k. Simulation results indicate that the loss due to the use of this suboptimal soft output is negligible, also at low $\overline{SNR}$, therefore this type of decoder will be employed in the following.

5.3 Related schemes

As indicated, a few schemes related to bit-interleaved coded modulation have recently been proposed , good performance is achieved by the use of separate encoders for the

in-phase and quadrature-phase components together with bit-interleaving. A block-diagram of the transmitter for such a system conveying

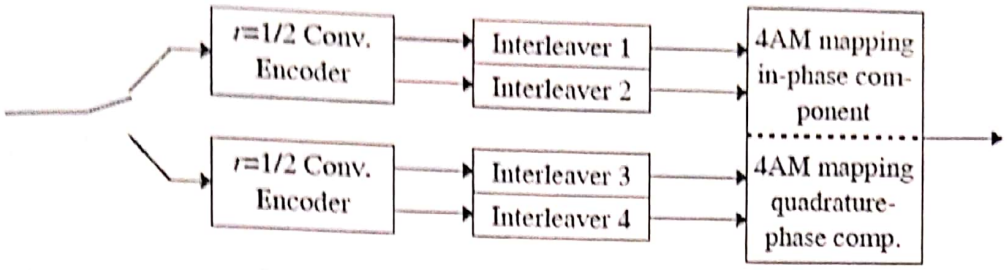


Figure 5.5 Example of a transmitter in QI-CSED.

two information bits per channel symbol is shown in figure 5.5. The same rate 1/2 code with eight states is used in each of the quadrature components and after bitwise interleaving the encoder outputs are mapped onto separate 4AM signal constellations resulting in a 16QAM signal. Since the two codes have eight states each, there is a total memory of 6 bits in the encoder. This implies that the transmitter can be viewed as being composed of a 64 state rate 2/4 encoder, which output bits are separately interleaved and mapped onto a 16QAM constellation. This system is nothing but a bit-interleaved CSED scheme, where the code is restricted to contain two separate parts corresponding to each of the input bit streams. As a result of this restriction, this system is expected to perform worse than the best 64 state bit-interleaved CSED scheme. This type of system will here be referred to as QI-CSED (Quadrature Interleaved).

QI-CSED is claimed to give a lower complexity than a general CSED scheme with the same mapping and a code having the same number of states as the individual codes in the QI-CSED system. There should be two main reasons for this. Firstly, a lower total number of comparisons is needed when using the Viterbi algorithm in two trellises instead of one. This is true since only two branches enter the same state in each of the individual trellises in QI-CSED instead of four branches for the single trellis. For 8 state codes and 16QAM mapping 16 and 24 comparisons are needed to select the survivors at each symbol interval for QI-CSED and CSED respectively. Even though less comparisons are needed in QI-CSED, the VA is applied twice requiring e.g. two instead of one set of survivor memories. Thus it is not obvious, which decoder is the more complex one.

The second reason is that, by treating 16QAM as two 4AM constellations, less computations and comparisons should be needed to obtain the soft outputs. When, however, Gray mapping according to figure 5.3 is used and the soft outputs are computed by eq. (5.7), that is not true. The reason is that for such mapping, the sets P_0^l and P_1^l (l=1,..., 4) differ either in the in-phase (I) or the quadrature-phase (Q) components (see figure 5.3). For two of the bits (l=1, 2) the Q-component can be disregarded, whereas the I-component is not needed for the other two (l=3, 4).

Another strategy related to BICM is coordinate-interleaving. As an example of this, the 8 state 8PSK scheme devised (repeating the same general approach) will be considered. It is based on a non-interleaved TCM scheme , where the channel symbols are split into their quadrature components, which are then separately interleaved. In order to achieve as high a diversity order as possible, it is important that the channel symbols have non-zero power in both quadrature components. A rotation of the signal constellation may therefore be necessary.

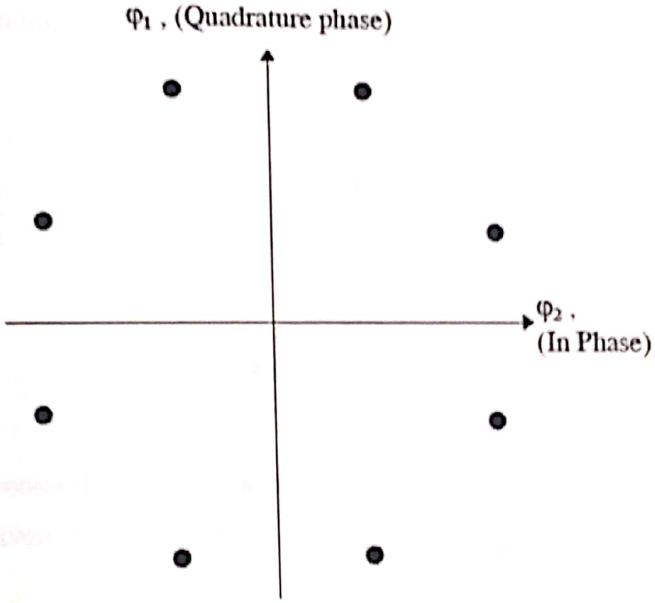


Figure 5.6 8PSK constellation

5.4 Detection performance

Both theoretical and simulation results of the detection performance of repetition coded CSED as well as convolutionally encoded CSED are presented. Comparisons will also be made with other strategies of coded modulation for the Rayleigh fading channel. The simulation results are presented as bit error rates (BER) at different

$\overline{SNR}$, defined as $\overline{E_b}/N_0$

5.4.1 Repetition coded CSED

A special case of the system introduced arises when the convolutional encoder is removed (see figure 5.2). This is in fact a repetition code where n copies of the same bit are mapped onto different M-ary channel symbols (M= 2^n). First consider the case, where the r=1/n repetition code words are mapped onto M-ary bi orthogonal signal constellations (e.g. BPSK or QPSK). Different bits will now be transmitted in distinct dimensions of a channel symbol and consequently be corrupted by statistically independent noise. Thus in contrast to detection in general CSED schemes, a joint treatment of the whole message is not necessary to achieve ML detection. Instead, the in-phase and quadrature-phase components of the received signal can be treated independently in the receiver.

Repetition coded CSED with QPSK-mapping can actually be viewed as a diversity system with n=2 statistically independent branches, where the total symbol energy is equally shared among the branches.

$$P(e)=\left(\frac{1-\mu}{2}\right)^n\sum_{k=0}^{n-1}\binom{n-1+k}{k}\left(\frac{1+\mu}{2}\right)^k \tag{5.8}$$

where

$$\mu=\sqrt{\frac{\overline{SNR}_b}{1+\overline{SNR}_b}}. \tag{5.9}$$

Here $\overline{SNR}_b$ denotes the average signal-to-noise ratio in each diversity branch. With energy sharing diversity this becomes,

$$\overline{SNR}_b=\frac{\overline{SNR}}{n}. \tag{5.10}$$

Note again that eq. (5.8) gives the error probability of the ML decoder (no constraints) for biorthogonal signals (e.g. QPSK) on the Rayleigh channel.

Figure 5.7 shows computed bit error probabilities and simulated bit error rates for different repetition coded CSED-schemes. When it comes to 8PSK (n=3) and 16QAM

(n=4) on the other hand, there are rather large differences between the computed error probabilities and the bit error rates. Here it is important to note that 3 and 4 signal space dimensions per bit are needed for the biorthogonal signals when n=3 and n=4 respectively, while only two dimensions can be used without an increase of the bandwidth. Thus the comparisons are a little unfair for the CSED schemes, where only two dimensions per bit are used.

The results in figure 5.7 indicate that repetition coded CSED is a promising low-complexity strategy to significantly improve the detection performance of uncoded signaling on the Rayleigh channel.

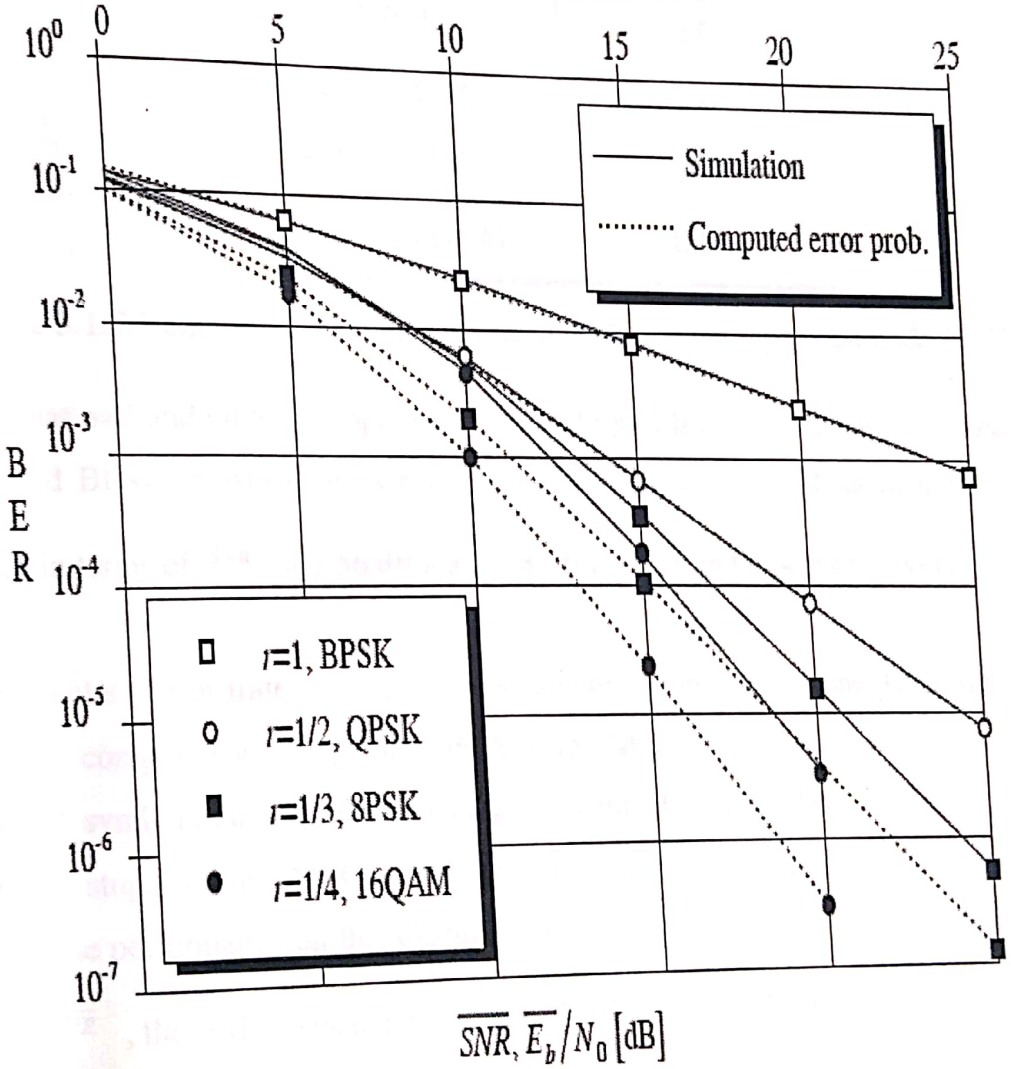


Figure 5.7 Simulation results for repetition coded CSED.

On the Gaussian channel the different schemes can be analyzed in terms of equivalent Euclidean distances. A strategy for computing those is given and it is summarized. For repetition coded CSED the equivalent squared and normalized (such that uncoded QPSK has d^2_{min} =2.0.) Euclidean distance becomes

$$d^2_{\min} = n \cdot d^2_{\text{symbol}}, \qquad (5.11)$$

where d^2_{symbol} is the minimum distance (squared and normalized) between neighboring points in the channel symbol constellation. In table 5.1 equivalent Euclidean distances are given for repetition coded CSED schemes.

System	$d^2_{\min}$
r=1, BPSK	2.0
r=1/2, QPSK	2.0
r=1/3, 8PSK	0.88
r=1/4, 16QAM	0.80

Table 5.1. Minimum equivalent Euclidean distances for repetition coded CSED

Note that n=2 and QPSK mapping implies no loss of Euclidean distance compared to n=1 and BPSK transmission, whereas the differences in Euclidean distances imply losses in terms of $\overline{SNR}$ of 3.56 dB and 3.98 dB for n=3 and n=4 respectively.

The results illustrate the differences between signaling on the Rayleigh fading channel compared with signaling in AWGN. Mapping the same bit onto different channel symbols and thereby reducing the equivalent Euclidean distance would be close to stupid on the AWGN channel. The reason for this is that the additive noise limits the performance on this channel. On the Rayleigh fading channel, especially at high $\overline{SNR}$, the performance is limited by the deep fades of the received signal power. In fact, when the $\overline{SNR}$ becomes sufficiently high, the effect of the additive noise can practically be disregarded. Transmission of the same bit in different signal constellations now makes sense, even though the resistance against the additive noise

is worse. This clearly indicates that different design criteria should be used for the AWGN and the Rayleigh fading channels.

5.4.2 Convolutionally encoded CSED

The conventional approach to performance analysis of TCM schemes is to use upper bounds on the error event probability. On the AWGN channel, the Euclidean distance d^2_{min} is a single good measure on the detection performance at moderate to high SNR. Unfortunately, on the Rayleigh channel there is no single parameter determining the performance over a large range of $\overline{SNR}$. Instead the performance is affected by different parameters such as the diversity order and both the sum and the product of the Euclidean distances along the error events. Sometimes not even these give good indications of the performance, since what is the most important parameter may vary with the $\overline{SNR}$. Of course this leads to difficulties when trying to analyze the behavior of TCM on the Rayleigh channel.

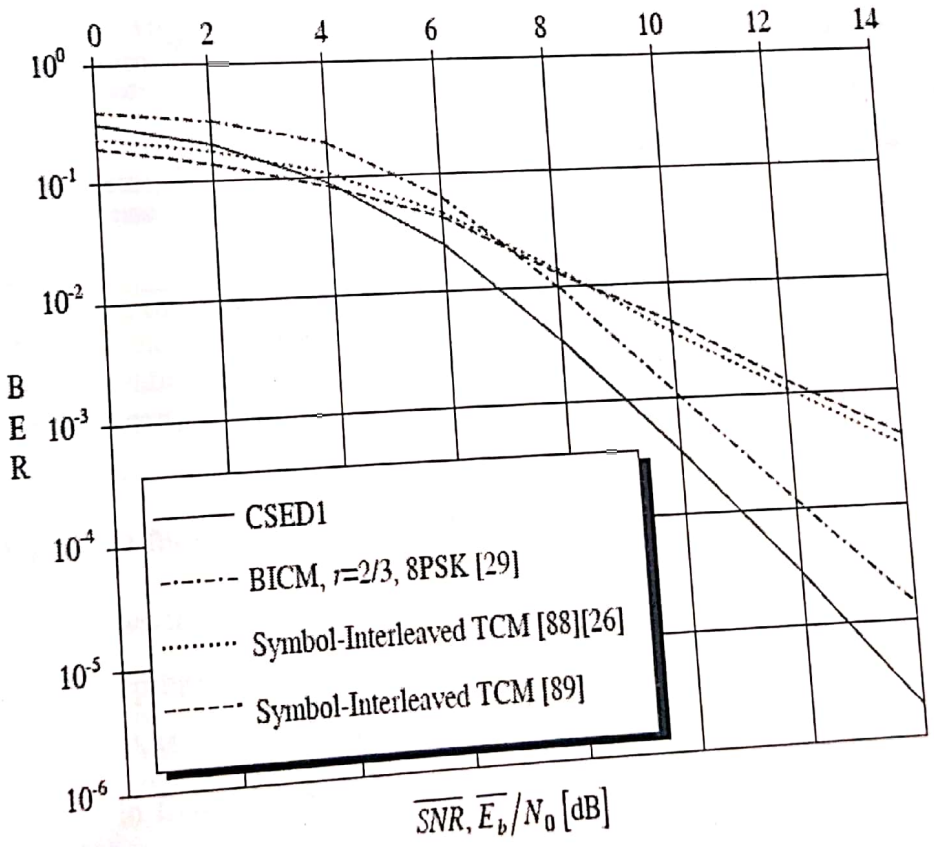
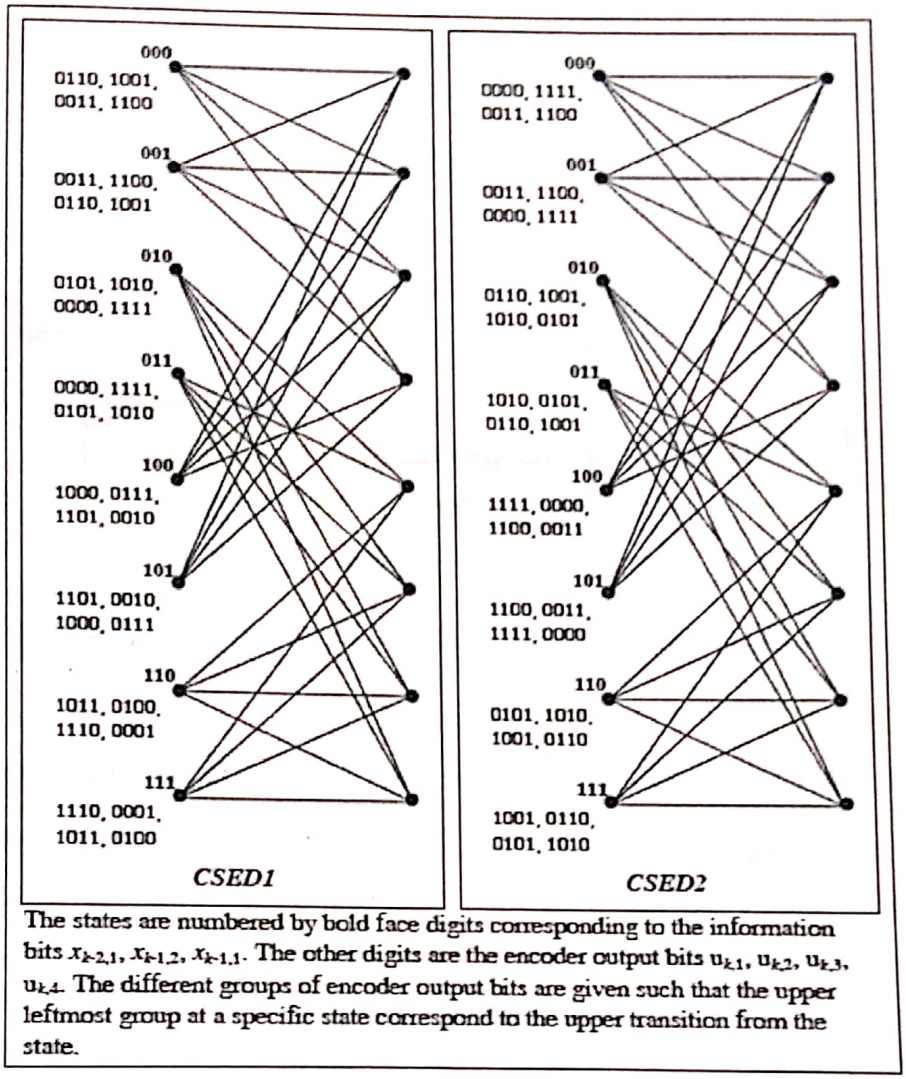


Figure 5.8 Simulation results for symbol- and bit-interleaved coded modulation on the Rayleigh fading channel.

bit-interleaved schemes (except at very low $\overline{SNR}$). Simulation results for some coded modulation schemes (all having encoders with 8 states and encoder rates r=2/n). The decoder uses the Viterbi algorithm.



The states are numbered by bold face digits corresponding to the information bits $x_{k-2,1}$, $x_{k-1,2}$, $x_{k-1,1}$. The other digits are the encoder output bits $u_{k,1}$, $u_{k,2}$, $u_{k,3}$, $u_{k,4}$. The different groups of encoder output bits are given such that the upper leftmost group at a specific state correspond to the upper transition from the state.

Figure 5.9 Trellis diagrams for CSED1 and CSED2.

Figure 5.8 shows the simulated bit error rates of the bit-interleaved (r=2/3, 8 states and 8PSK mapping). However, a convolutionally encoded CSED scheme (r=2/4, 8 states and 16QAM mapping) significantly outperforms that system. The trellis of that code is given in figure 5.9 and the system is referred to as CSED1. The difference in performance is due to the increased diversity order in the CSED scheme (Hamming distance 6 compared to 4). Also, simulation results for the two best symbol-

interleaved TCM-systems (r=2/3, 8 states and 8PSK mapping) presented so far in the literature are given (diversity order 2).

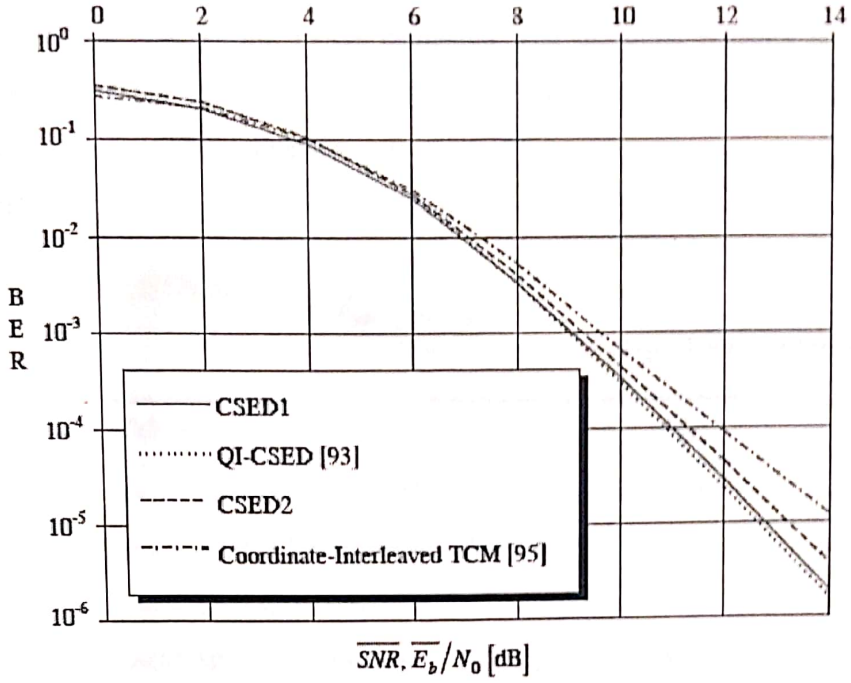


Figure 5.10 Simulation results for CSED, QI-CSED and coordinate interleaving.

Similar results for BICM are found], 64 state codes are compared and a rate 2/4 code with 16QAM mapping is about 1.5 dB better at interesting $\overline{SNR}$ than a rate 2/3 code with 8PSK mapping. Transmission of 3 information bits per channel symbol is investigated. Eight state codes of rates 3/4 and 3/5 having maximal Hamming distances (4 and 5) with 16QAM and 32Cross mapping are compared. Again the system with an expanded channel symbol set (rate 3/5 with 32Cross mapping) performs significantly better.

In figure 5.10, the above rate 2/4 code with 16QAM mapping (CSED1) is compared to the scheme, trellis diagram in figure 5.9,. Both codes are designed by hand and the difference between them is that the average multiplicity of the error event giving Hamming distance 6 is lower for CSED1 than for CSED2. This difference leads to a performance improvement, especially at moderate to large $\overline{SNR}$. Simulation results for the QI-CSED are also presented. It has the same Hamming distance as CSED1 and CSED2, but its performance is clearly better than that of CSED2. CSED1 on the other hand performs slightly better at low $\overline{SNR}$, whereas it performs a little worse at high

$\overline{SNR}$. In the same diagram, bit error rates for the coordinate-interleaved scheme are also displayed. That system performs well at low $\overline{SNR}$ due to its large Euclidean distance (d^2_{min} =4.59), but looses compared with the other schemes as the $\overline{SNR}$ becomes larger. The reason for this is that the effective diversity order is only 4, while the other systems all have diversity order 6.

By using the same technique as for the repetition coded , d^2_{min}

$$d^2_{min} = d_H \cdot d^2_{symbol}. \tag{5.12}$$

Scheme	d^2_{min}
r=2/3, 8PSK, bit-interleaved [29]	2.34
r=2/4, 16QAM, bit-interleaved CSED (CSED1)	2.40
r=2/4, 16QAM, bit-interleaved CSED (CSED2)	2.40
r=2/3, 8PSK, symbol-interleaved [26][88]	4.59
r=2/3, 8PSK, symbol-interleaved [89]	3.17
r=2/4, 16QAM, QI-CSED [93]	2.40
r=2/3, coordinate-interleaved 8PSK [95]	4.59

Table 5.2. Minimum equivalent Euclidean distances for convolutionally encoded CSED.

where d_H is the Hamming distance of the code. This formula for the equivalent Euclidean distance is also heuristically employed . Equivalent Euclidean distances for the 8 state schemes treated above are given in table 5.2. Note that the Euclidean distance is not affected by symbol-interleaving. From the table it can be concluded that the loss in terms of SNR for the CSED-schemes compared to Ungerboeck´s system is about 2.82 dB.

5.5 Summary

A new approach (bit-interleaved CSED) to coded modulation for the Rayleigh fading channel was treated. A discrete (one observable per symbol interval) channel model was employed and perfect knowledge of both the phase and the envelope of the fading process was assumed. By using codes with large Hamming distances, where the n bits

output from the encoder are mapped onto different M-ary channel symbols ($M=2^n$), a high diversity order and good spectral properties were achieved. The optimal (ML) as well as an asymptotically optimal ceiver for such systems were derived.

Both the theoretical and the simulation results presented above indicated that this scheme gives significant performance enhancement compared with TCM systems on the Rayleigh fading channel. This was achieved without sacrificing too much in complexity and performance on the AWGN channel. The CSED scheme was also found to outperform the best coordinate-interleaved system presented in the literature. Besides, comparison was made with the special type of CSED system . These schemes were found to give about the same performance at the same complexity and bandwidth. This makes systems based on the concept of CSED in combination with bit-interleaving, the best (in terms of detection performance) low-complexity schemes presented so far for use on the Rayleigh fading channel.

It was also found that a low-complexity system referred to as repetition coded CSED could give substantial performance improvements for uncoded signaling. All these results indicate that the coding strategies developed for the AWGN channel are not so well suited to the Rayleigh fading channel. For example it seems reasonable to augment the TCM structure, found by using channel capacity arguments that almost no coding gain was to be obtained by using convolutional encoders of rates other than m/(m+1). However, the above results indicate that encoders of other rates (such as r=m/(m+2)) and corresponding mappings in combination with bit-interleaving could give significant performance gains on the Rayleigh fading channel. Thus a general model of coded modulation for the Rayleigh channel should incorporate both bit-interleaving and CSED.

CHAPTER 6

Coded Modulation Time Continuous Rayleigh Fading Channel

CHAPTER 6

Coded Modulation on Time Continuous Rayleigh Fading Channel

Bit-interleaved CSED was found to give an excellent performance on the discrete channel. This channel is based on symbol rate sampling of a matched filter output, it was demonstrated that, when considering time continuous channels, its use leads to erroneous predictions of performance as well as to systems suffering from unnecessarily high error probabilities. The aim is now to combine the results into a bit-interleaved CSED scheme to be employed on the time continuous Rayleigh fading channel.

Little research has been presented in the area of combining coding, modulation and detection on the Rayleigh channel, not even for channel models less general than the one used here. Lately a few proposals of so called two-stage (concatenated) receivers have been presented. There is a certain degree of ad hoc in these procedures. In the first stage an estimate of the fading process for each symbol interval is produced. This estimate is then de interleaved and used in the second stage, which is the decoder. How the estimate is produced is the main difference between the strategies. The system employs orthogonal signals, while the schemes are based on pilot symbol assisted modulation (PSAM). The fading is estimated by interpolation, whereas interpolation in combination with a decision directed estimation scheme is employed. Decision directed strategies are used. Using a decision feedback detector, preliminary decisions of the channel symbols are made. These decisions are then used to obtain estimates of the fading. The difference is that the predictor based detector is applied to produce the preliminary decisions. The symbol-by-symbol MAP algorithm inherently produces symbol a posteriori probabilities and may therefore be used in the receiver front end.

A receiver front end producing optimal soft outputs for symbol- as well as bit-interleaved coded modulation on the Rayleigh fading channel is derived. By making appropriate approximations, a suboptimal solution of lower complexity based on the DA(B, ß) is obtained. The performance of this algorithm in combination with the coded modulation schemes in chapter 5 is investigated by means of computer simulations. The channel model is the same , i.e. a properly discretized time

continuous channel. An approximate lower bound on the performance of the optimal concatenated detector is also derived and compared with simulation results.

The starting point in the time continuous channel, together with the carefully designed receiver front end makes the treatment here unique. It is also unique in the sense that the performances of coded modulation schemes, which are among the best found so far for the channel used , now are demonstrated for the time continuous case.

This makes the results here valuable as predictions of the potential error performance for signaling on the fast Rayleigh fading channel.

6.1 System description

A model of a coded interleaved modulation system is shown in figure 6.1. Note that contrary to , the time continuous channel model is employed. The transition from the time continuous received signal to a discrete representation is made in the same way, N discrete observables per symbol interval are produced. Again it is assumed that the system

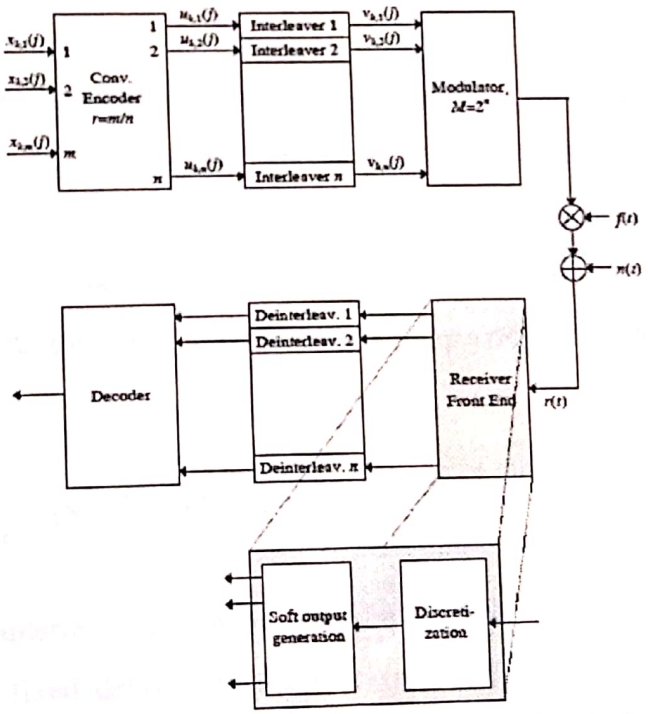


Figure 6.1 Model of a system for coded interleaved modulation on the time continuous Rayleigh fading channel.

operates with perfect knowledge of the covariances of both the fading and the noise. The optimal soft input to the decoder in a bit-interleaved system is bitwise aposteriori

probabilities ($P(u_{k,j}(j)|R,Z)$ (for the l:th en-coder output bit at time k and where j is the message number), while for a symbol-interleaved scheme it is instead symbol-wise aposteriori probabilities $P(v_{k,j}|R)$,which is to be computed for every bit position l=1,..., n and for each of the possible bit values. The problem is now to find a receiver front end, which produces these or accurate approximations thereof at a reasonably low complexity with a short fixed delay.

6.2 Receiver front end

The requirement of a short and fixed delay in the production of the soft outputs makes it feasible to replace R by R1 R_l^{k+d} , where d is a delay of an integer number of symbol intervals. The soft output for bit l being zero at time k can now be expressed as

$$P\left(v_{k,l}=0\middle|\mathbf{R}_1^{k+d}\right)=\frac{\sum_{\text{all } \mathbf{s}_1^{k+d}(i)} P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot P\left(v_{k,l}=0\middle|\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right)}{\sum_{\text{all } \mathbf{s}_1^{k+d}(i)} P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right)}=$$

$$=\frac{\sum_{\text{all } \mathbf{s}_1^{k+d}(i)\in P_0^{k,l}} P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right)}{\sum_{\text{all } \mathbf{s}_1^{k+d}(i)} P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right)}. \qquad (6.1)$$

Here $P_0^{k,j}$ is a set which contains all sequences S_l^{k+d} of channel symbols where bit l at time k equals zero, i.e. $v_{k,j}$=0. Apriori message probabilities are mainly included to allow for pilot symbol assisted modulation, but to enable an iterative soft output generation they can easily be replaced by their estimates. The denominator is here independent of the bit value, therefore only the numerator is needed. This gives the soft output

$$\sum_{\text{all } \mathbf{s}_1^{k+d}(i)\in P_0^{k,l}} P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right), \qquad (6.2)$$

which after deinterleaving can be used in an ML sequence detection (under the constraint of a fixed delay d in the receiver front end). If symbol- instead of bit-interleaving is used, symbol-wise soft outputs are required. The summation in the numerator shall then be made over sequences containing a certain symbol (instead of bit) at time k. A MAP symbol-by-symbol detector can also be based on eq. (6.2). In

this case a symbol decision, and not a soft output, is desired. As a decision, the symbol that maximizes eq. (6.2) is delivered.

Evaluation of the sums becomes prohibitively complex except for short sequences. Some strategy for complexity reduction is therefore needed. Another problem is how to evaluate the likelihood functions within the summation efficiently.

Starting with the second problem, note that the expression resembles the decision function for the MAP sequence, eq. (4.12). The difference is that messages up to time k+d instead of up to k are considered here. The strong resemblance implies that recursive computation can be obtained , in a similar way as in eq. (4.14), the likelihood function can be expressed

$$\begin{aligned} e^{\Gamma_{k+d}(i)} &= P\left(\mathbf{s}_1^{k+d}(i)\right)\cdot p\left(\mathbf{R}_1^{k+d}\middle|\mathbf{s}_1^{k+d}(i)\right) = \\ &= P\left(\mathbf{s}_1^{k+d-1}(i)\right)p\left(\mathbf{R}_1^{k+d-1}\middle|\mathbf{s}_1^{k+d-1}(i)\right)\cdot P\left(s_{k+d}(i)\right)p\left(\mathbf{r}_{k+d}\middle|\mathbf{s}_1^{k+d}(i),\mathbf{R}_1^{k+d-1}\right) = \quad (6.3) \\ &= e^{\Gamma_{k+d-1}(i)}\cdot e^{\gamma_{k+d}(i)}. \end{aligned}$$

This enables recursive computation through the relation $\Gamma_{k+d}(i) = \Gamma_{k+d-1}(i) + \gamma_{k+d}(i)$. $\Gamma_{k+d}(i)$ and $\gamma_{k+d}(i)$ are now named path and branch metrics respectively. The branch metric $\gamma_{k+d}(i)$ is computed according to eq. (4.30) giving

$$\gamma_{k+d}(i) = \log P\left(s_{k+d}(i)\right) + \log\frac{\left|\mathbf{C}_1^{k+d-1}(i)\right|}{\left|\mathbf{C}_1^{k+d}(i)\right|} + \xi_{k+d}(i). \qquad (6.4)$$

Crucial here is the computation of ξ_{k+d} .which can be made using either eq. (4.20) or (4.26) depending on the modulation format and discretization scheme. Evaluation of these equations can be realized by linear filtering as in figures 4.2 or 4.3. In the latter case the determinants in eq. (6.4) are independent of i and can then be omitted without loss of optimality. Again the same complexity problem arises , when ξ_{k+d}(i) is to be computed. The problem is that the sizes of the matrices in eq. (4.20) and (4.26) are proportional to the time index k, which leads to prohibitively large

filters as k grows large. By truncation of the time dependence , this problem is avoided at the price of suboptimality. Eq. (6.4) now becomes

$$\gamma_{k+d}(i) = \log P\left(s_{k+d}(i)\right) + \log \frac{\left|C_{k+d-b}^{k+d-1}(i)\right|}{\left|C_{k+d-b}^{k+d}(i)\right|} + \xi_{k+d}(i), \qquad (6.5)$$

where $\xi_{k+d}(i)$ can be computed by eq (4.34) or eq. (4.35).

Consider again the first problem mentioned above, i.e. the exponential increase of the number of terms in the summation in eq. (6.2). To avoid this that number must be reduced, without sacrificing too much in performance. The most accurate approximation (given a certain number of terms) is obtained when the largest terms are used. It is thus desirable to have an algorithm that finds a certain number of large terms in an efficient way. There is, however, little hope of finding say the . largest ones without performing an exhaustive search over all M^{k+d} possible sequences. Instead a suboptimal selection strategy, which finds reasonably large terms is needed. Here the DA(B,ß), can be used.

In DA(B, ß) B survivors at time k-1 are first extended by each of the M channel symbols up to time k. From all of the resulting BM sequences, ß branches grow up to time k+d giving a total of BMß sequences having large values of Γ_{k+d}(i). Both the selection of survivors at time k and the soft output generation can now be based on the branch- or path-metrics. The selection is made, choosing the B sequences maximizing eq. (4.28). This requires computation of $\wedge_k$.k()iaccording to eq. (4.15) or (4.31) for each of the BM sequences available at time k. Instead of using M^d sequences as specified by those equations, an approximate expression can be computed using only the ß sequences maintained by the DA(B, ß). Equation (4.31) can now be expressed and approximated using eq. (4.14) and (4.15) according to

$$\Lambda_k(i) = \log \sum_{\text{all } s_{k+1}^{k+d}(j)} \prod_{p=1}^{d} e^{\gamma_{k+p}(j)} \approx \log \sum_{j=1}^{\beta} \prod_{p=1}^{d} e^{\gamma_{k+p}(j)} . \qquad (6.6)$$

The sequences S_{k+l}^{k+d} (i) are all such that they originate in one of the available BM sequences at time k, i.e. s1 S_l^k(i).

Now consider generation of the soft output. For approximation of eq. (6.2) the available sequences are first divided in M groups of Bß sequences having the same symbol at time k. These groups are easily combined to the sets P0 P_o^{kj}, P_l^{kj}, (l=1,..., n) and the corresponding path metrics, Γ_{k+d}(i) are then used in the approximation of eq. (6.2), the logarithm is taken of eq. (6.2) to obtain the soft output. The soft output for bit l at time k being zero becomes

$$\log \sum_{\text{all } s_1^{k+d}(i) \in B_0^{k,j}} P\left(s_1^{k+d}(i)\right) \cdot p\left(\mathbf{R}_1^{k+d} \middle| s_1^{k+d}(i)\right) =$$

$$= \log \sum_{\text{all } s_1^{k+d}(i) \in B_0^{k,j}} e^{\Gamma_{k+d}(j)} \underset{\text{high } \overline{SNR}}{\approx} \max_{\text{all } s_1^{k+d}(i) \in B_0^{k,j}} \{\Gamma_{k+d}(i)\} \qquad (6.7)$$

where the approximation is good at high $\overline{SNR}$.

6.3 Numerical results

Here an approximate lower bound on the bit error probability for concatenated detection of interleaved coded modulation on the Rayleigh fading channel is derived. Simulation results for systems, where soft outputs are produced according to the above presentation are also given. This enables a comparison of some of the schemes on the time continuous channel.

6.3.1 Lower bound

To find a lower bound on the bit error probability, the error event probability in the decoder must first be addressed. The genie aided approach used in chapter 4.6.1 is applied here.

$$P_{\text{genie}}(e) = \sum_{\text{all } \mathbf{X}_1^k(i)} P\left(\mathbf{X}_1^k(i)\right) \cdot P_{\text{genie}}\left(\mathbf{X}_1^k(i) \rightarrow \mathbf{X}_1^k(j)\right). \qquad (6.8)$$

This is a lower bound to the error event probability of the sequence decoder. The tightest possible bound is obtained when for each message i, the sequence j is selected so that the pairwise error probability is maximized. Note, however, that finding the maximizing j for each i is a complex task, therefore a simpler strategy is needed. This can be done, since use of any specified competing sequence j, will yield a lower bound.

It seems reasonable that using an incorrect message at the minimum Hamming distance, from the correct one would give a fairly large value of the pairwise error event probability. By introduction of an indicator function , which equals one only when message I has a competing sequence at the minimum Hamming distance (otherwise it is zero) eq. (6.8) can be written

$$P_{\text{genie}}(e) = \sum_{\text{all } \mathbf{X}_1^k(i)} I\left(\mathbf{X}_1^k(i)\right) \cdot P\left(\mathbf{X}_1^k(i)\right) \cdot P_H\left(\mathbf{X}_1^k(i) \rightarrow \mathbf{X}_1^k(j)\right). \qquad (6.9)$$

$P_H\left(\mathbf{X}_1^k(i) \rightarrow \mathbf{X}_1^k(j)\right)$, denotes the probability of an incorrect choice of message j, which lies at the Hamming distance from the correct message i. Now recall that in a bit-interleaved system, the encoder output bits at a certain time are mapped onto n distinct channel symbols spaced far apart. Hence the channel symbols corresponding to message i over the error event depend not only on the bits associated with the error event, but also on other bits of message i. The concatenated receiver works without knowledge of those other bits, therefore they can be regarded being randomly elected. As a result the error event probability can be expressed in terms of the average pairwise error probability

$$P_{\text{genie}}(e) = \sum_{\text{all } \mathbf{X}_{t_1}^{t_2}(i)} I\left(\mathbf{X}_{t_1}^{t_2}(i)\right) \cdot P\left(\mathbf{X}_{t_1}^{t_2}(i)\right) \cdot \overline{P_H\left(\mathbf{X}_{t_1}^{t_2}(i) \rightarrow \mathbf{X}_{t_1}^{t_2}(j)\right)}. \qquad (6.10)$$

Here the overbar represents averaging with respect to all possible channel symbols corresponding to each of the bits of message i between t1 and t2, which are distinct from those of sequence j. Note that an average value computed over a set of events can be lower bounded by the largest value associated with an event in the set, provided it is multiplied by its probability of occurrence. Let P_{Hmax}(i, j) denote the maximal value (over all possible channel symbols) of the error event probability for message i and sequence j (at the Hamming distance from each other). Let further P_{EHmax}(i, j) denote its apriori probability. The error event probability of the genie aided decoder can now be lower bounded through

$$P_{\text{genie}}(e) \geq \sum_{\text{all } \mathbf{X}_{t_1}^{t_2}(i)} I\left(\mathbf{X}_{t_1}^{t_2}(i)\right) \cdot P\left(\mathbf{X}_{t_1}^{t_2}(i)\right) \cdot P\left(E_{H\max}(i,j)\right) \cdot P_{H\max}(i,j). \quad (6.11)$$

A lower bound to the bit error probability is obtained by multiplying eq. (6.11) by 1/m. The reason is that at least one out of m information bits in the first symbol interval of the error event must differ between any two diverging sequences.

Now consider the pairwise error event probability between two messages, where the error event starts at time t1 and ends at time t2. For a moment, general sequences not necessarily at the Hamming distance from each other are considered. The pairwise error event probability can be expressed in terms of . Λ_k (i) and . Λ_k (j) and . , which are the path metrics in the decoder for the messages i and j respectively. This probability becomes

$$P\left(\Lambda_k(j)-\Lambda_k(i)>0\middle|\mathbf{X}_1^k(i)\right)=P\left(\sum_{p=t_1}^{t_2}\sum_{l=1}^{n}\lambda_{pl}(j)-\lambda_{pl}(i)>0\middle|\mathbf{X}_1^k(i)\right). \quad (6.12)$$

The branch metrics for the l:th bit at time p of message i and j are here denoted λ_{pl}(i) λ_{pl}(i) respectively. The bitwise soft information can be used as branch metrics in the outer decoder. To simplify the computation, the optimal soft output (to be used as branch metrics) is approximated according to eq. (6.7). This equation can be rewritten in a slightly different way. The inter leavers spread and reorder the bits, therefore let bl(p) denote the time when encoder output bit l at time p appears at the inter leaver output. The notation $\mathbf{R}_1^{bl(p)}$ shall be interpreted as the set of channel symbols corresponding to the l:th bit value at time p of message i. The branch metric now becomes

$$\begin{aligned}\lambda_{pl}(i) &= \max_{\text{all } \mathbf{s}_1^{bl(p)+d}(i)\in\mathbf{R}_1^{bl(p)l}}\left\{\log P\left(\mathbf{s}_1^{bl(p)+d}(i)\right)p\left(\mathbf{R}_1^{bl(p)+d}\middle|\mathbf{s}_1^{bl(p)+d}(i)\right)\right\}= \\ &= \max_{\text{all } \mathbf{s}_1^{bl(p)+d}(i)\in\mathbf{R}_1^{bl(p)l}}\left(-\left(\mathbf{R}_1^{bl(p)+d}\right)^H\cdot\left(\mathbf{C}_1^{bl(p)+d}(i)\right)^{-1}\cdot\mathbf{R}_1^{bl(p)+d}\right)= \\ &= -\left(\mathbf{R}_1^{bl(p)+d}\right)^H\cdot\left(\mathbf{C}_1^{bl(p)+d}(i_{\max})\right)^{-1}\cdot\mathbf{R}_1^{bl(p)+d}\end{aligned} \quad (6.13)$$

where i_{max}denotes the message number maximizing eq. (6.13). In the same way let j_{max} denotes the message number maximizing eq. (6.13) for jλ_{pj}(j) . By the use of eq. (6.13) an approximation of eq. (6.12) can be obtained

$$P\left(\sum_{p=t_1}^{t_2}\sum_{l=1}^{n}\lambda_{p,l}(j)-\lambda_{p,l}(i)>0\middle|\mathbf{X}_1^k(i_{max})\right)\approx$$

$$\approx P\left(\sum_{p=t_1}^{t_2}\sum_{l=1}^{n}\left(\mathbf{R}_{bl(p)-\Delta}^{bl(p)+\Delta}\right)^H\left(\left(\mathbf{C}_{bl(p)-\Delta}^{bl(p)+\Delta}(i_{max})\right)^{-1}-\left(\mathbf{C}_{bl(p)-\Delta}^{bl(p)+\Delta}(j_{max})\right)^{-1}\right)\mathbf{R}_{bl(p)-\Delta}^{bl(p)+\Delta}>0\middle|\mathbf{X}_1^k(i_{max})\right)$$

(6.14)

Δ. is here a positive integer, introduced in order to limit the dimensionality of the vectors and matrices. This is required to make numerical computation possible. Note however, that due to this truncation and the approximation in eq. (6.13), only an approximate lower bound can be obtained.

Now consider again two sequences at the Hamming distance d_u from each other. For simplicity let R_q (q=1, 2, ..., d_u) denote the different vectors of (2Δ.+1)N observables in eq. (6.14). The right hand side of equation (6.14) can now be simplified to

$$P_{H,L}(e)=P\left(\sum_{q=1}^{d_H}\left(\mathbf{R}_q\right)^H\cdot\left(\left(\mathbf{C}_q(i_{max})\right)^{-1}-\left(\mathbf{C}_q(j_{max})\right)^{-1}\right)\cdot\mathbf{R}_q{}^H>0\middle|\mathbf{X}_1^k(i_{max})\right)$$

(6.15)

where

$$\mathbf{C}_q(i_{max})=E\left[\mathbf{R}_q\cdot\mathbf{R}_q{}^H\middle|\mathbf{X}_1^k(i_{max})\right]. \qquad (6.16)$$

This is the probability of a quadratic form in Gaussian random variables being larger than zero, which is the basic problem dealt with in the performance evaluation The quadratic form can be written

$$Q = \mathbf{R}^H \cdot \mathbf{A} \cdot \mathbf{R} \tag{6.17}$$

where $\mathbf{R} = \left[\mathbf{R}_1^T \quad \mathbf{R}_2^T \quad \cdots \quad \mathbf{R}_{d_H}^T\right]^T$ and

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}_1(i_{max})^{-1} - \mathbf{C}_1(j_{max})^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2(i_{max})^{-1} - \mathbf{C}_2(j_{max})^{-1} & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{d_H}(i_{max})^{-1} - \mathbf{C}_{d_H}(j_{max})^{-1} \end{bmatrix} \tag{6.18}$$

To employ the strategy in appendix A, the covariance of the vector R is needed. If a sufficiently large interleaver is used, the different vectors $\mathbf{R}_q$ (q=1, 2,..., d_H) are mutually statistically independent and the covariance matrix becomes

$$\mathbf{C}_\mathbf{R} = E\left[\mathbf{R} \cdot \mathbf{R}^H \middle| \mathbf{X}_1^k(i_{max})\right] = \begin{bmatrix} \mathbf{C}_1(i_{max}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2(i_{max}) & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{d_H}(i_{max}) \end{bmatrix}. \tag{6.19}$$

6.3.2 Simulation results

The time continuous channel and discretization are simulated. Discretization is performed using the ON-set. The decoder uses the Viterbi algorithm.

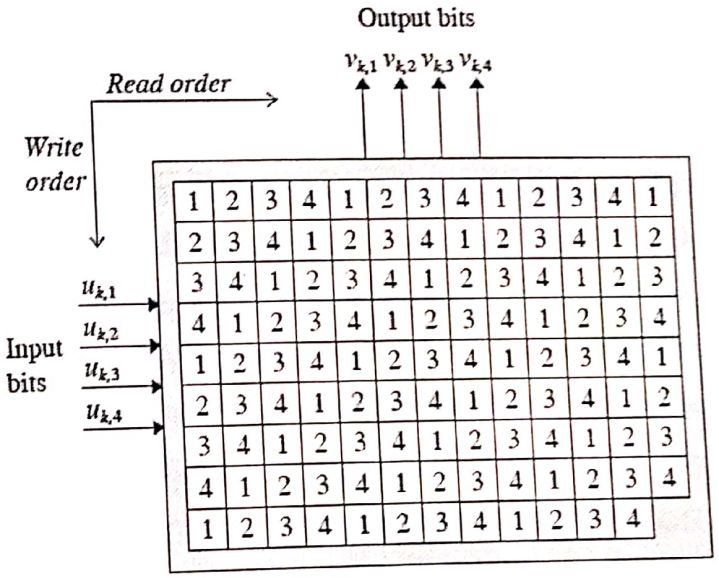


Figure 6.2 Bit-interleaver with r=2, c=3 and n=4.

Interleaving must be used to achieve a relatively high degree of independence of the

fading that affects nearby encoder outputs. This is crucial in obtaining the potential diversity effect of coded modulation on the Rayleigh fading channel. Ideal interleaving was assumed, but this is obtained only in the limit of infinitely large interleavers, i.e. for infinitely long delay. In a practical system, the size of the interleaver becomes a trade-off between delay and level of independence. For symbol-wise interleaving at fast fading, where the degree of correlation decays rapidly with time, the potential diversity effect should be attained at a low delay. Bitwise interleaving, on the other hand, requires a sufficient spreading of encoder output bits instead of symbols, therefore a larger interleaver may be needed to achieve the same degree of independence. Another problem with bitwise interleaving arises when it is used in conjunction with a fixed mapping of encoder output bits onto the channel symbols. If separate interleavers are used as in figure 6.1, then a fixed mapping is easily achieved. The problem is, however, that there is a possibility that different output bits from the same or from nearby symbol intervals are mapped onto the same channel symbol. Since interleaving is employed to avoid exactly this, using separate interleavers is no good solution. If, on the other hand, a single inter leaver is used, the fixed mapping is likely to be destroyed. The problem is that all bits jointly output from the encoder now may be mapped into a single bit position (in different symbol intervals) of the constellation.

To avoid this, a possible solution is to use a single bit-inter leaver, where some restrictions have been put on its dimensions. By choosing the number of rows and columns to rn+1 and cn+1 respectively, where r and c are integers, both problems are solved. This is exemplified in figure 6.2 for n=4, r=2 and c=3. The input bits are written column-wise and the output bits are read out row by row. The four bits input at each symbol interval are denoted 1, 2, 3 and 4. They correspond to the different encoder output bits. Note that the order of the bits when read from the inter leaver is still the same (although different bits are grouped together), which ensures that e.g. bit 1 is always mapped into the same bit position in the constellation.

Simulation of different coded modulation strategies.

Some of the systems presented are simulated here. In all cases rate 2/3, 8 state codes are used. The pilot symbol spacing is P=5, giving an effective information rate of 1.6 bits per symbol and the normalized fading bandwidth is again fn=0.1. Inter

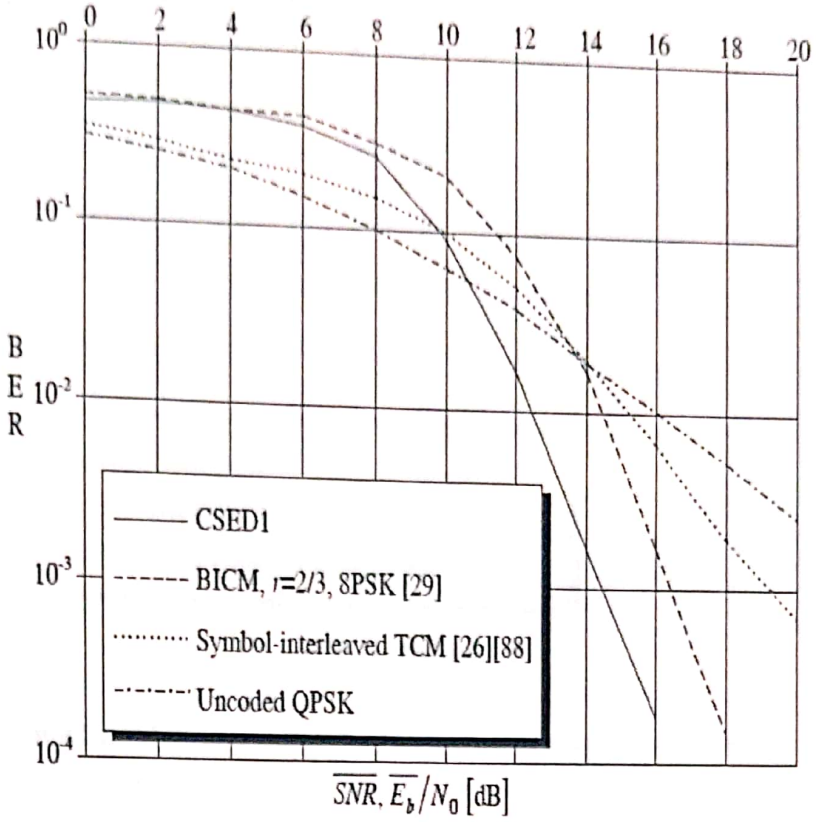


Figure 6.3 Simulation results for coded modulation systems conveying 1.6 bits per symbol.

leavers with 10 rows and columns (r=c=10) are found to give good performances for both bit- and symbol-interleaved systems at this fading bandwidth. In figure 6.3 bit-interleaved coded modulation with and without CSED is compared with symbol-interleaved TCM. The bit-interleaved schemes are the ones and CSED1 and the symbol-interleaved system is based on the TCM scheme. For comparison the performance of uncoded QPSK detected by DA(12, 0) is also displayed. In all cases the number of observables per symbol interval is N=4 and the truncation length is b=2. The receiver front end for the coded modulation schemes is based on DA(1, 12) with the delay d=5.. One interesting observation that was made in this process was that an increase of B barely affected the performance, provided ß was chosen sufficiently large.

At large SNR , on the other hand, the effective

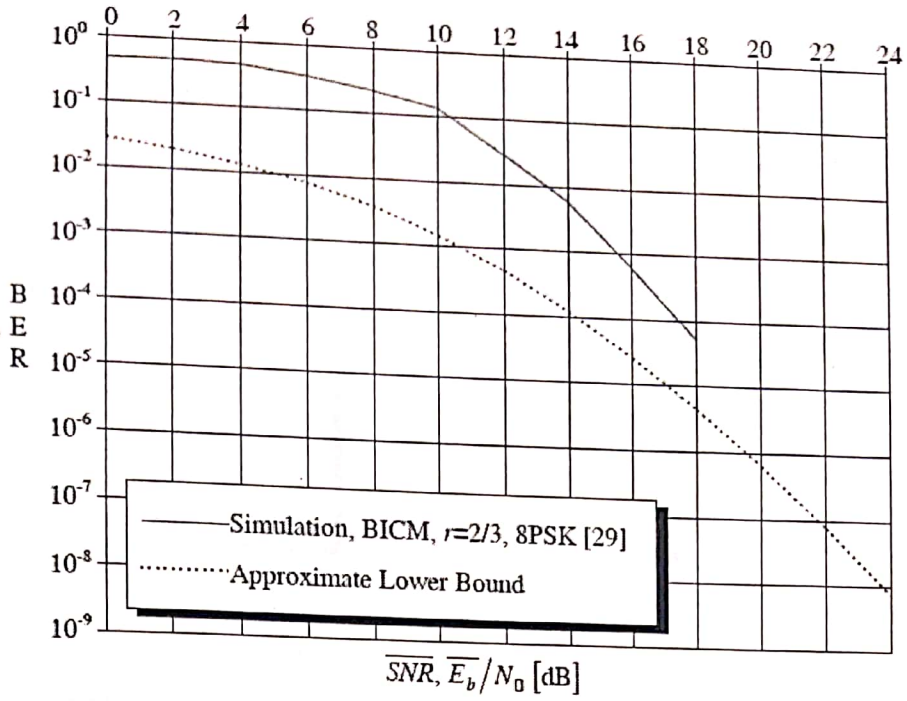


Figure 6.4 Simulation results compared with an approximate lower bound computed according to chapter 6.3.1.

diversity order dominates the performance and the bit-interleaved schemes perform significantly better than the other systems. Observe that the difference in performance between the two bit-interleaved schemes is about 2 dB, which is approximately twice as large as in chapter 5 (figure 5.8).

Figure 6.4 shows simulation results bit-interleaved coded modulation scheme (rate 2/3 encoder, 8 states and 8PSK mapping) together with an approximate lower bound computed as described in chapter 6.3.1. The pilot symbol spacing is here chosen to be P=2 (the other parameters are the same as above). At low SNR the simulation results are far from the lower bound, but the difference decreases when the SNR becomes larger and asymptotically the two curves possibly converge.

These results indicate that the approximations made in the derivation of the receiver front end are accurate, at least at high SNR . On the other hand, the very large difference between the bound and the simulation results at low SNR argue that this suboptimal receiver structure gives an unnecessarily bad performance in this region. The receiver front end when producing

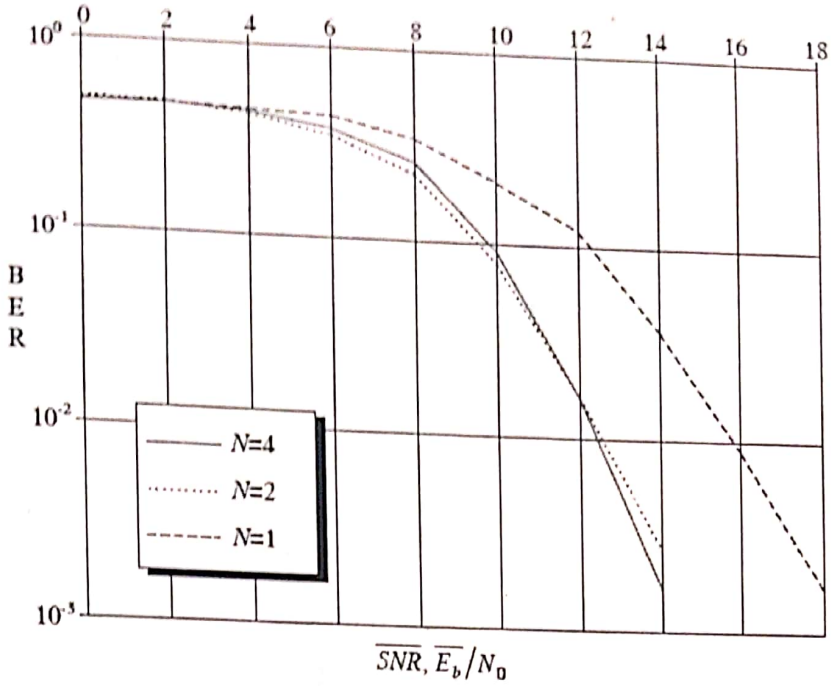


Figure 6.5 Influence of the number of observables for CSED1 when the receiver front end is based on DA(1, 12) and f_n=0.1.

soft output for say bit l at time k being zero has to consider all M/2 symbols for which that is the case, even though only one of them may have been transmitted. If some information could be provided telling the receiver front end which of the M/2 symbols is most likely to have been transmitted, the performance should be improved. A rather high price in terms of receiver complexity and delay has to be paid, but it may be rewarded by a good performance at low and moderate SNR . This is supported by the results, where a good performance is obtained by iterative decoding of symbol-interleaved TCM.

Dependence on N and on the fading bandwidth

The influence of the number of observables per symbol interval N is investigated here for CSED1. Again the receiver front end is based on DA(B, ß) and the soft outputs are produced according to eq. (6.7). The other parameters are B=1, ß=12, b=2, d=5, P=5, r=10 and c=10. Figure 6.5 indicates that for signaling at the normalized fading bandwidth fn=0.1, the performance of the matched filter based receiver (N=1) is significantly worse than for those using N=2 or 4. Note that the complexity is quadratic in N , therefore N=2 seems to give a good tradeoff between performance and complexity.

Another observation is that increasing fading bandwidths actually give better error performances for uncoded signaling, provided a sufficiently good de

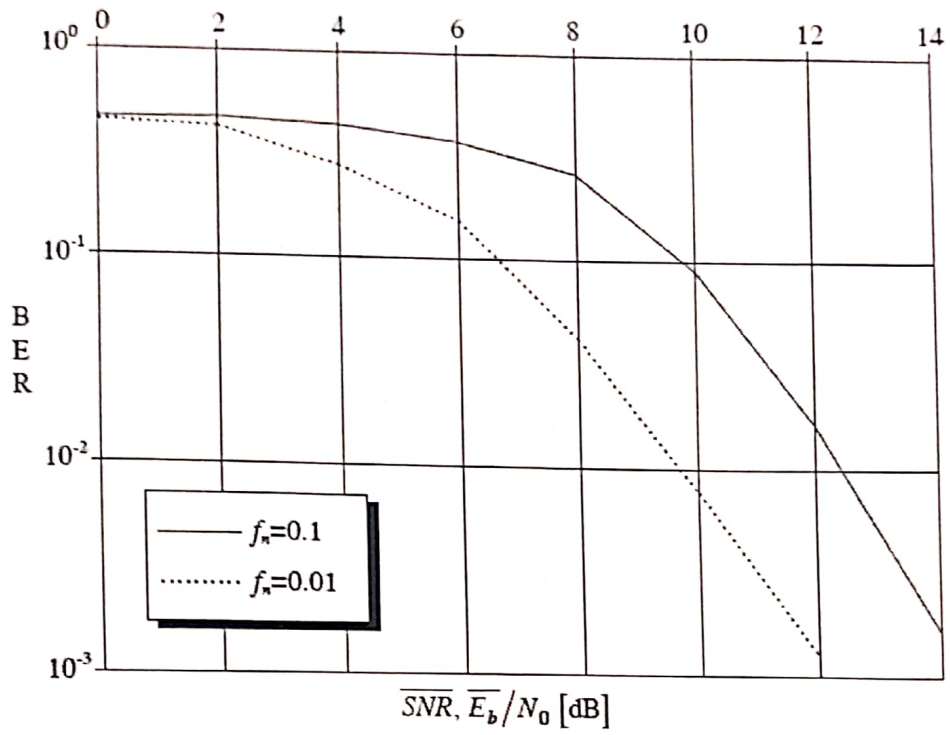


Figure 6.6 Influence of the fading bandwidth on CSED1 when a receiver front end based on DA(1, 12) is used.

tector is used. The performance improvement for sequence detection of binary PSAM at fn=0.1 compared with fn=0.01 occurred for SNR >15 dB. These improvements are due to the implicit diversity, which becomes more pronounced as the fading bandwidth increases. At average signal to noise ratios below 15 dB, on the other hand, transmission at lower fading bandwidths gives a better performance. It is at these signal to noise ratios the coded systems in this chapter operate, therefore performance improvements for signaling at higher fading bandwidths do not result here. This is illustrated in figure 6.6, where the performance of CSED1 is displayed at fn=0.1 and 0.01. PSAM with P=5 is used and the receiver front end is based on DA(1, 12). A 10 by 10 interleaver is used again and the other receiver parameters are b=2, d=5 and N=4. Clearly, the performance at fn=0.01 is significantly better than at fn=0.1. Thus if a sufficiently good code is used, signaling in slow fading seem to have a higher potential performance than signaling at fast fading.

Symbol-interleaved TCM in combination with a receiver front

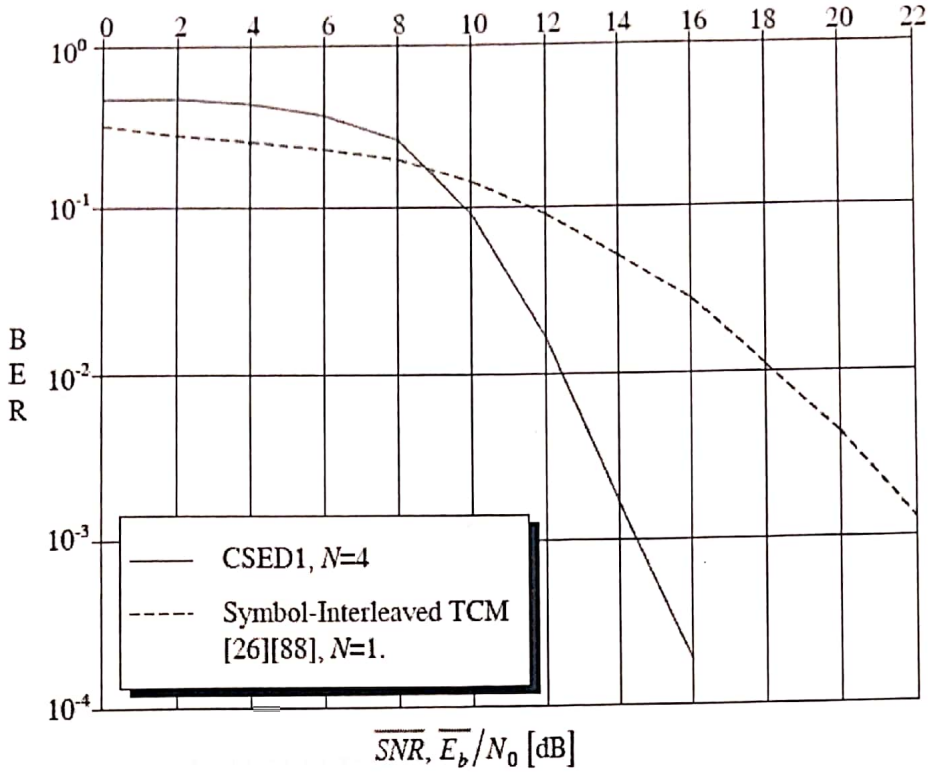


Figure 6.7 Comparison of a coded modulation system designed for the Rayleigh channel with a scheme based on strategies developed for AWGN.

end based on matched filtering is an example from the latter category. In figure 6.7 simulation results of the TCM system used above is displayed when discretization is made using a matched filter (N=1). Simulation results for CSED1 with discretization to four observables per symbol interval are also presented. The receiver front end is in both cases based on DA(B, ß), the soft outputs are generated according to eq. (6.7) and the other receiver parameters are again B=1, ß=12, b=2, d=5. A 10 by 10 interleaver is used, the pilot symbol spacing is P=5 and fn=0.1. Clearly the performance of the system especially designed for use on the Rayleigh fading channel is considerably much better than that only adapted for use on this channel.

Robustness

The robustness against erroneous estimates of channel parameters was investigated for single symbol and sequence detection respectively. Similar observations were made in the two cases. The detectors were found to be insensitive to erroneous

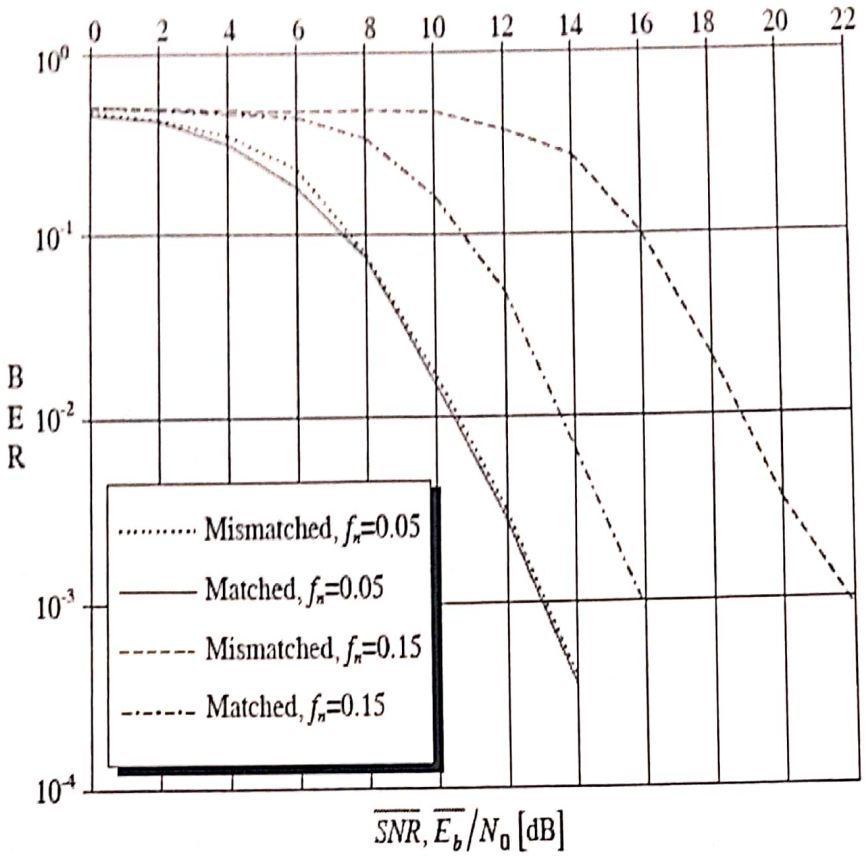


Figure 6.8 Influences on the BER of mismatches in the receiver.

estimates of the shape of the fading spectrum as well as of the signal to noise ratio. When the estimate of the fading bandwidth was too low, however, the performance loss compared with the matched detector became rather large.

Pilot symbol assisted modulation with P=5 is employed. The receiver front end uses DA(1, 12) with N=4, b=2 and d=5. It is designed for a rectangular fading spectrum with fn=0.1. The mismatch arises since the receiver is used on channels having the U-shaped spectrum with autocorrelation given, where the fading bandwidths are either fn=0.05 or fn=0.15. The results largely agree with those ,but the loss when the fading bandwidth is underestimated is even larger here.

6.4 Summary

Detection of coded interleaved modulation on the time continuous Rayleigh fading channel was considered. A receiver front end producing the soft outputs required to

make an optimal decision (ML criterion) in the decoder was derived. By suitable approximations a simplified receiver front end based on the DA(B, ß) was obtained. An approximate lower bound on the performance of the optimal concatenated receiver was also derived.

Simulations were used to evaluate some of the coded modulation schemes introduced. The bit-interleaved systems were again found to outperform the symbol-interleaved schemes and the use of CSED led to even larger improvements of the performance on this channel. The difference between CSED1 and the scheme was about twice as large (2 dB) on this channel than on the time discrete channel. This makes the system based on CSED1 with the receiver front end presented in this chapter, the most efficient coded modulation system presented so far for the time continuous frequency flat Rayleigh fading channel. This "state-of-the-art" system was about 8 dB better (at BER= 10-3) than a more conventional solution based on TCM and matched filtering (figure 6.7). Again it was found that using a sufficiently large number of discrete observables per symbol interval (at least two observables) in the receiver front end is of vital importance for the performance. The receiver was found to be robust against imperfect knowledge of the covariance of the received process. It was observed that iterative decoding might be a way to improve the performance even further at low SNR . This, however, affects the receiver and leads to a higher complexity and a larger delay. It should be noted that there are some similarities between such a scheme and multilevel coded modulation with multistage decoding . The main difference is that while a single encoder is employed in CSED, separate encoders are used for each of the bits in the channel symbol constellation in multilevel coded modulation. Thus at a given code-complexity a better performance can always be achieved by CSED, since that encoder is more general (less restricted) than the encoder jointly describing the separate encoders in multilevel coded modulation.

Contrary to , it was found that the performance is no longer improved by an increase of the fading bandwidth. The reason for this is that the introduction of coded modulation moves the operational range of the receiver towards lower signal to noise ratios, where the implicit diversity does not occur.

CHAPTER 7

Conclusions and Discussion

7.1 Conclusions and research contributions
7.2 Discussion
7.3 Future work

CHAPTER 7

CONCLUSION & FUTURE SCOPE

The problem of how to achieve efficient digital communication over the time continuous frequency flat Rayleigh fading channel has been treated. The basic motivation behind this work is the lack of knowledge of how to communicate efficiently on this channel. The problem has in fact, apart from basic studies in the 50´s and early 60´s, been largely avoided. The reason for this is probably to find in a widespread belief that efficient communication over fast fading channels is practically impossible. That is certainly the case if systems originally designed for AWGN are used. Problems arise since e.g. phase-locked loops do not work satisfactorily, matched filter detectors lead to high error floors and coded modulation schemes give performance improvements far from those needed.

In that respect, a conclusion of this work is that communication over the fast fading channel is indeed possible. However, it is not only possible, but a good performance can also be achieved, provided the properties of the channel are carefully considered in the design process. A thorough treatment of each system component with the time continuous frequency flat Rayleigh fading channel as a starting point, has for that reason been the main thread through this thesis.

7.1 Conclusions and research contributions

To achieve efficient communication over the Rayleigh fading channel, a system based on coded modulation was considered. Both the coded interleaved modulation scheme and the transition from the time continuous channel to a discrete representation in the receiver were found to be crucial to the system performance. By combining the concept of CSED (devised in this thesis) with bitwise interleaving a high effective diversity was achieved. This gives a high resistance against the deep fades of the received signal power, which leads to a good potential performance. To actually achieve this performance, the number of discrete observables per symbol interval was found to be the most important parameter. By using only a few more observables than in a matched filter based receiver, an excellent performance was achieved (far better than the system based on matched filtering). The bit-interleaved CSED scheme performed about 2 dB better than a system based on TCM in combination with

bitwise interleaving . This difference is twice as large as the difference between the schemes on a less general, but often employed time discrete channel. For this time discrete channel bit-interleaved CSED, as a general concept for coded modulation, was found to give the best performance among other known strategies.

The system was designed under the assumption of perfect knowledge of the statistical properties of the channel. Despite this, it was found to be rather robust against imperfect knowledge of the channel parameters. Estimation errors of the noise variance and the shape of the fading spectrum led to very small performance losses, whereas underestimation of the fading bandwidth gave a larger deterioration. This is no large problem, however, since overestimation of the fading bandwidth led to minor losses. Thus a practical receiver should use a biased estimate, so that a larger bandwidth than the actual one is assumed in the receiver.

The treatment of coded modulation schemes on the time continuous channel here is unique, therefore no comparisons with other investigations are made. Based on the above results it seems, however, reasonable to conclude that bit-interleaved CSED with a proper number of discrete observables enables efficient digital communication over the time continuous Rayleigh fading channel.

To reach a system with this level of performance, each part of the communication link was thoroughly designed and investigated. Apart from being steps in the system development, these studies also constitute research topics in their own. The results obtained here must be regarded equally important as those obtained for the whole system. Some of these results and conclusions are listed below (further conclusions are presented in the summaries of chapter 4, 5 and 6).

Chapter 3: Transition from the time continuous channel to a discrete representation in the receiver was investigated for transmission of a single symbol. A very general class of receivers was studied. The tool is numerical error probability calculations. The performance is highly dependent on the shapes of the modulator waveforms. This is in contrast with signaling on the AWGN channel, where the signal space constellation determines the performance.

Chapter 4: Here sequence detection of uncoded signals on the time continuous Rayleigh channel was considered. The discretization strategy was used in the receiver. Starting with the optimal detector (MAP) a suboptimal detector of lower complexity was obtained. This detector can be considered a generalization of the detector . A lower bound on the error probability of the optimal detector was derived and simulations indicated that the error probability of the suboptimal detector closely approached this when the signal to noise ratio becomes large. Thus its performance is close to that of the MAP sequence detector, even though its complexity is significantly lower.

Extensive simulation results were presented. Again, the number of observables were found to be crucial to the performance. Another observation made in chapters 3 and 4 was that the performance of the optimal receiver is improved by an increase of the normalized fading bandwidth. This effect is a result of the implicit diversity and is obtained provided a sufficiently good detector is used. This result also indicates that a decrease of the symbol rate on the channel can improve the performance for uncoded signaling. In chapter 6 it was observed that this does not hold when efficient coded modulation schemes are used.

Chapter 5: Coded modulation for a time discrete channel model was studied. Bit-interleaved CSED was proposed and was found to give significantly lower error probabilities than other known low complexity strategies. The comparison with symbol-interleaved TCM indicated that the TCM structure developed for AWGN may be too restrictive to achieve a good performance in AWGN. A more general and less restrictive structure should therefore include both bitwise interleaving and CSED. The receiver front end producing soft information enabling an ML decision in the decoder was also derived.

Chapter 6: Here coded modulation for the time continuous frequency flat Rayleigh fading channel was treated. A new (compared with the one in chapter 5) derivation of the receiver front end was presented. After deriving the receiver front end that enables ML decisions in the detector, a suboptimal solution of lower complexity was presented. It is based on the sequence detection algorithm presented in chapter 4. The

conclusions drawn in chapter 6 largely coincides with the general conclusions for the entire system presented above.

7.2 Discussion

A main thread through this thesis is the starting point in the most general model of the Rayleigh fading, i.e. the time continuous description. This is in contrast to the common approach of using methods developed for the AWGN channel. These are often adapted to the channel through the discrete time channel model presented, which was found to be useful only on slowly fading channels.

One example of this is the (mis) use of matched filter based detectors. Those are actually the basis for the time discrete channel model, but as mentioned above they lead to unnecessarily bad performances at moderate or fast fading channels.

Another example is Trellis Coded Modulation (TCM). Not until the main design rule for AWGN, i.e. set-partitioning, was relaxed by the introduction of bit-interleaving a reasonably good performance was achieved. By also increasing the size of the channel symbol constellation further compared with that of TCM an even better performance was obtained.

One can of course argue that few, if any, systems will be used solely on the fast Rayleigh fading channel. A likely situation is that the Rayleigh channel is a worst case encountered only occasionally. A design, as in this thesis, which only focus on that special case may give a bad overall performance. However, there is no reason whatsoever that the results obtained here cannot be accommodated within a more general channel model, which certainly would be desired.

The transition from the time continuous received signal to a discrete representation in the receiver was here made using projection of the received signal onto a receiver signal space of finite dimensionality. Note that even though sampling, as a method for discretization, was used only in a single case it fits well in such a signal space description. The principal results about discretization, should thus hold for sampling as well as for the other strategies employed in this thesis.

Space diversity, i.e. multiple receiver (or sometimes also transmitter) antennae, are often used in communication systems. This is not considered in this thesis, but by a straightforward extension in the derivations it can be accommodated. The only thing needed is a concatenation of vectors of observables obtained from each diversity path into a joint received vector. The observables are, as before jointly zero mean complex Gaussian, and for that reason all derivations in the thesis apply from this point.

7.3 Future work

Extension of this work to include more general channels, where both time dispersion and a strong component are present in the received signal is an interesting topic for further research. Essentially the same strategy as the one used here would be applicable.

The starting point should be in a time continuous model, probably with known statistical properties. After derivation of the optimal detector, simplified suboptimal solutions would almost certainly be desired. The techniques for performance analysis should also be extended to this more general situation. Detection in presence of time-varying inter symbol interference is yet another situation, where methods and results developed for AWGN are not directly applicable. It is shown that, unless some structure of an unknown channel impulse response is assumed, it is impossible to derive an ML sequence detector in AWGN. This is, however, no problem on the frequency selective fading channel when the receiver operates with perfect knowledge of the second order statistics of the channel.

In chapter 6, reasons were presented for a potential performance improvement to be achieved by the use of iterative decoding. The receiver front end developed in that chapter would be almost directly applicable as a key component in such a decoder. The price to be paid by the use of iterative decoding is a large delay and a high receiver complexity. Iterative decoding is also a key component of a turbo coded modulation system.

A completely different, but technically very interesting approach would be an effort to further decrease the complexity of the system in chapter 6. One approach could be to make further approximations in the algorithm used in the receiver front end. Another interesting strategy would be to combine interpolation of the fading values between

the pilot symbols with the system devised in this thesis. The utility of such an approach would be that expensive filter operations may be avoided, at least partly. In any case, the results presented in this thesis, would be useful for performance comparisons.

Another interesting issue is that of finding good upper bounds on the error probabilities for bit-interleaved coded modulation on the time discrete as well as the time continuous fading channel. The problem is that most of the bounds developed for coded modulation are far from being asymptotically (at high signal to noise ratios) tight on the Rayleigh channel.

References

REFRENCES

1. P. A. Dighe, R. K. Mallik, and S. S. Jamuar, ``Analysis of transmit-receive diversity in Rayleigh fading," IEEE Transactions on Communications, vol. 51, no. 4, pp. 94-703, April 2003.

2. R. K. Mallik, ``On multivariate Rayleigh and exponential distributions," IEEE Transactions on Information Theory, vol. 49, no. 6, pp. 1499-1515, June 2003.

3. R. K. Mallik, D. Singh, and A. Kumari, ``Analysis of rake reception with multiple symbol weight estimation for antipodal signaling," IEEE Transactions on Communications, vol. 51, no. 10, pp. 1721-1729, Oct. 2003.

4. R. K. Mallik, P. Gupta, and Q. T. Zhang, ``Minimum selection GSC in independent Rayleigh fading," IEEE Transactions on Vehicular Technology, May 2005.

5. U. Hansson, T. Aulin, "Channel Symbol Expansion Diversity", IEE Electronics Letters, Vol. 31, No. 18, pp. 1545-1546, 31st August 1995.

6. U. Hansson, T. Aulin, "Channel Symbol Expansion Diversity for Trellis Coded Modulation on the Rayleigh Fading Channel", in Conf. Records of RVK´96, Luleå,, Sweden, pp. 380-384, June 1996.

7. U. Hansson, T. Aulin, "Channel Symbol Expansion Diversity-Improved Coded Modulation for the Rayleigh Fading Channel", in Conf. Records of ICC´96, Dallas, TX, pp. 891-895, June 1996.

8. U. Hansson, T. Aulin, "Aspects on Binary Signaling on the Rayleigh Fading Channel", Tech. Report No. 272, Dept. of Computer Eng., Chalmers Univ. of Techn., Nov. 1996.

9. U. Hansson, T. Aulin, "Aspects on Detection of Binary Signals on the Rayleigh Fading Channel", in Conf. Records of ISIT´97, Ulm, Germany, p. 194, June 1997.

10. U. Hansson, T. Aulin, "Aspects on Single Symbol Signaling on the Rayleigh Fading Channel", Submitted to IEEE Trans. Commun., June 1997.

11. U. Hansson, T. Aulin, "Soft Information Transfer for Sequence Detection with Concatenated Receivers", IEEE Trans. Commun., Vol. 44, No. 9, pp. 10861095, Sept. 1996.

12. U. Hansson, T. Aulin, "Theoretical Performance Evaluation of Different Soft-Output Algorithms", in Conf. Records of ISITA´94, Sydney, Australia, pp. 875880, Nov. 1994.

13. www.wikipedia.com