Sequential Monte Carlo Methods for Digital Communications

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

Elena Punskaya, St. John's College

July, 2003

Signal Processing and Communications Laboratory
Department of Engineering
University of Cambridge

Declaration

I hereby declare that this dissertation is the result of my own work, performed at Cambridge University between October 1999 and July 2003, and includes nothing which is the outcome of the work done in collaboration. It is not substantially the same as any dissertation that I have submitted, or am currently submitting, for a degree or diploma or other qualification at any other University. The dissertation contains less than 65 figures and its length does not exceed 40000 words.

Acknowledgments

My sincerest thanks are due to all the people who have inspired and supported me during the course of this thesis. My gratitude goes to Prof. Bill Fitzgerald who gave me the opportunity to explore my own ideas, while tactfully ensuring that I follow a constructive path. I greatly value this freedom and especially appreciate his positive attitude. Beyond the call of duty, his advice has been always readily available, his enthusiasm true source of inspiration, his dedicated friendship more than I could have asked for.

I own my deepest thanks to Dr. Arnaud Doucet for his constant guidance through the evolution of this thesis, for his willingness to share his thoughts and ideas, and for believing in me when so many would not. Without him, this work would *never* be possible. No matter where he was, his help has been always at hand, his fascination with science infectious, his care, support and friendship invaluable. So many times his words turned the moments of desperation into moments of inspiration with new insights and goals to pursue. I am and always will be indebted to him.

I am also thankful to Dr. Christophe Andrieu for his companionship and help when I started investigating MCMC and SMC methods. From him, through the numerous fruitful discussions and helpful comments, I benefited immensely. I am grateful to Simon Maskell for constructive and insightful comments which led us to a more critical evaluation of our methodology, to Prof. Peter Rayner for creating a professional and stimulating atmosphere and ensuring we are on the right track, to Arnaud Doucet, Tarik Tutill and Nick Haan for proof-reading this tome so carefully, to Dr. Kate Pretty and Dr. Helen Watson, whose professionalism inspired me, and many others.

I gratefully acknowledge the financial support of the EPSRC, ORS Award scheme, St. John's College, and Homerton College, University of Cambridge.

Last, but by no means least, I would like to thank *all* my family for bearing me (and all those papers everywhere), for unselfishly sharing all my ups and downs, and for all the unconditional love, happiness and support they brought into my life. My especial thanks go to my husband and my daughter Anastasia who participated keenly in the preparation of this thesis.

Summary

Due to the dramatic increase in the number of users and their demand for more advanced services, the need for fast and accurate filtering techniques in digital communications, capable of coping with challenging transmission conditions, is becoming more and more prevalent. This research investigates the application of a relatively new and very promising approach - Sequential Monte Carlo methods - to the problems associated with the transmission over a digital link.

In this thesis, a general framework for addressing a number of digital communications applications is proposed, and several specific tasks are considered as an illustration of this general approach. First, the problem of demodulation of M-ary modulated signals under conditions of flat Rayleigh fading channels in the presence of non-Gaussian additive noise is addressed. The particle filtering receiver is then extended to include joint symbol decoding and demodulation, space diversity combining and multiuser transmission. Joint symbol detection and propagation delay estimation in direct sequence spread spectrum systems in multipath environment is treated at the end of the thesis.

We develop an efficient simulation-based algorithm, based on particle filtering, which combines sequential importance sampling, a selection scheme and several variance reduction techniques. Computer simulations show a good performance of the proposed method for a number of applications. For the class of problems involving the discrete unknown parameters only, several other deterministic and randomized techniques applicable in this scenario are reviewed and tested. To the best of our knowledge, such comparison has never been made before. The conclusion is in favour of the simplest deterministic algorithm, which, unfortunately, cannot be applied straightforwardly to the problems with both discrete and continuous-valued parameters being unknown. A new particle filtering receiver incorporating both deterministic and stochastic schemes to consider this more complicated scenario is derived at the end. An extensive simulation study is carried out throughout the thesis.

Keywords

Bayesian methods, decoding, demodulation, digital communications, Rayleigh fading, multipath, multiuser detection, multiple antennas, particle filter, Sequential Monte Carlo, spread spectrum systems.

NOTATION

a	Scalar
$\mathbf{a}$	Column vector
a_i	ith element of a
$a_{k,i:j}$	Sequence $a_{k,i:j} \triangleq (a_{k,i}, a_{k,i+1}, ..., a_{k,j})^{\mathsf{T}}$ if $i < j$, $a_{k,i:j} \triangleq (a_{k,i}, a_{k,i-1}, ..., a_{k,j})^{\mathsf{T}}$ otherwise
$\mathbf{a}_{k,i:j}$	Sequence $\mathbf{a}_{k,i:j} \triangleq (\mathbf{a}_{k,i}, \mathbf{a}_{k,i+1}, ..., \mathbf{a}_{k,j})^{\mathsf{T}}$ if $i < j$, $\mathbf{a}_{k,i:j} \triangleq (\mathbf{a}_{k,i}, \mathbf{a}_{k,i-1}, ..., \mathbf{a}_{k,j})^{\mathsf{T}}$ otherwise
$\mathbf{a}^{(i:j)}$	Sequence $\mathbf{a}^{(i:j)} \triangleq \left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}, ..., \mathbf{a}^{(j)}\right)^{\mathsf{T}}$ if $i < j$, $\mathbf{a}^{(i:j)} \triangleq \left(\mathbf{a}^{(i)}, \mathbf{a}^{(i-1)}, ..., \mathbf{a}^{(j)}\right)^{\mathsf{T}}$ otherwise
$\mathbf{I}_n$	Identity matrix of dimension $n \times n$
$\mathbf{A}$	Matrix
$\mathbf{A}^{\mathsf{T}}$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	Determinant of matrix $\mathbf{A}$
$\mathbf{a} \sim p\left(\mathbf{a}\right)$	$\mathbf{a}$ is distributed according to distribution $p\left(\mathbf{a}\right)$
$\mathbf{a}\rvert\, \mathbf{y} \sim p\left(\mathbf{a}\right)$	The conditional distribution of $\mathbf{a}$ given $\mathbf{y}$ is $p\left(\mathbf{a}\right)$
$\mathcal{N}\left(\mathbf{m}, \mathbf{\Sigma}\right)$	Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$
$\mathcal{N}_c\left(m, \Sigma\right)$	Complex Gaussian distribution
$\lfloor a \rfloor$	Highest integer strictly less than a
$\{a\}$	$\{a\} \triangleq a - \lfloor a \rfloor$
$\mathbb{E}(\mathbf{a})$	Expectation of the random variable $\mathbf{a}$
$var(\mathbf{a})$	Variance of the random variable $\mathbf{a}$
$\delta(t)$	Dirac delta function
$\arg\max\limits_{\mathbf{a}}$	The argument $\mathbf{a}$ that maximizes the operand

OTHER NOTATION

k	denotes kth information source,	$k = 1, \ldots, K$
l	denotes lth antenna (subchannels),	$l = 1, \ldots, L$
h	denotes hth chip,	$h = 1, \ldots, H$
g	denotes gth channel path,	$g = 0, \ldots, G-1$
m	denotes one of possible κ-bit symbols, $M = 2^\kappa$,	$m = 1, \ldots, M$
$\xi^{(l)}$	denotes $\xi^{(l)}$th component of Gaussian mixture,	$\xi^{(l)} = 1, \ldots, Z$
i	denotes ith particle,	$i = 1, \ldots, N$

Other symbols used throughout the thesis

$d_{k,n}$	nth information κ-bit symbol from the kth information source
$r_{k,q}$	qth κ-bit coded symbol from the kth information source
$s_{k,q}(\cdot)$	Mapping function (digital symbols $\rightarrow$ analogue waveforms)
$c_{k,1:H}$	Spreading code sequence
$f_{k,t}^{(g,l)}$	Channel coefficients for the kth user, lth subchannel, gth path at instant t
$\theta_{k,t}$	Propagation delay for the kth user at instant t
$\varepsilon_t^{(l)}$	Additive noise at lth subchannel (mixture of zero-mean Gaussians) $\varepsilon_t^{(l)} = \sigma_{z_t}^{(l)} \epsilon_t^{(l)}$, $\epsilon_t^{(l)} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$
$z_t^{(l)}$	Allocation variable for the lth subchannel (identifies component in the Gaussian mixture)
$y_t^{(l)}$	Output of the lth subchannel at instant t
T_n	Symbol transmitted interval
T_q	Coded symbol transmitted interval
T_{ch}	Chip interval
T_s	Nyquist sampling rate
$1/R$	Code ratio
$w(\cdot)$	Importance weight
$\pi(\cdot\vert\cdot)$	Importance distribution

ABBREVIATIONS

AGC	Automatic gain controller
APF	Auxiliary particle filter
APP	*a posteriori* probability
AR	Auto-regressive
ARMA	Auto-regressive moving average
BDPSK	Binary differential phase shift keyed
BER	Bit-error-rate
CDMA	Code division multiple access
DDE	Discounted detection estimation
DEA	Detection-Estimation Algorithm
DFE	Decision feedback estimator
DML	Deterministic most likely
DPSK	Differential phase shift keyed
DS	Direct sequence
DSR	Deterministic stratified resampling
EKF	Extended Kalman filter
EM	Expectation-maximization
FEC	Forward error correcting
GPB	Generalised Pseudo-Bayes
HMM	Hidden Markov model
ISI	Intersymbol interference
IMM	Interacting multiple model
JMLS	Jump Markov linear system
LPI	Low probability of being intercepted
MA	Moving average
MAP	Maximum *a posteriori* probability
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MF	Matched filter

MAI	Multiple-access interference
MMAP	Marginal maximum *a posteriori*
MMSE	Minimum mean square estimate
MSE	Mean-square error
OFDM	Orthogonal frequency division multiplexing
PFO	Particle filter with optimal importance distribution
PFP	Particle filter with prior importance distribution
PFS	Particle filter with suboptimal importance distribution
PLL	Phase locked loop
PN	Pseudo-noise
PSAM	Pilot symbol-aided schemes
PSK	Phase shift keyed
QAM	Quadrature amplitude modulated
RLW	Resampling particles with low weights
RSA	Random Sampling Algorithm
SIR	Sequential importance resampling
SMC	Sequential Monte Carlo
WSS	Wide-sense stationary

Contents

1 Introduction

Modern communication systems have undergone a remarkable development over the past decades, with digital communication techniques dominating analogue methods. Mobile phones and the internet have become such a great part of our lives that we simply cannot imagine what would we do without them. Speed, safety and reliability are now matters of primary concern, and the need for an advanced technology to meet these often challenging requirements is ever growing. The Sequential Monte Carlo methods show great potential in providing such a powerful mathematical technique, and their applications in the field of digital communications are the subject of this thesis.

1.1 Motivation, Scope and Contributions

Filtering in one form or another has been with us for centuries. In the past man tried to remove the impurities in his water to obtain a pure drink. Similarly, in the modern age, we seek to remove the unwanted, usually random, noise corrupting the signals. The most obvious situation where this is required is in communication systems, where the transmitted signal is corrupted in a random manner by a variety of possible mechanisms, such as thermal noise or interferences of different kinds. Recovering the original message from the observations in this case is a complicated (filtering) problem, which, except for several special cases, still presents a major challenge to researchers.

The task has proved to be especially difficult under conditions of fading communication channels. With rapid advances in modern technologies and increasing access to, and use of, digital communications, the demand for fast and accurate filtering techniques capable of coping with such conditions is becoming more and more prevalent. However, physical limitations and impairments of the transmission channels such as multiple-access and co-channel interferences, time-variation and frequency selectivity, amplitude fading and phase distortion, multiplied by the potential user mobility and the random nature of the system in general, make the situation extremely problematic. The task can be greatly facilitated by the use of an efficient signal processing technique, and the Sequential Monte Carlo (SMC) (or *particle filtering*) approach, recently emerged in the field of statistics and engineering, is believed to be such a key leading technology.

The Sequential Monte Carlo methods are a set of powerful and versatile simulation-based algorithms to perform optimal state estimation in nonlinear non-Gaussian state space models. The approach has recently received a lot of interest since it allows a large number of challenging non-linear estimation tasks to be addressed in an efficient on-line manner; see [Doucet *et al.*, 2001a] for a survey. The idea is to approximate the posterior distribution of interest by swarms of weighted points in the sample space, called particles, which evolve randomly in time according to a simulation-based rule, and either give birth to offspring particles or die according to their ability to represent the different zones of interest of the state space.

Since many problems arising in digital communications can be considered as optimal filtering problems, the SMC seems to be a sensible choice, and this research investigates the application of these promising techniques to a variety of problems associated with digital communications. In order to develop a general framework for addressing a number of applications, we formulate the model in a general form, including such essential concepts as *channel coding*, *modulation*, *spread spectrum* techniques, *space diversity* and *multiuser* transmission, and represent each individual task as a particular case of this general system. The transmission in fading environment is considered throughout the thesis, and we begin our treatment with demodulation of symbols in flat fading conditions with a possibly non-Gaussian additive channel noise. We then show how to extend the proposed approach for the detection of coded sequences, space diversity and multiuser demodulation. Joint symbol detection and propagation delay estimation in direct sequence (DS) spread spectrum systems is considered later in the thesis. The advantage of the proposed approach is the fact that many other applications could be addressed in the same framework, such as channel and code delay estimation, for instance. Moreover, further extensions including more complicated scenarios (asynchronous code division multiple access (CDMA) is one of the examples) could be obtained straightforwardly using the developed methods.

During our research, we identified two classes of problems typical for digital communications. In the *first* one, the unknown state of the model - usually the transmitted symbol(s) - takes its values in a finite set; this includes, for example, demodulation in fading channels (Chapter 4 and 6), orthogonal frequency division multiplexing (OFDM) systems and multiuser detection in synchronous CDMA (Chapter 5). In the *second* class, one faces a more challenging problem where the unknown state of interest consists not only of the symbol(s) but also some continuous-valued parameters such as code delays as in DS spread-spectrum system analyses (Chapter 7).

For the first class, we begin with an efficient particle filtering algorithm, which is designed to make use of the structure of the model, and incorporates efficient variance reduction strategies based on Kalman filtering techniques and the use of importance distribution proven to be *optimal* in this framework (Chapter 4). The detection problem in this case can be reformulated as an optimal filtering problem for jump Markov linear systems; i.e. linear Gaussian state space models switching according to an unobserved finite state space Markov chain. Several other deterministic and stochastic approaches are applicable in this scenario, and we review and compare them in this thesis (Chapter 6). Such comparison has never been done before, and one might find the obtained results surprising.

Unfortunately, these detection schemes cannot be applied straightforwardly to solve the problems of the second class. However, taking these results into account, we develop and test several algorithms based on these approaches, to obtain an efficient SMC receiver for the applications involving both discrete and continuous-valued unknown parameters. An extensive simulation study is presented for all examples considered in the thesis.

To summarize, this thesis proposes:

- a general framework for addressing a variety of digital communication applications using Sequential Monte Carlo methods;
- an efficient particle filtering receiver for demodulation of symbols in flat fading conditions, incorporating variance reduction technique and optimal importance distribution;
- its straightforward extensions to the non-Gaussian additive noise scenario, joint symbol decoding and demodulation, space diversity and multiuser detection among others;
- a general review and comparison of deterministic and randomized approaches for the problems involving jump Markov linear systems;
- an efficient SMC algorithm based on a combination of deterministic and stochastic schemes for the problems involving both discrete and continuous-valued parameters;
- an efficient technique for joint symbol/code delay estimation in DS spread spectrum systems in a multipath environment with potential extension to asynchronous CDMA;
- an extensive simulation study of all the algorithms developed.

1.2 Thesis Outline and Reading Directions

As a guide to reading this thesis, we now briefly summarize its structure. The thesis consists of eight chapters, presenting introductory and review material (Chapters 1, 2 and 3), our analysis and novel contributions (Chapters 4, 5, 6 and 7), and some conclusions and directions for future research (Chapter 8).

Mainly, the work might be of interest to two groups of people with a background in communications and in simulation-based methods, and is organised with that in mind. Although we hope everyone finds the overview of a digital communication system (Chapter 2) and non-linear and non-Gaussian filters, including particle filters, (Chapter 3) useful, these chapters could be skipped if necessary. Chapters 4, 6 and 7, describing the algorithms developed in this thesis, are self-contained and can be read independently. For each chapter, at the beginning, we briefly recall the model for the digital communications application under consideration, then describe the proposed algorithm(s), and, finally, present corresponding computer simulations followed by a discussion. It is, however, preferable to read Chapter 5 in conjunction with the preceding chapter, since it is based on the material presented there. In general, those readers, whose main research activities lie in the area of Sequential Monte Carlo Methods might be interested in the results obtained in Chapters 6 and 7. The receivers developed in Chapters 4, 5, 6 and 7, hopefully, will be of some benefit to the readers with communications background.

Chapter 1: Introduction

This chapter has begun the thesis with an exposition of the topic, providing the motivation and perspective of the work as well as highlighting its main contributions. The chapter also summarizes the structure of the thesis, suggests some reading directions and concludes with a list of publications based on the described work.

Chapter 2: A Digital Communication System

The basics of a modern digital communication system employing channel coding, modulation, spread spectrum techniques, space diversity and multiuser transmission are described in this chapter. A general mathematical model representing typical characteristics of these main components as well as the multipath fading transmission channel with possibly non-Gaussian additive noise is then presented. The estimation objectives for this general set-up, and several particular digital communication problems, are stated in the following section, which also serves as a *road map* to the rest of the thesis. Two classes of problems typical for digital communications are discussed at the end.

Chapter 3: Bayesian Filtering

The chapter introduces the Bayesian approach to filtering, with several nonlinear and non-Gaussian filters previously proposed in the literature being reviewed. The mathematical formulation of the general state-space model studied in this dissertation is presented. A particular emphasis is placed on the relatively novel Sequential Monte Carlo approach, and the basic ideas of particle filtering techniques and some strategies for their efficient implementation are discussed.

Chapter 4: Particle Filtering for Demodulation in Flat Channels

The problem of demodulation of M-ary modulated signals in flat Rayleigh fading channels in the presence of non-Gaussian additive noise is addressed here. An efficient simulation-based algorithm based on SMC methods and combining sequential importance sampling, a selection scheme, and several variance reduction techniques is developed. An application of the algorithm to uncoded and coded M-ary phase shift keyed (PSK), differential phase shift keyed (DPSK) and quadrature amplitude modulated (QAM) signals is presented and an extensive simulation study is carried out.

Chapter 5: Extended particle filtering receiver

The chapter extends the proposed particle filtering demodulator to perform joint demodulation and decoding, optimal (space) diversity combining, and multiuser detection. The results obtained by means of computer simulations are discussed at the end of the chapter.

Chapter 6: Detection Estimation Algorithm for Demodulation

The particle filtering algorithm developed in the previous chapters is reviewed and compared with alternative deterministic and stochastic algorithms previously presented in the literature for this class of problems. We discuss and test these approaches for demodulation and CDMA detection in flat fading channels, and demonstrate their performance by means of computer simulations.

Chapter 7: Joint Symbol/Delay Estimation in Spread Spectrum Systems

Finally, we develop a new receiver for joint symbol, channel characteristics and code delay estimation for DS spread spectrum systems under conditions of multipath fading. The proposed algorithms are based on the results obtained in the previous chapter and combine sequential importance sampling, deterministic scheme and selection. An extensive simulation study is carried out and demonstrates the performance of the suggested approaches.

Chapter 8: Conclusion

This final chapter closes the thesis by summarizing and discussing the results obtained during this research, and considering some possible future directions arising from this work.

1.3 Publications

Some of the material in this thesis has been, or will be, published elsewhere. Technical reports and non-reviewed material are not included. In some cases, the conference papers contain material overlapping with the journal publications.

Journal Papers

- E. Punskaya, C. Andrieu, A. Doucet and W. Fitzgerald, "Particle Filtering for Demodulation in Fading Channels with Non-Gaussian Additive Noise," *IEEE Transactions on Communications*, vol. 49, no. 4, pp. 579-582, 2001.
- E. Punskaya, C. Andrieu, A. Doucet and W. Fitzgerald, "Optimal Demodulation of QAM in Fading Channels using Particle Filtering," *Signal Processing,* under review.
- E. Punskaya, A. Doucet and W. Fitzgerald, "Particle Filtering for Joint Symbol and Code Delay Estimation in DS Spread Spectrum Systems in Multipath Environment," *EURASIP Journal on Applied Signal Processing,* under review.

Book Chapters

- C. Andrieu, A. Doucet, E. Punskaya, "Sequential Monte Carlo Methods for Optimal Filtering," in *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, 2001.

Conference papers

- E. Punskaya, C.Andrieu, A. Doucet and W. Fitzgerald, "Particle Filters for Demodulation of M-ary Modulated Signals in Noisy Fading Communication Channels", in Proc. *IEEE Conf. ICASSP*, 2000.
- E. Punskaya, C.Andrieu, A. Doucet and W. Fitzgerald, "Particle filtering for multiuser detection in fading CDMA channels", in Proc. *IEEE SSP* 2001.
- E. Punskaya, A. Doucet and W. Fitzgerald, "On the use and misuse of particle filtering in digital communications", in Proc. *EUSIPCO* 2002.
- E. Punskaya, A. Doucet and W. Fitzgerald, "Particle Filtering for Joint Symbol and Parameter Estimation in DS Spread Spectrum Systems", in Proc. *IEEE Conf. ICASSP*, 2003.

2 A Digital Communication System

In a *digital* communication system, the information generated by a source or a number of sources, whether analog or discrete, is processed in a digital form, i.e. as a sequence of binary digits obtained after source encoding. A *communication channel* providing a link between the transmitting and receiving ends of the system is an important consideration in the system construction. Its characteristics generally affect the design of the basic elements of the system, a detailed description of which is given in this section.

The overview below incorporates such essential concepts as *channel coding*, *modulation*, *spread spectrum* techniques, *space diversity* and *multiuser* transmission. A specified mathematical model represents typical characteristics of these main components and the transmission channel.

We consider here a rather general, modern system responsible for the transmission of information generated by *several* information sources to one or *more* destinations. The basic elements of such a system are illustrated in Figure 2.1, and we start with the treatment of these main components. We then define the estimation objectives and the digital communications problems associated with them, and, finally, discuss two classes of problems identified in this thesis.

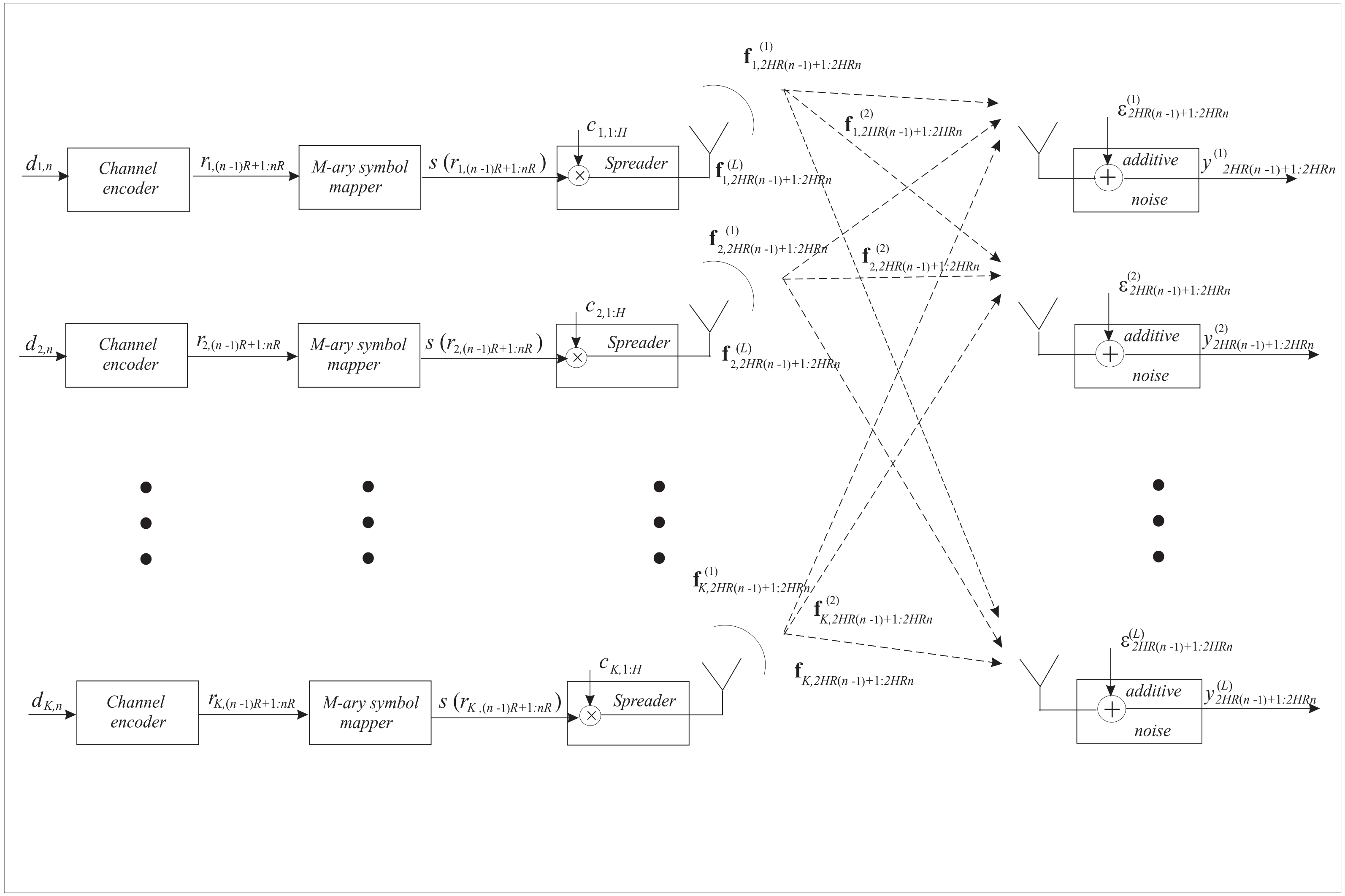

Figure 2.1: Digital communication system.

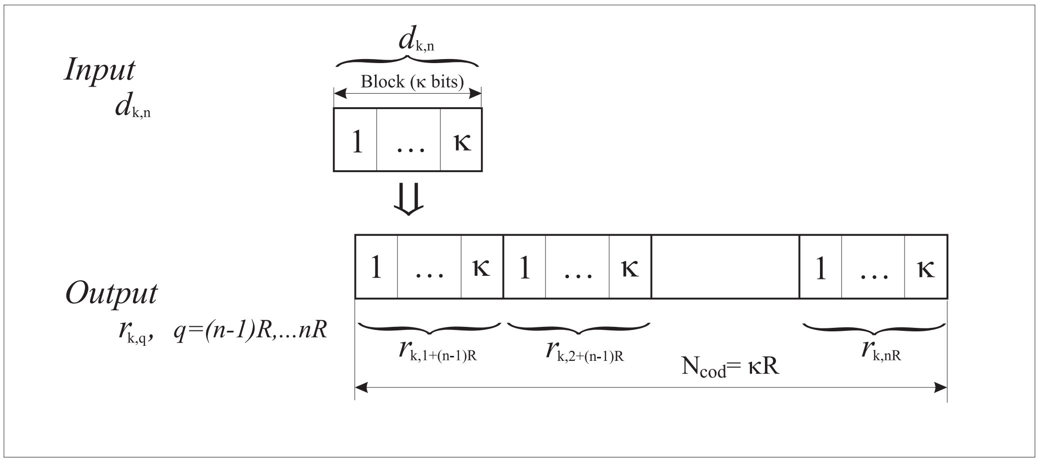

Figure 2.2: Channel Encoding.

2.1 Digital communication system model

Hereafter, we denote for any sequence α_n, $\alpha_{i:j} \triangleq (\alpha_i, \alpha_{i+1}, \ldots, \alpha_j)^{\mathsf{T}}$, and for any sequence $\alpha_{k,n}$, $\alpha_{k,i:j} \triangleq (\alpha_{k,i}, \alpha_{k,i+1}, \ldots, \alpha_{k,j})^{\mathsf{T}}$.

2.1.1 Channel encoding

In a well-designed digital system, each information sequence is first passed to a channel encoder which introduces, in a controlled manner, some carefully designed structure to a data word in order to protect it from transmission errors [Lin & Costello, 1983]. This added redundancy enables a receiving system to detect and possibly correct errors caused by corruption from the channel, and, thus, improve the accuracy and reliability of the system in the presence of noise and interferences. This process is also known as forward error correcting (FEC), and can be employed as follows.

The encoder takes κ information bits at a time from the original data stream, and maps them in a certain known way into a unique N_{code} -bit sequence called a *code word* (see Figure 2.2 for an example of channel encoding). For clarity of presentation, we assume that $N_{code} = \kappa R$, thus resulting in R output 2^{κ}-ary symbols per input data word and a *code ratio* of $1/R$.

Let $d_{k,n}$ be an indicator variable associated with one of $M = 2^{\kappa}$ possible κ-bit symbols for the nth input to the encoder from the kth information

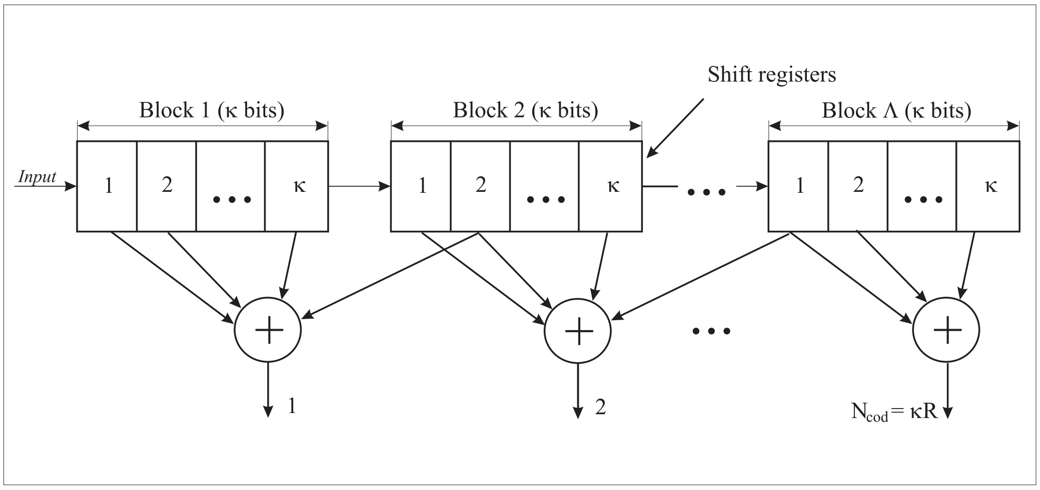

Figure 2.3: Convolutional encoder.

source:

$$d_{k,n} \in \mathcal{M} = \{1, 2, \ldots, M\}, \quad n = 1, 2, \ldots,$$

and $r_{k,1+(n-1)R:nR}$ be the corresponding output, which in the general case of the code with memory depends on several preceding data words:

$$r_{k,1+(n-1)R:nR} = \Phi\left(d_{k,n-\Lambda+1:n}\right), \tag{2.1}$$

where $\Phi(\bullet)$ is a known mapping function, and Λ is, for example, the *constraint length* for the convolutional code in κ-bit bytes; $r_{k,q}$ indicates one of $M = 2^\kappa$ possible symbols (see Figure 2.3):

$$r_{k,q} \in \mathcal{M} = \{1, 2, \ldots, M\}, \quad q = 1, 2, ...,$$

Thus, $q = 1 + (n-1)\,R, \ldots, nR$ output 2^κ-ary symbols correspond to the original input symbol $d_{k,n}$:

$$d_{k,n} \leftrightarrow r_{k,1+(n-1)R:nR} \quad n = 1, 2, \ldots$$

2.1.2 Interleaving

Although usually transmitted data are interleaved as well as coded in order to avoid the *burst errors* exhibited by different classes of channels, no interleaving is assumed for the system described in this thesis for reasons specified in Chapter 5.

2.1.3 Modulation

The non-binary code described above is particularly matched to a M-ary modulation technique, for which the κ-bit symbols are mapped into $M = 2^{\kappa}$ deterministic, finite energy analogue waveforms $\{s^{(m)}_{\text{trans, }k}(\tau),\ m = 1, 2, \ldots, M\}$ so that the information could be transmitted over the communication channel. The signal waveform for the kth user corresponding to the qth symbol may be represented as

$$s_{\text{trans, }k}(\tau) = \mathrm{Re}[s_{k,q}(r_{k,1:q})u_k(\tau)\exp(j2\pi \mathrm{f}_{\text{car}}\tau)], \quad (q-1)T_q \leq \tau \leq qT_q, \tag{2.2}$$

where f_{car} is a carrier frequency, $s_{k,q}(\cdot)$ performs the mapping from the digital sequence to waveforms and, in a general case of a modulator with memory (for example, when *differential encoding* is employed, see Chapter 4), it depends on one or more previously transmitted symbols $r_{k,1:q}$, and T_q is the symbol duration (the duration of the coded word $r_{(n-1)R+1:nR}$ is then T_qR). The form of the waveform $u_k(\tau)$ is discussed in the next subsection.

In this thesis we assume that a linear modulation scheme such as phase shift keying (PSK) or quadrature amplitude modulation is employed. Differentially encoded symbols are also considered in Chapter 4. The corresponding forms of the function $s_{k,q}(\cdot)$ are presented in Chapter 4.

2.1.4 DS Spread-Spectrum techniques

In order to enable the transmission of several distinctive information-bearing signals through the same channel, direct sequence (DS) spectrum spreading is used in the system.

Originally, this technique was developed for military purposes for low-

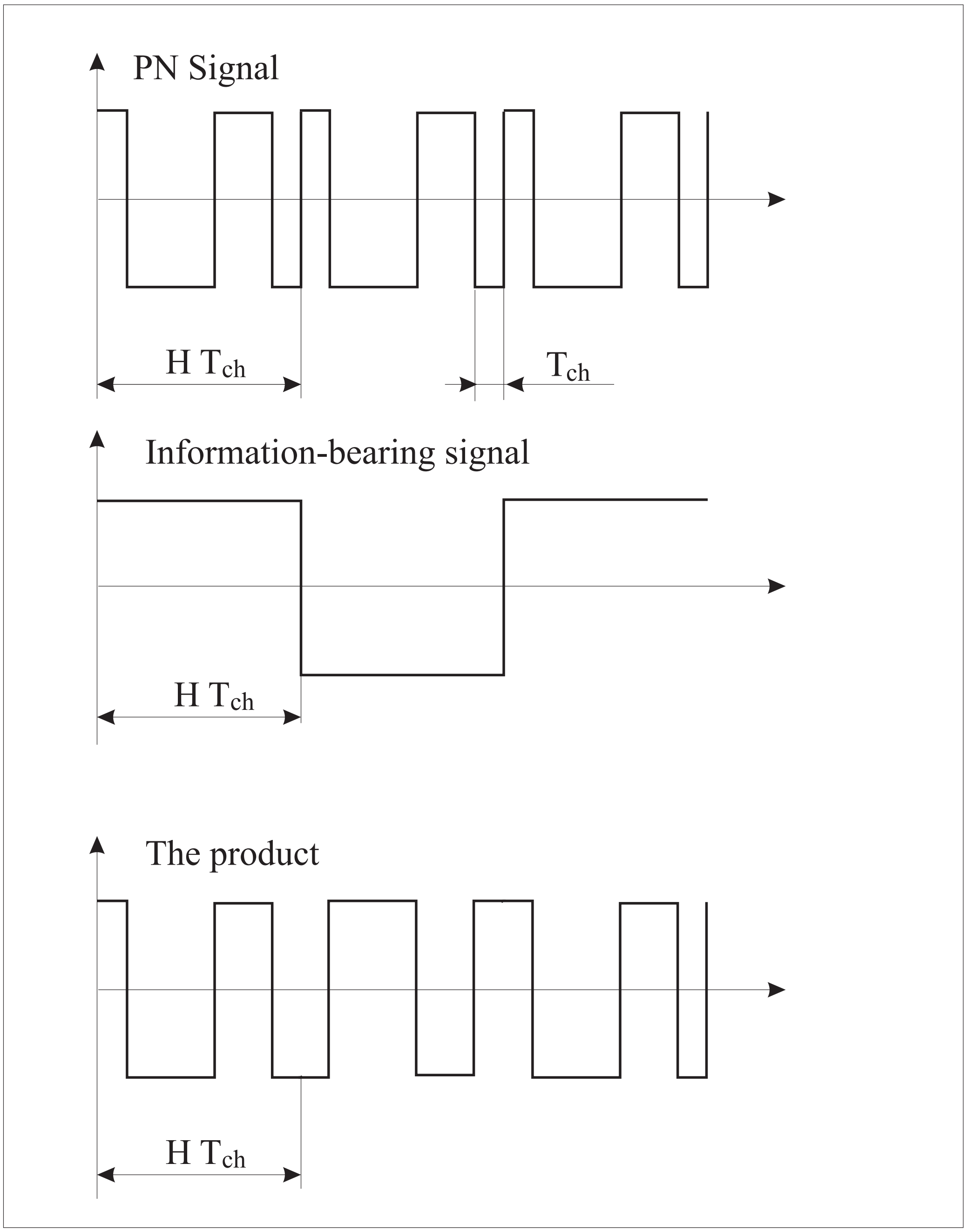

Figure 2.4: Generation of DS spread spectrum signal.

detectability signal transmission. By using an auxiliary pseudo-noise (PN) process that "looks" random to an "outsider", a weak low-powered message is hidden in the background noise. The message could be recovered by the intended receiver, however, anyone else who has no knowledge of the PN code sequence is unable to detect its presence in the signal. Hence, the signal has a low probability of being intercepted (LPI) and is called correspondingly.

It is not only the recovery of the message by an unauthorized receiver that could be prevented, the resulting signal is inherently robust to many channel impairments due to its antijamming capability. Indeed, as the name suggests, the spectrum of the original narrowband signal is spread over a much greater frequency range, thus resulting in a significantly wider band (*wideband* signal). In a DS spectrum spreading system, this is due to the use of the shorter symbols called "chips" in PN waveform $u(\tau)$:

$$u(\tau) = \sum_{h=1}^{H} c_h \eta(\tau - hT_{ch}).$$

Here $c_{1:H}$ is a spreading code sequence consisting of H chips (with values $\{\pm 1\}$) per symbol, $\eta(\tau - hT_{ch})$ is a rectangular pulse of unit height and duration T_{ch} transmitted at $(h-1)T_{ch} < \tau \leq hT_{ch}$ and T_{ch} is the chip interval satisfying the relation $T_{ch} = T_q/H$ (see Figure 2.4). The original narrowband message is multiplied by $u(\tau)$, resulting in a much faster transmitted signal with larger frequency components and a wider spectrum. However, multiplication by the same $u(\tau)$ is performed at the receiver end again, while the information-bearing signal is de-spreading (as illustrated in Figure 2.5). This, in turn, spreads the spectrum of any interference, hence, lowering its power density, and, therefore, allowing the design of the system to be more robust to channel impairments.

The same idea allows several (in our case K) simultaneous users to share the same channel bandwidth in a *code division multiple access* (CDMA) system [Verdu, 1998]. This is performed by providing each user with its own unique signature sequence (spreading code) $c_{k,1:H}$, resulting in a spreading

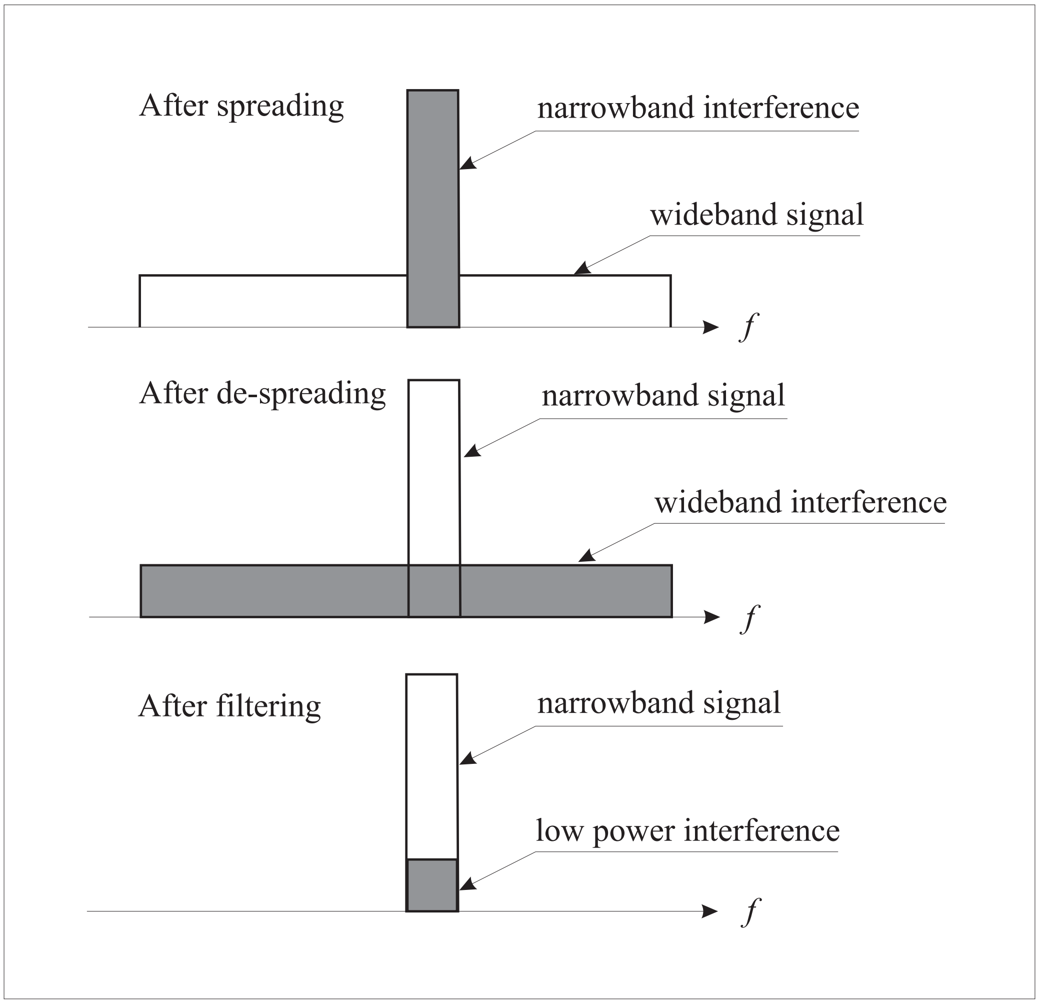

Figure 2.5: Antijamming capability of the spread spectrum signals.

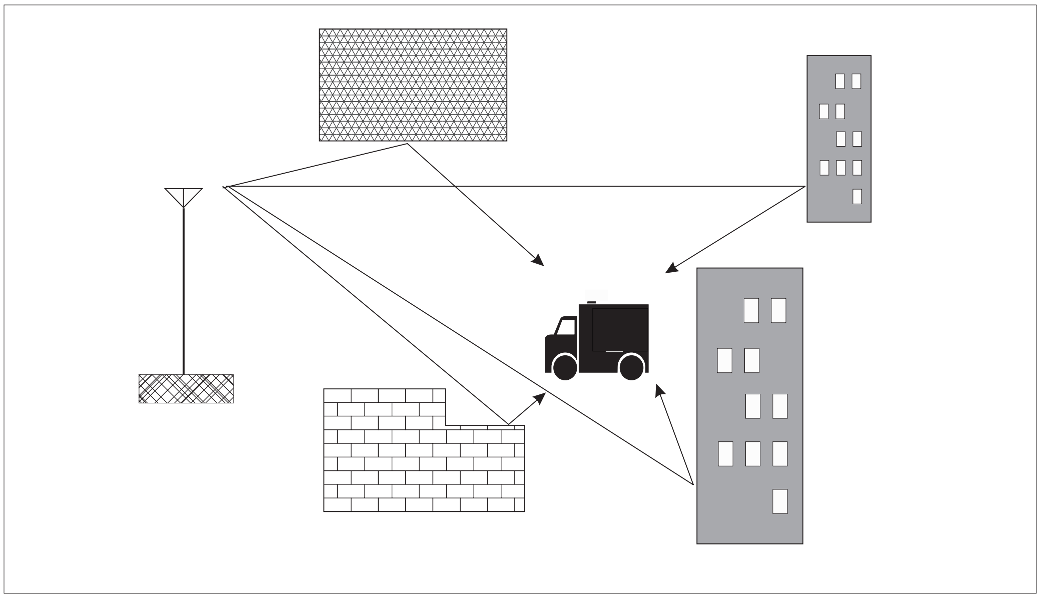

Figure 2.6: Multipath environment.

waveform for the kth user:

$$u_k(\tau) = \sum_{h=1}^{H} c_{k,h}\eta(\tau - hT_{ch}) \quad k = 1, \ldots, K.$$

In order to ensure low level of interference between the users and efficient separation of the signals, the sequences should have good cross-correlation properties (see [Viterbi, 1995] for the choice and generation of PN sequences).

If no spectrum spreading is employed in the system it is assumed that

$$u(\tau) = \eta(\tau - T_q),$$

with $T_q = T_{ch}$ and $H = 1$. This is not a practical choice for a pulse shaping filter but the following derivations could be extended to other pulse shapes.

2.1.5 Communication channel

The signal is passed from the transmitter to the receiver through a *communication channel*, which can be any physical medium used for the transmission of information, from a pair of wires that carry an electrical signal

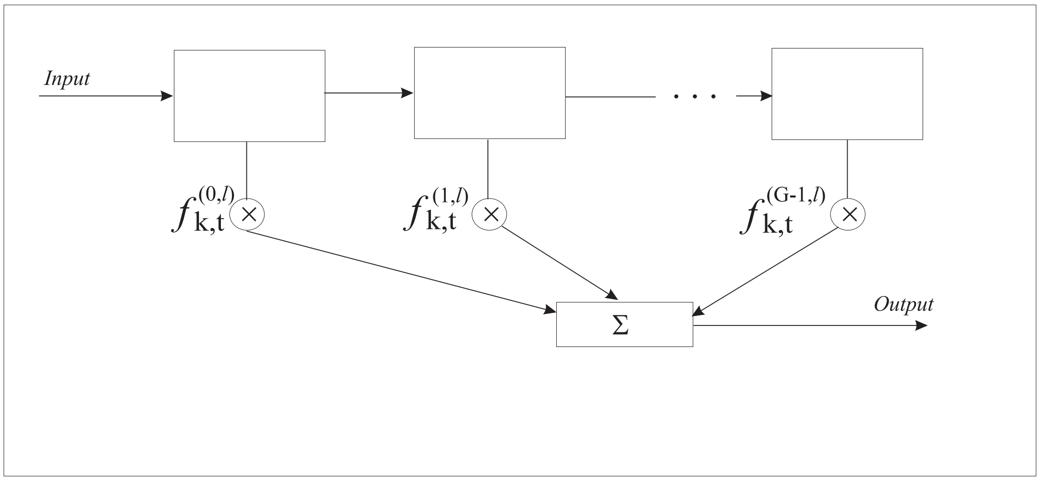

Figure 2.7: A tapped-delay-line model for multipath channel.

to free space over which the information is radiated by antennas. Many of them, of course, are not purpose-built for communications, and, therefore, are characterized by different kinds of amplitude and phase distortion as well as *additive* (or *thermal*) noise.

Furthermore, in some scenarios, several transmission paths from the transmitter to the receiver may be distinguishable due to various signal reflections (*multipath*), which, in turn, can add constructively or destructively resulting in signal *fading* (Figure 2.6). In addition, the different time delays of the individual component waves can cause interference between transmitted symbols (*intersymbol interference*).

From the modelling point of view, the channel is a description of this impact of the communication medium on the transmitted signal.

2.1.5.1 Channel response

This impact can be characterized by the impulse response of the channel which, due to the nature of multipath, is *complex* (causing both phase and amplitude distortions) and *time-variant*.

A *tapped-delay-line* (Figure 2.7) can be used for channel representation since it explicitly shows that the received signal is a superposition of the delayed and attenuated versions of the input signal. The taps are spaced T_s

seconds apart, and the tap weights indicate the contributions of the signal reflections arriving with different time delays. The equivalent discrete-time impulse response of the lth subchannel of the kth user can then be represented as follows:

$$h^{(l)}_{\text{channel } k,t} = \sum_{g=0}^{G-1} f^{(g,l)}_{k,t} \delta(t-g),$$

where t is a discrete time index, G is the number of paths of the channel, $f^{(g,l)}_{k,t}$ are the complex-valued time-varying multipath coefficients for the kth user and lth antenna arranged into vector $\mathbf{f}^{(l)}_{k,t}$, and $\delta(t-g)$ stands for Dirac function. The channel length G is considered to be the same for all subchannels $l = 1, \ldots, L$.

In general, it is impossible to provide a generic description of the channel characteristics in a real-life scenario – all kinds of unpredictable changes constantly alter the signal propagation conditions and, as a result, multipath coefficients (see Figure 2.8 for an example of the response of the multipath channel to a very narrow pulse, $\theta_{i,j}$ denotes the delay of the jth path at time t_i). The transmission channel, therefore, has to be described statistically.

2.1.5.2 A statistical description of the transmission channel

Over the years, considerable research has been carried out into the modelling and classification of communication channels which is reflected by the large number of publications in the area (a detailed overview on the subject is given in [Biglieri *et al.*, 1998]). Due to the rapid evolution of mobile systems, a significant amount of this research was directed into characterization of multipath channels typical for wireless systems. Several statistical models based on multiple experiments and analyses of the physics of the channels have been proposed and validated.

One of the widely acceptable scenarios, which is under consideration in this thesis, is *Rayleigh fading*. It is based on the assumption of a large number of scatters in the channel, which after the application of the central limit theorem results in a Gaussian process model, leading to the Rayleigh distribution of the channel response envelope for the zero-mean channel [Jakes, 1974]. A zero-mean channel assumption particularly applies when the re-

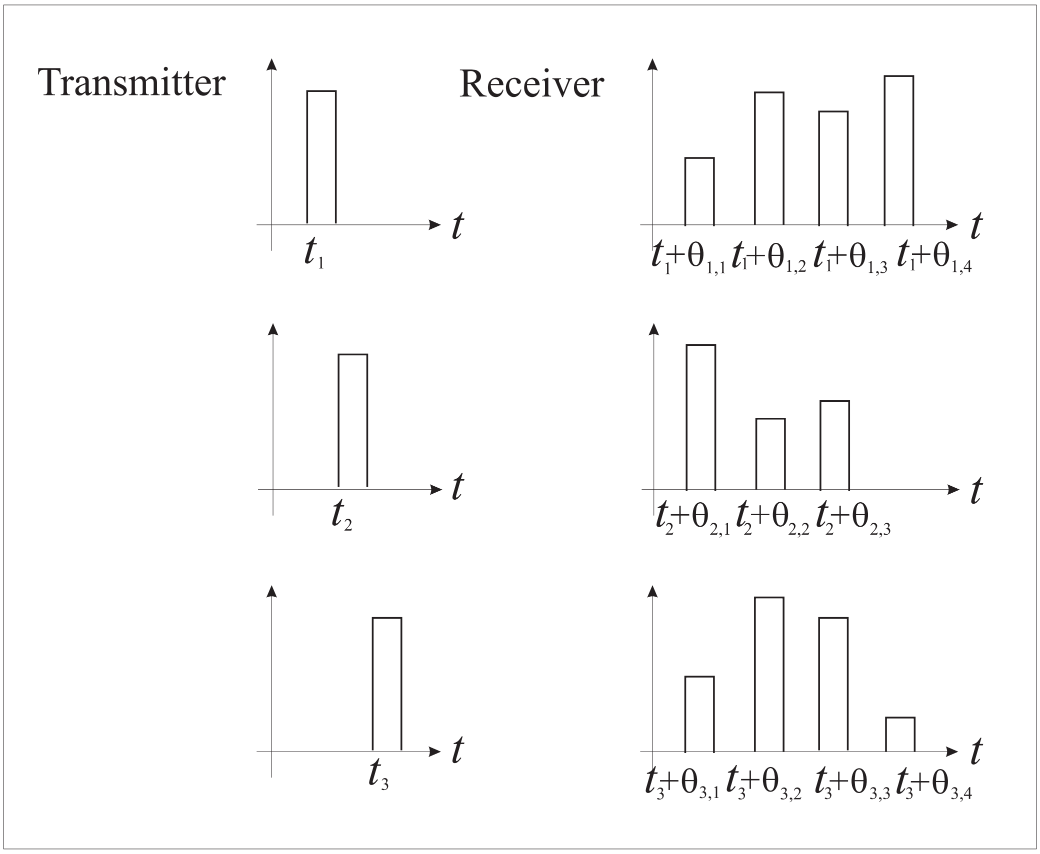

Figure 2.8: Multipath channel response to a very narrow pulse.

ceiver and the transmitter have no *line-of-sight* connection because of the lack of dominance of any particular reflected wave. This is typical for urban environment, where a large number of buildings and trees obstruct the transmitted signal.

When there is a predominant line-of-sight between the transmitter and the receiver the above no longer holds since the mean value of (at least) one component is non-zero due to a strong wave (in addition to the Rayleigh fading), and the channel is described as *Rician* (with the channel response envelop distributed according to Rice distribution).

An alternative probability function, which provides more flexibility and accuracy in matching the observed signal statistics and includes both Rayleigh and Rician models as a special case, is the Nakagami-m distribution.

2.1.5.3 Fading types

All statistical models presented in the previous subsection describe so-called small-scale fading. Large-scale fading refers to the path loss caused by motion over large areas due to considerably large physical objects (like hills and forests), and is described in terms of a mean loss and variations around it. Large-scale fading is reflected only on the strength of the received signal, and will not be considered in this thesis (see [Rappaport, 1996] for more details).

We will look, however, at different types of small-scale fading, both in terms of the signal dispersion and time variance of the channel (see Figure 2.9).

Due to the *signal dispersion* a transmitted digital pulse is spread in time, and if this maximum spread T_m is greater than the duration of the symbol, it overlaps with the other pulses transmitted at adjacent times (a phenomenon known as intersymbol interference), and *frequency selective* fading occurs. This happens when the *coherence bandwidth* of the channel $\mathrm{f_{co}}$ (a range of frequencies over which the channel passes all spectral components with approximately equal gain and linear phase, $\mathrm{f_{co}} \propto \frac{1}{T_m}$) is smaller than the signal bandwidth W_s: $T_s < T_m$, i.e. $\mathrm{f_{co}} < \frac{1}{T_s} \approx W_s$, and, as a result, the spectral components of the signal are affected in a different way by the channel.

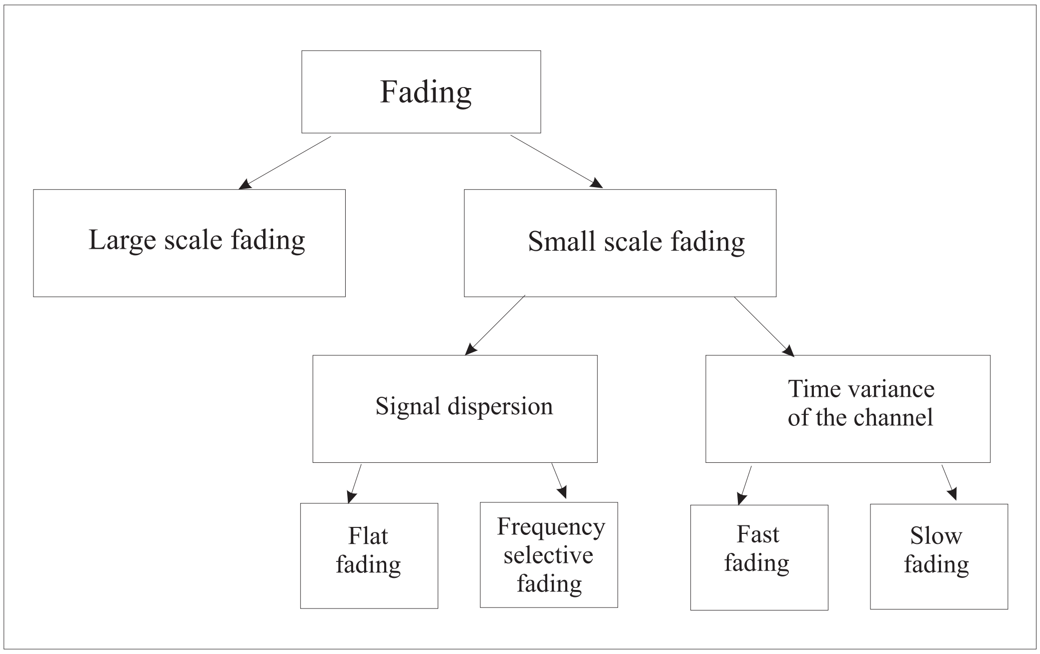

Figure 2.9: Fading types.

If this does not occur, the fading is called *flat,* meaning that the channel characteristics are approximately flat for all frequencies; such a channel has actually only one path, $G = 1$.

The time variability of the channel depends on its *coherence time* of the channel T_{co} (an expected time during which the channels response is essentially invariant); if it is smaller than the symbol duration T_s, the fading is considered to be *fast.* This occurs when the signal bandwidth W_s is less than the maximum Doppler frequency shift $\mathrm{f_d}$ of the signals in the mobile environment, where $\mathrm{f_d}$ depends on the relative velocity between the transmitter and the receiver ν, and the wavelength of the transmitted signal λ: $\mathrm{f_d} = \nu/\lambda$. Otherwise, the *slow* fading conditions are exhibited.

A common way to quantify this is to refer to the normalized Doppler frequency (or *Doppler rate*) defined as the product of the maximum Doppler frequency shift and the transmission symbol period $\mathrm{f_d}T_s$. The higher Doppler rates lead to the faster varying channels.

All types of fading described above will be considered in this thesis.

2.1.5.4 Channel modelling

As it was already mentioned, the channel is completely described by its impulse response, which is usually regarded as a stochastic process due to many random effects that occur in nature.

For many physical channels, including mobile communication channels, the statistics of the fading can be considered approximately stationary over long enough time intervals, and as such can be regarded as *wide-sense stationary* (WSS) [Bello, 1963]. The time-evolution of such processes can be effectively modelled by a finite-order, complex, auto-regressive moving average (ARMA) model.

For the flat Rayleigh fading channel with just one path $G = 1$, it will take the following form [Bello, 1963]:

$$f_{k,t}^{(l)} = \mathbf{a}_k^{(l)\mathrm{T}} f_{k,t-1:t-\rho_{k,1}^{(l)}}^{(l)} + \mathbf{b}_k^{(l)\mathrm{T}} v_{k,t:t-\rho_{k,2}^{(l)}}^{(l)}, \tag{2.3}$$

where $\mathbf{a}_k^{(l)}$ and $\mathbf{b}_k^{(l)}$ are the vectors of length $\rho_{k,1}^{(l)}$ and $\rho_{k,2}^{(l)}$ containing ARMA coefficients, $f_{k,t}^{(l)}$ is a complex channel disturbance for the lth subchannel kth user, and $v_{k,t}^{(l)}$ is complex Gaussian noise with independent and identically distributed (i.i.d.) zero-mean real and imaginary components of variance 1, $v_{k,t}^{(l)} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$.

This model, however, is difficult to use in a multipath environment due to its high complexity for the number of taps $G > 1$. Therefore, for the frequency-selective fading scenario, we adopt a simple first order autoregressive (AR) model ($\rho_{k,1}^{()} = 1, \rho_{k,2}^{(l)} = 1$), which, nevertheless [Iltis, 1990; Komninakis *et al.*, 1999] captures most of the channel tap dynamics:

$$f_{k,t}^{(g,l)} = a_k^{(g,l)} f_{k,t-1}^{(g,l)} + b_k^{(g,l)} v_t^{(g,l)}, \quad g = 0, \ldots G-1.$$

In this thesis, the coefficients of the model are assumed known, and their choice is discussed in Chapters 4 and 7.

2.1.6 Diversity receiver

A major problem with the fading channels is when the signal has a large magnitude fade and becomes unreadable (*deep fade*). In this case, unless another replica of the same transmitted signal are available, it might be impossible to recover transmitted symbols, and they might be lost.

To overcome this problem, a receiver with multiple antennas is employed in the system. If the antennas are spaced sufficiently far apart (more than half of the wavelength), the transmitted signals have different propagation paths and their fading statistics can be considered independent. Since deep fades rarely occur simultaneously during the same time intervals on two or more paths, the receiver level could be considerably improved.

We assume that a diversity receiver with L antennas is employed. The complex output of the lth subchannel $y_t^{(l)}$ at instant t can, thus, be expressed as

$$\begin{aligned} y_t^{(l)} = \textstyle\sum_{k=1}^{K} \sum_{g=0}^{G-1} f_{k,t}^{(g,l)} s_{\text{trans, } k} \left((t-g) \, T_s - \theta_{k,t} \right) + \varepsilon_t^{(l)}, \\ l = 1, \ldots, L, \end{aligned} \tag{2.4}$$

where the channel length G is considered to be the same for each subchannel, $\theta_{k,t}$ represents the unknown propagation delay for the kth user transmitted waveform $s_{\text{trans, } k}(\tau)$ and $\varepsilon_t^{(l)}$ is complex zero-mean additive noise at each subchannel. We assume that the delay corresponding to a user is the same for all receiver antennas.

The output $y_t^{(l)}$ is sampled at the Nyquist rate T_s^{-1}, with $T_s = T_{ch}/2$ due to the PN bandwidth being approximately equal to $1/T_{ch}$, resulting in two samples per chip. This means that $2H$ samples ($t = 2H\,(q-1)+1, \ldots, 2Hq$) correspond to each ($q$th) 2^κ-ary symbol obtained after coding, i.e.

$$r_{k,q} \leftrightarrow y^{(l)}_{2H(q-1)+1:2Hq}, \quad l = 1, \ldots, L, \quad k = 1, \ldots, K,$$

and $2HR$ samples ($t = 2HR\,(n-1)+1, \ldots, 2HRn$) correspond to the original nth symbol transmitted

$$d_{k,n} \leftrightarrow y^{(l)}_{2HR(n-1)+1:2HRn} \quad l = 1, \ldots, L, \quad k = 1, \ldots, K.$$

For channels with no delay $\theta_{k,t} = 0$, $k = 1, \ldots K$, the transmitted information sequence for each user should be oversampled correspondingly, and a *discrete time equivalent* model [Forney, 1972; Hoeher, 1992; Dai & Schwedyk, 1994; Proakis, 1995] can, thus, be obtained (see Chapter 5 for more details).

2.1.7 Propagation delay

Following [Iltis, 1990], we assume that the code delay also propagates according to a first-order AR model

$$\theta_{k,t} = \gamma_k \theta_{k,t-1} + \sigma_{k,\theta} \vartheta_{k,t}, \ \vartheta_{k,t} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1),$$

with the coefficients γ and $\sigma_{k,\theta}$ adjusted to account for transmitter and receiver timing jitter (see [Iltis, 1990] for details). In the above expression, $\vartheta_{k,t}$ is a complex white Gaussian process with i.i.d. real and imaginary components of variance 1, $\vartheta_{k,t} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$.

2.1.8 Non-Gaussian additive noise

As mentioned above, the signal at each receive antenna is corrupted by complex zero-mean additive noise $\varepsilon_t^{(l)}$. We assume that real and imaginary parts of $\varepsilon_t^{(l)}$ are mutually independent i.i.d. random sequences distributed as a mixture of zero-mean[1] Gaussians with a known number of components Z, the same for all subchannels $l = 1, \ldots, L$. This assumption allows us to model non-Gaussian noise and, in particular, outliers.

In order to identify the parameters (the variance in our case) of the distribution from which the noise samples are drawn, it is convenient to introduce a latent allocation variable $z_t^{(l)}$,

$$z_t^{(l)} \in \mathcal{Z} = \{1, 2, \ldots, Z\}, \quad t = 1, 2, \ldots,$$

such that

$$\Pr(z_t^{(l)} = \xi^{(l)}) = \lambda_{\xi^{(l)}}, \text{ for } \xi^{(l)} = 1, \ldots, Z, \quad l = 1, \ldots, L,$$

[1] The extension to non-zero mean Gaussian components is straightforward.

$$\sum_{\xi^{(l)}=1}^{Z} \lambda_{\xi^{(l)}} = 1, \quad l = 1, \ldots, L.$$

Given the values of $z_t^{(l)}$, the additive noise component at time t is drawn from the complex Gaussian distribution with standard deviation $\sigma_{z_t}^{(l)}$ correspondingly. Thus, one obtains

$$\varepsilon_t^{(l)} = \sigma_{z_t}^{(l)} \epsilon_t^{(l)}, \quad \epsilon_t^{(l)} \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0, 1\right), \tag{2.5}$$

2.2 Digital communications problems

In this thesis we assume that the symbols $d_{k,n}$, the channel characteristics $\mathbf{f}_{k,t}^{(l)}$ and the code delays $\theta_{k,t}$ are unknown for $k = 1, \ldots, K,\ \ l = 1, \ldots, L$, $n > 0$ and $t > 0$; and our aim is to obtain sequentially in time the estimates of these parameters based on the currently available data $y_{1:2HRn}$ – the process known as *filtering*.

2.2.1 Selected digital communications applications

Formulated in this general form, the problem has many particular cases that are of great interest in digital communications. A diagram presented in Figure 2.10 indicates some of them, mainly, the ones that this dissertation concerns. However, the results, and, indeed, the diagram itself, can be significantly extended.

2.2.2 Classification of digital communications problems

From a methodological point of view, these cases fall into two general classes (Figure 2.11). In the *first* one, the unknown state of the model – typically the transmitted symbols – takes its values in a finite set and one is concerned with exponentially increasing number of terms in the summations; this includes, for example, demodulation in fading channels, or multiuser detection in synchronous CDMA. This type of models is also known as jump Markov linear systems, i.e. linear Gaussian state space models switching ac-

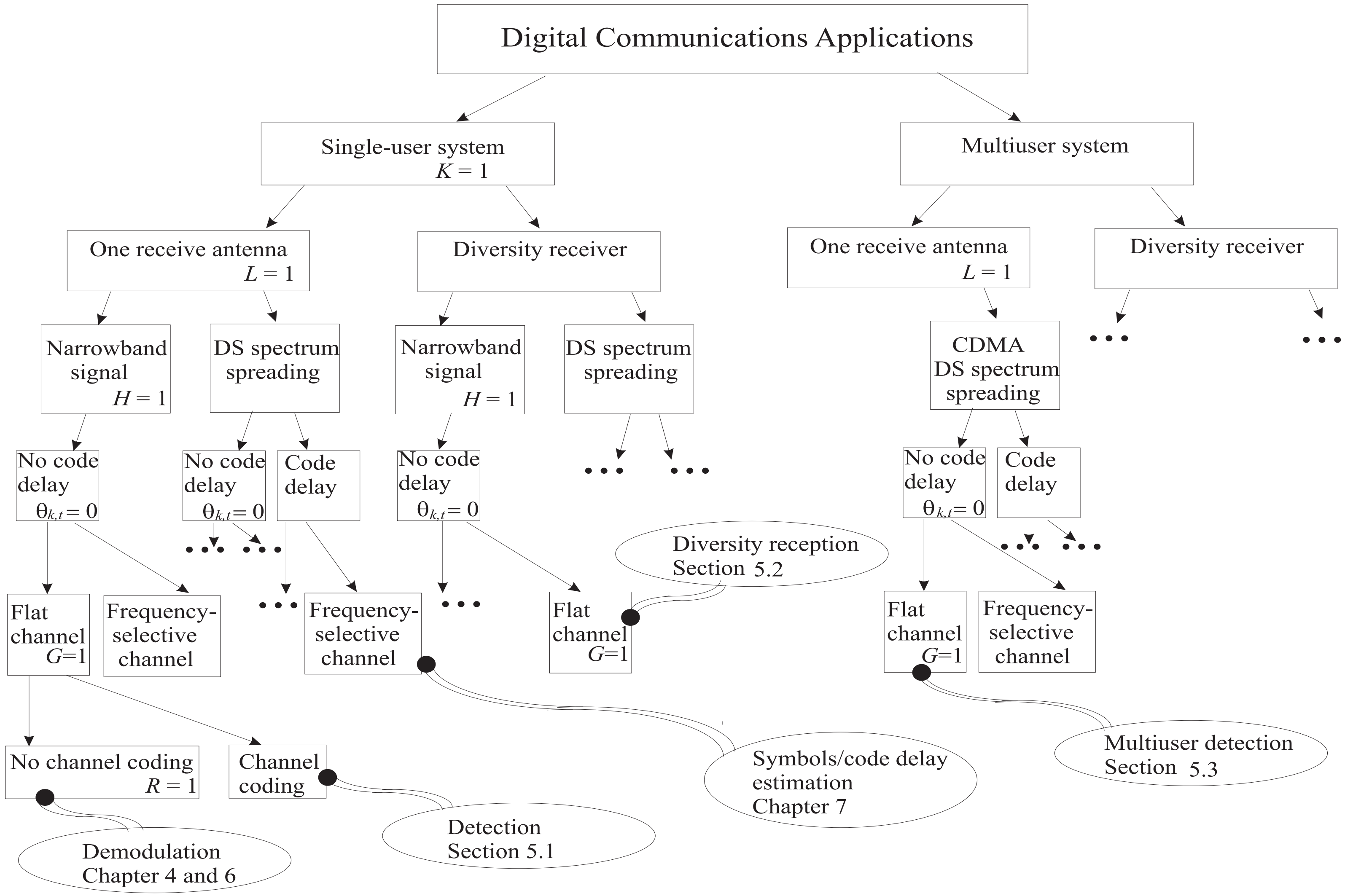

Figure 2.10: Selected digital communications applications.

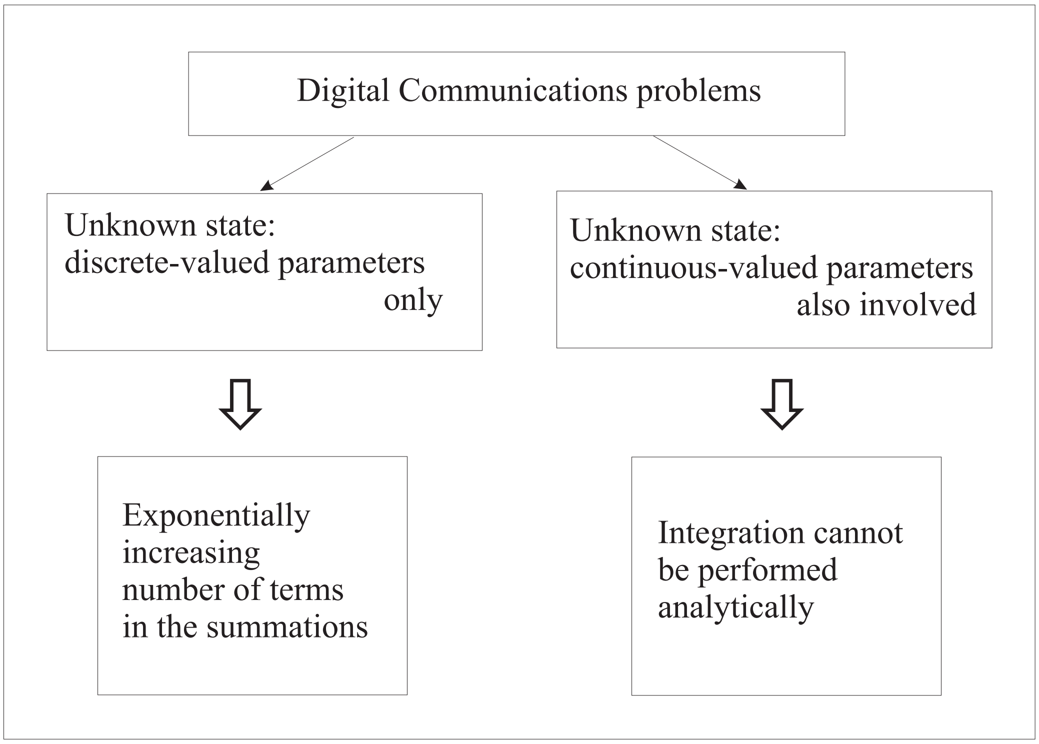

Figure 2.11: Classification of digital communications problems.

cording to an unobserved finite state space Markov chain. In the *second* class, one faces a more challenging problem where the unknown state of interest consists not only of the symbols but also some continuous-valued parameters, such as the code delays as in DS spread-spectrum system analyses, and the integration required to obtain the estimates of interest cannot be performed analytically. A development of efficient filtering algorithms for both these classes is of interest in this thesis.

Even in its simplest form (demodulation in flat fading channels), the problem still presents a great challenge for researchers, and requires the use of an accurate filtering technique capable of coping with the difficult transmission conditions described above. We will now review the non-linear non-Gaussian filters previously proposed in the literature with particular emphasis on *particle filters* – the technique which has recently gained much interest in the literature and proved to be useful in multiple complicated scenarios. This approach shows a great promise for our purposes and is investigated in this dissertation.

3 Bayesian Filtering

In many engineering applications, one needs to extract the signal from the data corrupted by additive random noise and interferences of different kinds in order to recover the unknown quantities of interest. The data often arrive sequentially in time and, therefore, require on-line decision-making responses. Except for several special cases restricted to a narrow linear Gaussian class of models, this problem still presents a major challenge to researchers. A number of approximate filters have been devised for more complicated scenarios. However, until recently, there existed no general methodology whenever non-linearity or non-Gaussianity was involved. The *Sequential Monte Carlo* (SMC) also known as *particle filtering* methods, recently emerged in the field of statistics and engineering, are believed to be such a key leading technology, and are the subject of this chapter.

We begin with the mathematical formulation of the general model studied in this dissertation and the introduction of the Bayesian approach to filtering. Several filters previously proposed in the literature are then reviewed, and, finally, basic ideas of particle filtering techniques and some strategies for their efficient implementation are discussed.

3.1 Problem statement

3.1.1 State-space model

For the sake of clarity, we restrict ourselves here to the following state-space model, which is of interest in this thesis[2]. The unobserved $n_{\mathbf{x}}$ - dimensional state of interest, $\{\mathbf{x}_n; n = 0, 1, 2, \ldots\}$, changes over time according to the *system equation* of the following form:

$$\mathbf{x}_n = \Psi\left(\mathbf{x}_{n-1}, \mathbf{v}_n\right), \tag{3.1}$$

and the $n_{\mathbf{y}}$-dimensional measurements $\mathbf{y}_n$, taken at discrete time points $n = 1, 2, \ldots$, are given by

$$\mathbf{y}_n = \Omega\left(\mathbf{x}_n, \boldsymbol{\varepsilon}_n\right), \tag{3.2}$$

where $\Psi(\cdot)$ is a system transition function, $\Omega(\cdot)$ is a measurement function, and $\mathbf{v}_n$ and $\boldsymbol{\varepsilon}_n$ are independent noise vectors with a known distribution. For simplicity, the same notation is used throughout for both random variables and their realization, and we also assume that the analytical forms of the functions $\Psi(\cdot), \Omega(\cdot)$ and the initial probability density of the state $p(\mathbf{x}_0)$ are known.

3.1.2 Bayesian inference

In this thesis, we follow the Bayesian approach, which provides an elegant and consistent method of dealing with uncertainty. The Bayesian posterior, $p\left(\mathbf{x}_{0:n} \middle| \mathbf{y}_{1:n}\right)$, reflects all the information we have about the state of the system $\mathbf{x}_{0:n}$, contained in the measurements $\mathbf{y}_{1:n}$ and the prior $p\left(\mathbf{x}_{0:n}\right)$, and gives a direct and easily applicable means of combining the two last-mentioned densities (*Bayes' theorem*)[3]:

$$p\left(\mathbf{x}_{0:n} \middle| \mathbf{y}_{1:n}\right) = \frac{p\left(\mathbf{y}_{1:n} \middle| \mathbf{x}_{0:n}\right) p\left(\mathbf{x}_{0:n}\right)}{p\left(\mathbf{y}_{1:n}\right)}. \tag{3.3}$$

[2] SMC can be applied in a more general setting.

[3] Recall that the sequences $\mathbf{x}_{0:n}$ and $\mathbf{y}_{0:n}$ denote respectively the signal and the observations up to time n: $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$ and $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$.

Taking into account that the observations up to time n are independent given $\mathbf{x}_{0:n}$, the likelihood $p\left(\mathbf{y}_{1:n} \middle| \mathbf{x}_{0:n}\right)$ in the above equation can be factorized as follows:

$$p\left(\mathbf{y}_{1:n} \middle| \mathbf{x}_{0:n}\right) = \prod_{i=1}^{n} p\left(\mathbf{y}_{i} \middle| \mathbf{x}_{0:n}\right), \tag{3.4}$$

and, since, conditional on $\mathbf{x}_i$, the measurement $\mathbf{y}_i$ is independent of the states at all other times, it is given by:

$$p\left(\mathbf{y}_{1:n} \middle| \mathbf{x}_{0:n}\right) = \prod_{i=1}^{n} p\left(\mathbf{y}_{i} \middle| \mathbf{x}_{i}\right). \tag{3.5}$$

In addition, as a result of the Markov structure of the system Equation (3.1), the prior $p\left(\mathbf{x}_{0:n}\right)$ takes the following form:

$$p\left(\mathbf{x}_{0:n}\right) = p\left(\mathbf{x}_{0}\right) \prod_{i=1}^{n} p\left(\mathbf{x}_{i} \middle| \mathbf{x}_{i-1}\right), \tag{3.6}$$

resulting in the posterior probability density being equal to

$$p\left(\mathbf{x}_{0:n} \middle| \mathbf{y}_{1:n}\right) = \frac{p\left(\mathbf{x}_{0}\right) \prod_{i=1}^{n} p\left(\mathbf{y}_{i} \middle| \mathbf{x}_{i}\right) p\left(\mathbf{x}_{i} \middle| \mathbf{x}_{i-1}\right)}{p\left(\mathbf{y}_{1:n}\right)}. \tag{3.7}$$

3.1.3 Filtering objectives

Our objective is to obtain the estimates of the state at time n, conditional upon the measurements up to time $\breve{n}$, such as, for example, *Minimum Mean Square Estimate* (MMSE) of $\mathbf{x}_n$:

$$\hat{\mathbf{x}}_{n}^{MMSE} = \mathbb{E}_{p\left(\mathbf{x}_{n} | \mathbf{y}_{1:\breve{n}}\right)}[\mathbf{x}_{n}] = \int \mathbf{x}_{n} p\left(\mathbf{x}_{n} \middle| \mathbf{y}_{1:\breve{n}}\right) d\mathbf{x}_{n}, \tag{3.8}$$

or *Marginal Maximum A Posteriori* (MMAP) given by:

$$\hat{\mathbf{x}}_{n}^{MMAP} = \arg\max_{\mathbf{x}_{n}} p\left(\mathbf{x}_{n} \middle| \mathbf{y}_{1:\breve{n}}\right). \tag{3.9}$$

These can obviously be extended to estimating the functions of the state instead of the state itself:

$$\widehat{\Upsilon(\mathbf{x}_n)}^{MMSE} = \mathbb{E}_{p(\mathbf{x}_n|\mathbf{y}_{1:\breve{n}})}\left[\Upsilon(\mathbf{x}_n)\right] = \int \Upsilon(\mathbf{x}_n)\, p(\mathbf{x}_n|\mathbf{y}_{1:\breve{n}})\, d\mathbf{x}_n.$$

Calculating $p(\mathbf{x}_n|\mathbf{y}_{1:\breve{n}})$ for $\breve{n} = n$, and hence the estimates of $\mathbf{x}_n$, given the data up to time n, is the aim of Bayesian *filtering* and a primary subject of this dissertation. One might also be interested in the related problems of evaluating $p(\mathbf{x}_{0:n}|\mathbf{y}_{1:\breve{n}})$ for $\breve{n} > n$ (*smoothing*) or for $\breve{n} < n$ *(prediction)*.

3.1.4 Sequential scheme

The probability density of interest $p(\mathbf{x}_n|\mathbf{y}_{1:n})$ can be obtained by marginalization of (3.7), however, the dimension of the integration in this case grows as n increases. This can be avoided by using a sequential scheme.

A recursive formula for the joint probability density can be obtained straightforwardly from Equation (3.7):

$$p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n}) = p(\mathbf{x}_{0:n-1}|\mathbf{y}_{1:n-1}) \frac{p(\mathbf{y}_n|\mathbf{x}_n)\, p(\mathbf{x}_n|\mathbf{x}_{n-1})}{p(\mathbf{y}_n|\mathbf{y}_{1:n-1})}, \tag{3.10}$$

with the marginal $p(\mathbf{x}_n|\mathbf{y}_{1:n})$ also satisfying the recursion [Sorenson, 1988]:

$$p(\mathbf{x}_n|\mathbf{y}_{1:n-1}) = \int p(\mathbf{x}_n|\mathbf{x}_{n-1})\, p(\mathbf{x}_{n-1}|\mathbf{y}_{1:n-1})\, d\mathbf{x}_{n-1}, \tag{3.11}$$

$$p(\mathbf{x}_n|\mathbf{y}_{1:n}) = \frac{p(\mathbf{y}_n|\mathbf{x}_n)\, p(\mathbf{x}_n|\mathbf{y}_{1:n-1})}{p(\mathbf{y}_n|\mathbf{y}_{1:n-1})}, \tag{3.12}$$

where

$$p(\mathbf{y}_n|\mathbf{y}_{1:n-1}) = \int p(\mathbf{y}_n|\mathbf{x}_n)\, p(\mathbf{x}_n|\mathbf{y}_{1:n-1})\, d\mathbf{x}_n. \tag{3.13}$$

Equations (3.11) and (3.12) are called respectively *prediction* and *updating.*

The above expressions are deceptively simple, however, since the integrations involved are usually intractable. One cannot typically compute the normalizing constant $p(\mathbf{y}_{1:n})$ and the marginals of $p(\mathbf{x}_n|\mathbf{y}_{1:n})$, particularly, $p(\mathbf{x}_n|\mathbf{y}_n)$, except for several special cases when the integration can be performed exactly. The problem is of great importance, though, which is why a great number of different approaches and filters have been proposed, and in the next subsection we will briefly review some of them.

3.2 Standard approach to filtering

The major breakthrough in the filter theory[4] was due to Kalman and Bucy [Kalman & Bucy, 1961], who noticed that Equations (3.11-3.13) could be solved to produce the Kalman filter for a linear Gaussian class of problems. The Kalman filter was then extended (EKF) to consider more general non-linear non-Gaussian scenario. Later, with the increase in computer power, more computationally expensive filters were introduced approximating the posterior of interest by mixture distributions, with the Gaussian sum filter and Interacting Multiple Model algorithm among others. Finally, the grid-based methods evaluating the required density as a set of nodes covering the state space appeared. All these methods are briefly described in this section, where we also mention some problems associated with their use.

3.2.1 Kalman filter

If the state space model is linear, with uncorrelated system and observation Gaussian noise and a Gaussian prior, i.e. Equations (3.1-3.2) are of the following form:

$$\begin{aligned} &\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{B}\mathbf{v}_n, \quad \mathbf{v}_n \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_\mathbf{v}}\right), \\ &\mathbf{y}_n = \mathbf{C}\mathbf{x}_n + \mathbf{D}\boldsymbol{\varepsilon}_n, \qquad \boldsymbol{\varepsilon}_n \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_{\boldsymbol{\varepsilon}}}\right), \\ &\text{with} \quad \mathbf{x}_0 \sim \mathcal{N}\left(\hat{\mathbf{x}}_0, \boldsymbol{\Sigma}_0\right), \end{aligned} \tag{3.14}$$

the probability densities $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n-1}\right)$ and $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n}\right)$, which can be regarded as the *prior* and *posterior* at time n, are themselves Gaussian random variables. Here, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are, correspondingly, $n_\mathbf{x} \times n_\mathbf{x}$, $n_\mathbf{x} \times n_\mathbf{v}$, $n_\mathbf{y} \times n_\mathbf{x}$, and $n_\mathbf{y} \times n_{\boldsymbol{\varepsilon}}$,matrices[5]. A probability distribution of a Gaussian random vari-

[4] for the state-space models

[5] Recall also the assumptions specified in Subsection 3.1.1.

able is uniquely defined by its mean vector and covariance matrix, which, in turn, can be updated according to Equation (3.11-3.13).

If one sets $\mathbf{x}_{0|0} = \hat{\mathbf{x}}_0$ and $\Sigma_{0|0} = \mathbf{\Sigma}_0$, and denotes the mean and covariance of the prior Gaussian $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n-1}\right)$ at time n by $\mathbf{x}_{n|n-1}$ and $\mathbf{\Sigma}_{n|n-1}$, and the mean and covariance of the posterior Gaussian $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n}\right)$ at time n by $\mathbf{x}_{n|n}$ and $\mathbf{\Sigma}_{n|n}$, then, for $n = 1, ..., T$, the Kalman filter proceeds as follows

$$\begin{aligned}
\mathbf{x}_{n|n-1} &= \mathbf{A}\mathbf{x}_{n-1|n-1}, \\
\mathbf{\Sigma}_{n|n-1} &= \mathbf{A}\mathbf{\Sigma}_{n-1|n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}}, \\
\\
\mathbf{x}_{n|n} &= \mathbf{x}_{n|n-1} + \mathbf{\Sigma}_{n|n-1}\mathbf{C}^{\mathrm{T}}\mathbf{K}_n^{-1}\left(\mathbf{y}_n - \mathbf{C}\mathbf{x}_{n|n-1}\right), \\
\mathbf{\Sigma}_{n|n} &= \mathbf{\Sigma}_{n|n-1} - \mathbf{\Sigma}_{n|n-1}\mathbf{C}^{\mathrm{T}}\mathbf{K}_n^{-1}\mathbf{C}\mathbf{\Sigma}_{n|n-1},
\end{aligned} \tag{3.15}$$

where the $n_{\mathbf{y}} \times n_{\mathbf{y}}$ matrix $\mathbf{K}_n$ is the Kalman gain:

$$\mathbf{K}_n = \mathbf{C}\mathbf{P}_{n|n-1}\mathbf{C}^{\mathrm{T}} + \mathbf{D}\mathbf{D}^{\mathrm{T}}. \tag{3.16}$$

These equations could be easily adapted to consider prediction and smoothing [Anderson & Moore, 1979]. For the other cases when Equations (3.11-3.13) could be solved see, for example, [Daum, 1988].

3.2.2 Extended Kalman Filter

Although the Kalman filter is extremely simple and optimal in the linear Gaussian scenario, new methods of filtering were required to consider more general case. Therefore, a number of approximate filters have been devised. Extended Kalman filter (EKF) is historically the first [Jazwinski, 1973], and, probably, most used one. The idea is to use a Taylor expansion to linearize the system and observation Equations (3.1-3.2), and, thus, approximate the system and noise distributions as Gaussians. The Kalman filter can then be used to obtain the solution for the resulting system.

Let us consider the following more general state space model :

$$\begin{aligned}
\mathbf{x}_n &= \Psi_1\left(\mathbf{x}_{n-1}\right) + \mathbf{B}\mathbf{v}_n, & \mathbf{v}_n &\overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_{\mathbf{v}}}\right), \\
\mathbf{y}_n &= \Omega_1\left(\mathbf{x}_n\right) + \mathbf{D}\boldsymbol{\varepsilon}_n, & \boldsymbol{\varepsilon}_n &\overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_{\varepsilon}}\right),
\end{aligned} \tag{3.17}$$

where $\mathbf{B}$ and $\mathbf{D}$ are the $n_{\mathbf{x}} \times n_{\mathbf{v}}$, and $n_{\mathbf{y}} \times n_{\boldsymbol{\varepsilon}}$ matrices.

Using a Taylor expansion about the current estimate of the mean of the state $\mathbf{x}_{n-1|n-1}$, one obtains

$$\mathbf{x}_n \approx \Psi_1(\mathbf{x}_{n-1|n-1}) + \underbrace{\left.\frac{\partial \Psi_1(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}_{n-1|n-1}}}_{\mathbf{A}_{n-1}} \left(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1}\right) + \mathbf{B}\mathbf{v}_n, \tag{3.18}$$

Similarly, the observation equation can be linearized about $\mathbf{x}_{n|n-1}$

$$\mathbf{y}_n \approx \Omega_1(\mathbf{x}_{n|n-1}) + \underbrace{\left.\frac{\partial \Omega_1(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}_{n|n-1}}}_{\mathbf{C}_n} \left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right) + \mathbf{D}\boldsymbol{\varepsilon}_n, \tag{3.19}$$

Applying the Kalman filter to these approximations (3.18-3.19) gives the EKF equations:

$$\begin{aligned}
\mathbf{x}_{n|n-1} &= \Psi_1\left(\mathbf{x}_{n-1|n-1}\right), \\
\boldsymbol{\Sigma}_{n|n-1} &= \mathbf{A}_{n-1}\boldsymbol{\Sigma}_{n-1|n-1}\mathbf{A}_{n-1}^{\mathsf{T}} + \mathbf{B}\mathbf{B}^{\mathsf{T}}, \\
\\
\mathbf{x}_{n|n} &= \mathbf{x}_{n|n-1} + \boldsymbol{\Sigma}_{n|n-1}\mathbf{C}_n^{\mathsf{T}}\mathbf{K}_n^{-1}\left(\mathbf{y}_n - \Omega_1\left(\mathbf{x}_{n|n-1}\right)\right), \\
\boldsymbol{\Sigma}_{n|n} &= \boldsymbol{\Sigma}_{n|n-1} - \boldsymbol{\Sigma}_{n|n-1}\mathbf{C}_n^{\mathsf{T}}\mathbf{K}_n^{-1}\mathbf{C}_n\boldsymbol{\Sigma}_{n|n-1},
\end{aligned} \tag{3.20}$$

where

$$\mathbf{K}_n = \mathbf{C}_n\mathbf{P}_{n|n-1}\mathbf{C}_n^{\mathsf{T}} + \mathbf{D}\mathbf{D}^{\mathsf{T}}. \tag{3.21}$$

The filter works quite well for a weakly non-linear system. For a systems with higher degree of non-linearity, the accuracy of the linearization could be increased, for example, by using Iterated EKF. The idea is to improve the estimates $\mathbf{x}_{n-1|n-1}$, $\mathbf{x}_{n|n-1}$ by smoothing, i.e. using the filter to estimate $\mathbf{x}_{n-1|n}$ as well as $\mathbf{x}_{n|n}$, and then re-linearizing the equations at time $n-1$ about the new estimate. This, in turn, will lead to a new $\mathbf{x}_{n-1|n-1}$, and a, hopefully, better $\mathbf{x}_{n|n}$ and $\mathbf{x}_{n-1|n}$. The algorithm could cycle repeatedly through the data until there are no significant changes in the estimates [Wishner *et al.*, 1968].

A different way of reducing the error in the linearization approximation is by intelligent choice of the co-ordinates: for instance, using polar co-ordinates instead of Cartesian as in [Collings & Moore, 1994, 1995]. Another suggestion is to include quadratic terms in the Taylor expansion, although, at the cost of increased computational load and complexity [Sorenson & Stubberud, 1968], see also [Wishner *et al.*, 1969] for the comparison of the EKF, iterated EKF and EKF with quadratic terms.

Divergence is a further problem with the EKF occurring due to model mis-specification. Since the model that the Kalman filter is solving is inaccurate, the EKF can often considerably underestimate the covariance of its estimate of the state, and, if the current estimate is poor, will struggle to improve the accuracy with time. To overcome this, the old data could be exponentially weighted [Fagin, 1964]. Alternatively, some extra uncertainty could be included in the system by increasing $\mathbf{\Sigma}_{n|n-1}$ [Fitzgerald, 1968; Wolf, 1968], or fixing it to a predetermined value obtained after processing a limited amount of data [Fitzgerald, 1968]. A more structured way to evaluate filter performance [Jazwinski, 1973] is to include some small and random residuals, and make the modifications to the noise input level according to their size: if the residuals become large, which indicates the divergence, the noise input increases.

3.2.3 Gaussian Sum filter

One way or another, the EKF and its variants always approximate the probability densities as Gaussian ones, which is itself a false assumption. With the increase in computer power, therefore, the efforts were directed into developing the filters estimating the densities as mixture distributions; and the Gaussian sum filter [Sorenson & Alspach, 1971; Alspach & Sorenson, 1972] is one of them.

The filter approximates the probability densities of all the non-Gaussian model parameters by mixture of Gaussians, and, in a way, is an extension of the Kalman filter or EKF allowing to cope with non-Gaussianity. The probability densities $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n-1}\right)$ and $p\left(\mathbf{x}_n|\,\mathbf{y}_{1:n}\right)$ are also approximated by a Gaussian mixture with N_n terms at each time step:

$$\hat{p}\left(\mathbf{x}_n|\,\mathbf{y}_{1:n-1}\right) = \sum_{i=1}^{N_n} w_{n|n-1}^{(i)} \mathcal{N}\left(\mathbf{x}_n; \mathbf{x}_{n|n-1}^{(i)}, \mathbf{\Sigma}_{n|n-1}^{(i)}\right), \tag{3.22}$$

$$\hat{p}\left(\mathbf{x}_n|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N_n} w_{n|n}^{(i)} \mathcal{N}\left(\mathbf{x}_n; \mathbf{x}_{n|n}^{(i)}, \mathbf{\Sigma}_{n|n}^{(i)}\right), \tag{3.23}$$

where $w_{n|n-1}^{(i)}$ and $w_{n|n}^{(i)}$ are the sets of weights, such that $\sum_{i=1}^{N_n} w_{n|n-1}^{(i)} = 1$ and $\sum_{i=1}^{N_n} w_{n|n}^{(i)} = 1$. The means $\mathbf{x}_{n|n}^{(i)}$ and covariances $\mathbf{\Sigma}_{n|n}^{(i)}$ for the ith term can be updated using the EKF, conditional on both, the term in the densities mixture, Equation (3.22), and the term in the noise mixture (when the noise as well is non-Gaussian). This means that at each time step, the number of means and covariances calculated is $N_n = N_{n-1}N_{\mathbf{v}}N_{\boldsymbol{\varepsilon}}$, where $N_{\mathbf{v}}$, $N_{\boldsymbol{\varepsilon}}$ are correspondingly the number of terms in the system and observation noise mixture approximation. This number grows exponentially with time, and should be reduced at each step either by removing the terms with low weight or by merging two "close" terms together. The details are given in [Sorenson & Alspach, 1971], where the update equation for the weights could also be found, and the issues with "best" Gaussian mixture approximations are discussed.

As such, the Gaussian sum filter still uses the EKF, and, therefore, suffers from its inaccuracies. The linearization error in this framework, however, might be substantially reduced by using larger number of terms in the mixture approximations.

3.2.4 Jump Markov Linear System Filters

The filters estimating the densities of interest by the mixtures of Gaussians are particularly suitable for the Jump Markov Linear Systems (JMLS), and a lot of work has been done in this direction.

The JMLS can be modelled as

$$\begin{aligned} \mathbf{x}_n &= \mathbf{A}(d_n)\mathbf{x}_{n-1} + \mathbf{B}(d_n)\mathbf{v}_n, \quad & \mathbf{v}_n &\overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_{\mathbf{v}}}\right), \\ \mathbf{y}_n &= \mathbf{C}(d_n)\mathbf{x}_n + \mathbf{D}(d_n)\boldsymbol{\varepsilon}_n, & \boldsymbol{\varepsilon}_n &\overset{i.i.d.}{\sim} \mathcal{N}\left(0, \mathbf{I}_{n_{\boldsymbol{\varepsilon}}}\right), \end{aligned} \tag{3.24}$$

where $\mathbf{v}_n$, and $\boldsymbol{\varepsilon}_n$ denote independent white Gaussian noise vectors, and d_n is an unobserved indicator variable, which takes its values in a discrete set $\mathcal{M}$

$$d_n \in \mathcal{M} = \{1, 2, \ldots, M\}, \quad n = 1, 2, \ldots, \tag{3.25}$$

and changes over time according to known transition probabilities:

$$p_{ij} = \Pr\{\, d_n = j |\, d_{n-1} = i\}\,, i, j \in \mathcal{M}. \tag{3.26}$$

For this model, if the value of $d_{1:n}$ was known[6], it would be possible to calculate the posterior $p(\mathbf{x}_n|\, \mathbf{y}_{1:n})$ straightforwardly in a sequential manner with the help of the Kalman filter. For the unknown $d_{1:n}$, however, the true posterior at time n is a mixture with M^n Gaussian terms, each corresponding to a specific value of $d_{1:n}$ and weighted accordingly:

$$p(\mathbf{x}_n|\, \mathbf{y}_{1:n-1}) = \sum_{d_{1:n} \in \mathcal{M}^n} p(\mathbf{x}_n|\, d_{1:n}, \mathbf{y}_{1:n})\, p(d_{1:n}|\, \mathbf{y}_{1:n})\,. \tag{3.27}$$

The weights $p(d_{1:n}|\, \mathbf{y}_{1:n})$ of the Gaussians $p(\mathbf{x}_n|\, d_{1:n}, \mathbf{y}_{1:n})$ could be estimated recursively similar to Equation (3.11-3.12):

$$p(d_{1:n}|\, \mathbf{y}_{1:n-1}) = p(d_n|\, d_{1:n-1})\, p(d_{1:n-1}|\, \mathbf{y}_{1:n-1})\,, \tag{3.28}$$

$$p(d_{1:n}|\, \mathbf{y}_{1:n}) \propto p(\mathbf{y}_n|\, d_{1:n})\, p(d_{1:n}|\, \mathbf{y}_{1:n-1})\,, \tag{3.29}$$

and should sum to 1, with $p(\mathbf{y}_n|\, d_{1:n})$ being the likelihood of the observations. The means and covariances of the Gaussians could be propagated using the Kalman filter.

Unfortunately, one faces the same problem of the exponentially increasing computational complexity of the filter, therefore, the number of terms in the mixture should be limited. A number of algorithms developed in order to solve this problem will be briefly reviewed now.

---

[6]Recall that the sequences $d_{1:n}$ denotes $d_{1:n} = \{d_1, \ldots, d_n\}$.

3.2.4.1 Detection-Estimation Algorithm

The Detection-Estimation Algorithm (DEA), first, proposed in [Tugnait & Haddad, 1979] employs the simplest approach. The exponentially growing computational cost of the filter is reduced by keeping at each stage only a certain predetermined number of terms with the highest weights in the mixture. The rest of the components are discarded. The number of Kalman filters required in this approach is equal to the number of terms in the mixture kept multiplied by the number of possible states M. This algorithm is studied in more details in Chapter 6.

3.2.4.2 Random Sampling Algorithm

The Random Sampling Algorithm (RSA) of [Akashi & Kumamoto, 1977] is based on an alternative approach. A sufficiently large number, say N, of possible Markov chain sequences $\{d_{1:n}^{(i)}\}_{i=1}^{N}$ is sampled, and, given these samples, an approximation of the filtering density function is obtained as a mixture of N corresponding Gaussians:

$$\hat{p}_{RSA}\left(\mathbf{x}_n|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}(d_{1:n}^{(i)}) p\left(\mathbf{x}_n|\, d_{1:n}^{(i)}, \mathbf{y}_{1:n}\right), \tag{3.30}$$

where $\tilde{w}(d_{1:n}^{(i)})$ denote normalized weights. The probability of any term being sampled is proportional to its weight in the mixture. Similar to the DEA, the computational load of this algorithm is proportional to MN.

3.2.4.3 Generalized Pseudo-Bayes Algorithm

The Generalised Pseudo-Bayes (GPB) algorithm assumes that the filtering probability density, conditioned on l_{GPB} last values of the indicator variable $d_{n-l_{\mathrm{GPB}}+1:n}$, is Gaussian:

$$\hat{p}_{\mathrm{GPB}}\left(\mathbf{x}_n|\, d_{n-l_{\mathrm{GPB}}+1:n}, \mathbf{y}_{1:n}\right) \sim \mathcal{N}(\hat{\mathbf{x}}_{n|n}, \hat{\mathbf{\Sigma}}_{n|n}), \tag{3.31}$$

with the mean $\hat{\mathbf{x}}_{n|n}$ and covariance $\hat{\boldsymbol{\Sigma}}_{n|n}$ of the Gaussian sum it is approximating:

$$\hat{\mathbf{x}}_{n|n} = \sum_{i=1}^{M^{l_{\mathrm{GPB}}}} \hat{\mathbf{x}}_{n|n}^{(i)} p\left(d_{n-l_{\mathrm{GPB}}+1:n}^{(i)} \middle| \mathbf{y}_{1:n}\right), \tag{3.32}$$

$$\hat{\boldsymbol{\Sigma}}_{n|n} = \sum_{i=1}^{M^{l_{\mathrm{GPB}}}} \left[\hat{\boldsymbol{\Sigma}}_{n|n}^{(i)} + \left(\hat{\mathbf{x}}_{n|n}^{(i)} - \hat{\mathbf{x}}_{n|n}\right)^{\mathrm{T}} \left(\hat{\mathbf{x}}_{n|n}^{(i)} - \hat{\mathbf{x}}_{n|n}\right)\right] p\left(d_{n-l_{\mathrm{GPB}}+1:n}^{(i)} \middle| \mathbf{y}_{1:n}\right). \tag{3.33}$$

Here, $\hat{\mathbf{x}}_{n|n}^{(i)}$ and $\hat{\boldsymbol{\Sigma}}_{n|n}^{(i)}$ are obtained by using the Kalman filter conditional on one of $M^{l_{\mathrm{GPB}}}$ possible sequences $d_{n-l_{\mathrm{GPB}}+1:n}^{(i)}$, $i = 1, \dots M^{l_{\mathrm{GPB}}}$. At the next time step, these mean and covariance are updated for the new sequence $\left\{d_{n-l_{\mathrm{GPB}}+2:n}^{(i)}, d_{n+1}\right\}$ that contain all but first parameters of the sequence $d_{n-l_{\mathrm{GPB}}+1:n}^{(i)}$, and one of the possible values of the indicator variable d_{n+1} at time $n+1$. A total number of $M^{l_{\mathrm{GPB}}+1}$ Kalman steps is required at each stage to perform these calculations. As $l_{\mathrm{GPB}} \to n$, the GPB approaches the optimal solution.

The basic form of the algorithm was first proposed in [Ackerson & Fu, 1970], and later several more general versions appeared in the literature. The relevant references can be found in [Tugnait, 1982] who also compared this approach with previously described DEA and RSA methods. The conclusion was in favour of GPB, although, in general, the performance was found to be problem dependent.

3.2.4.4 Interacting Multiple Model Algorithm

The Interacting Multiple Model (IMM) method, which is, probably, the most popular among the others, is, in fact, quite similar to the GPB with $l_{\mathrm{GPB}} = 0$. It is slightly more accurate, however, as it delays the mixture reduction, as a result of which, the probability density is never reduced to a single Gaussian. This makes a big difference if the transition probabilities p_{ij}, Equation (3.26), vary for different values of i, j, however, the two filters are identical if they are equal.

The detailed description of the algorithm is given in [Blom & Bar-Shalom, 1988], where the comparison of the two approaches is also presented.

3.2.5 Grid-based methods

A different strategy for non-linear non-Gaussian filtering is to evaluate the filtering density function at a number of prescribed grid points [Sorenson, 1988], i.e. approximate $p(\mathbf{x}_n | \mathbf{y}_{1:n-1})$ by a discrete distribution, or use these points as the basis for an approximation by a simple function, such as splines [Bucy & Youssef, 1974], or a piecewise linear function [Kitagawa, 1987].

Such approximation simplifies integration in Equation (3.11-3.13), however, the calculations at each point might be quite time-consuming, and the number of them needs to be increased exponentially as the dimension of the state space grows. This results in a quite computationally expensive filter not suitable for high dimensional problems. Moreover, it is not at all clear how to choose the grid points, or, indeed, what constitutes a good choice. In addition, due to the moving state, a new choice might have to be made at each time step.

As one can see, while a large number of different filters have been proposed in the literature, all of them suffer from quite serious drawbacks and do not provide a general methodology for non-linear non-Gaussian filtering. The *Sequential Monte Carlo* (SMC) also known as *particle filtering* approach, described in details in the next section, shows a great promise in providing such a fundamental technology.

3.3 Sequential Monte Carlo

The idea to use Monte Carlo integration methods in filters can be traced back to [Handschin & Mayne, 1969; Akashi & Kumamoto, 1977; Zaritskii *et al.*, 1975]. However, it is not until recently, with the increase of computational power, that Monte Carlo based filters gained much interest in different areas of statistics and engineering. The interest has arisen with the proposal of the so called *bootstrap* filter [Gordon *et al.*, 1993], simultaneously developed by [Kitagawa, 1993]. Since then, Sequential Monte Carlo (SMC) algorithms, under the names of *particle filters*, sequential importance resampling (SIR) and condensation trackers, have been applied to a wide range of problems in the fields of engineering, financial data analyses, genetics, medicine, biology, to name a few (see [Doucet *et al.*, 2001a] for the list of references).

The Sequential Monte Carlo methods approximate the posterior distribution of interest by swarms of points in the sample space, called *particles*, which evolve randomly in time according to a simulation-based rule, and either give birth to offspring particles or die according to their ability to represent the different zones of interest of the state space dictated by the observation process and the dynamics of the underlying system. Some similarities can be seen with the grid based methods, however, particle filters use adaptive stochastic grid approximation, i.e. naturally follow the movement of the state instead of being chosen arbitrary by the user. The probability density of interest is, thus, represented more accurately, and, in addition, the rate of convergence of the approximation error towards zero is theoretically not sensitive to the size of the state space [Crisan *et al.*, 1999].

This section describes the basic particle filtering algorithm based on sequential importance sampling and resampling (SIR), and points out some of its limitations. The issues concerning some improvements on SIR are discussed in the next section.

3.3.1 Monte Carlo Methods

Monte Carlo methods are commonly used for approximation of intractable integrals and rely on the ability to draw a random sample from the required probability distribution. The idea is to simulate N independent identically distributed (*i.i.d.*) samples $\left\{\mathbf{x}_{0:n}^{(i)}\right\}_{i=1}^{N}$ from the distribution of interest, which is in our case the posterior $p\left(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$, and use them to obtain an empirical estimate of the distribution:

$$\hat{p}_N\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(d\mathbf{x}_{0:n} - \mathbf{x}_{0:n}^{(i)}\right). \tag{3.34}$$

The function $\delta\left(d\mathbf{x}_{0:n} - d\mathbf{x}_{0:n}^{(i)}\right)$ here denotes the Dirac delta function.

The expected value of $\mathbf{x}_{0:n}$

$$\mathbb{E}_{p(\,\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\mathbf{x}_{0:n}\right] = \int \mathbf{x}_{0:n} p\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right), \tag{3.35}$$

or, indeed, of any function $\Upsilon(\cdot)$ of $\mathbf{x}_{0:n}$:

$$\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}[\Upsilon(\mathbf{x}_{0:n})] = \int \Upsilon(\mathbf{x}_{0:n})\, p(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}), \tag{3.36}$$

can be obtained consequently by approximating the corresponding integrals by the sums:

$$\mathbb{E}_{\hat{p}_N(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}[\mathbf{x}_{0:n}] = \int \mathbf{x}_{0:n}\hat{p}_N(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{0:n}^{(i)}, \tag{3.37}$$

$$\mathbb{E}_{\hat{p}_N(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}[\Upsilon(\mathbf{x}_{0:n})] = \int \Upsilon(\mathbf{x}_{0:n})\,\hat{p}_N(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) = \frac{1}{N}\sum_{i=1}^{N}\Upsilon\left(\mathbf{x}_{0:n}^{(i)}\right). \tag{3.38}$$

The estimate (3.38) is unbiased with the variance proportional to $1/N$ for the finite variance of $\Upsilon(\mathbf{x}_{0:n})$ (see [Doucet *et al.*, 2000] for more details), and is easily obtained providing one can sample from $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$. This is usually not the case, however, with $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$ being multivariate, non standard and typically only known up to a normalizing constant, and in the next section we briefly discuss an alternative approach for drawing random samples from such distribution.

3.3.2 Bayesian Importance Sampling

The approach is based on the following remark. Suppose one cannot efficiently sample from $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$, however, there is another arbitrary convenient probability distribution $\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$ (such that $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) > 0$ implies $\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) > 0$) which is easy to sample from. Then the estimate of the function $\Upsilon(\cdot)$ of $\mathbf{x}_{0:n}$ can be represented as

$$\begin{aligned}
\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}[\Upsilon(\mathbf{x}_{0:n})] &= \int \Upsilon(\mathbf{x}_{0:n})\frac{p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})}{\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})}\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})\,d\mathbf{x}_{0:n} \\
&= \mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\frac{p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})}{\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})}\right],
\end{aligned} \tag{3.39}$$

with $w(\mathbf{x}_{0:n})$ being the *importance weight*

$$w(\mathbf{x}_{0:n}) = \frac{p\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}{\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}. \tag{3.40}$$

Taking into account the Bayes' theorem (3.3) one obtains:

$$\begin{aligned}
\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right] &= \int \Upsilon\left(\mathbf{x}_{0:n}\right) \frac{p\left(\mathbf{y}_{1:n}\middle|\,\mathbf{x}_{0:n}\right) p\left(\mathbf{x}_{0:n}\right)}{p\left(\mathbf{y}_{1:n}\right) \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)} \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n} \\
&= \frac{1}{p\left(\mathbf{y}_{1:n}\right)} \int \Upsilon\left(\mathbf{x}_{0:n}\right) w_n \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n}, \qquad (3.41)
\end{aligned}$$

where w_n are so called *unnormalized importance weights*:

$$w_n = \frac{p\left(\mathbf{y}_{1:n}\middle|\,\mathbf{x}_{0:n}\right) p\left(\mathbf{x}_{0:n}\right)}{\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}, \tag{3.42}$$

and $\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)$ is called the *importance distribution.*

From (3.41) follows

$$\begin{aligned}
\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\mathbf{x}_{0:n}\right] &= \frac{1}{p\left(\mathbf{y}_{1:n}\right)} \int \Upsilon\left(\mathbf{x}_{0:n}\right) w_n \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n} \\
&= \frac{\int \Upsilon\left(\mathbf{x}_{0:n}\right) w_n \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n}}{\int p\left(\mathbf{y}_{1:n}\middle|\,\mathbf{x}_{0:n}\right) p\left(\mathbf{x}_{0:n}\right) \dfrac{\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}{\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)} d\mathbf{x}_{0:n}} \\
&= \frac{\int \Upsilon\left(\mathbf{x}_{0:n}\right) w_n \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n}}{\int w_n \pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) d\mathbf{x}_{0:n}} \qquad (3.43) \\
&= \frac{\mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[\Upsilon\left(\mathbf{x}_{0:n}\right) w_n\right]}{\mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[w_n\right]}.
\end{aligned}$$

The estimate of interest can, thus, be approximated by

$$\begin{aligned}
\mathbb{E}_{\hat{p}_N(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right] &= \frac{\frac{1}{N}\sum_{i=1}^{N} \Upsilon\left(\mathbf{x}_{0:n}^{(i)}\right) w_n^{(i)}}{\frac{1}{N}\sum_{i=1}^{N} w_n^{(i)}} \qquad (3.44) \\
&= \sum_{i=1}^{N} \Upsilon\left(\mathbf{x}_{0:n}^{(i)}\right) \tilde{w}_n^{(i)},
\end{aligned}$$

where $\mathbf{x}_{0:n}^{(i)}$ denote samples drawn from $\pi\left(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$, and $\tilde{w}_n^{(i)}$ are the *normalized importance weights*:

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{i=1}^{N} w_n^{(i)}}. \tag{3.45}$$

This estimate is biased since it involves a ratio of estimates. However, under the assumptions of $\mathbf{x}_{0:n}^{(i)}$ being a set of i.i.d. samples drawn from $\pi\left(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$, the support of $\pi\left(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$ including the support of $p\left(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$, and $\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right]$ existing and being finite, one could obtain a convergence of the empirical distribution $\sum_{i=1}^{N} \tilde{w}\left(\mathbf{x}_{0:n}^{(i)}\right)\delta\left(d\mathbf{x}_{0:n} - \mathbf{x}_{0:n}^{(i)}\right)$ towards $p\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$ in the sense of almost sure convergence of $\mathbb{E}_{\hat{p}_N(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right]$ towards $\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right]$ as $N \rightarrow +\infty$. Under the additional assumptions of the expectations $\mathbb{E}_{p(\cdot|\mathbf{y}_{1:n})}\left[\Upsilon^2(\mathbf{x}_{0:n})w(\mathbf{x}_{0:n})\right]$ and $\mathbb{E}_{p(\cdot|\mathbf{y}_{1:n})}\left[w(\mathbf{x}_{0:n})\right]$ existing and being finite, a central limit theorem also holds. The details are given in [Doucet *et al.*, 2000; Geweke, 1989].

The above means that the algorithm could be interpreted as a simulation based method for sampling from $p\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$, with $p\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$ being approximated by point mass estimate:

$$\hat{p}_N\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_n^{(i)}\delta\left(d\mathbf{x}_{0:n} - \mathbf{x}_{0:n}^{(i)}\right). \tag{3.46}$$

The "*perfect*" simulation case would correspond to

$$\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) = p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}), \tag{3.47}$$

and $\tilde{w}_n^{(i)} = N^{-1}$ for all i.

3.3.3 Sequential Importance Sampling

The method described up to now is a batch method. In order to obtain the estimate of $p\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$ sequentially, one should be able to propagate $\hat{p}_N\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)$ in time without modifying the past simulated states $\{\mathbf{x}_{0:n}^{(i)}\}_{i=1}^{N}$.

In general, the importance function $\pi\left(d\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)$ can be expanded as:

$$\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)=\pi\left(\mathbf{x}_{0}\middle|\,\mathbf{y}_{1:n}\right)\prod_{j=1}^{n}\pi\left(\mathbf{x}_{j}\middle|\,\mathbf{x}_{0:j-1},\mathbf{y}_{1:n}\right). \tag{3.48}$$

In order to fulfil the above condition, though, the adopted distribution should be of the following form:

$$\begin{aligned}\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) &= \pi\left(\mathbf{x}_{0}\right)\prod_{j=1}^{n}\pi\left(\mathbf{x}_{j}\middle|\,\mathbf{x}_{0:j-1},\mathbf{y}_{1:j}\right) \qquad (3.49)\\ &= \pi\left(\mathbf{x}_{0:n-1}\middle|\,\mathbf{y}_{1:n-1}\right)\pi\left(\mathbf{x}_{n}\middle|\,\mathbf{x}_{0:n-1},\mathbf{y}_{1:n}\right),\end{aligned}$$

i.e. it should admit $\pi\left(\mathbf{x}_{0:n-1}\middle|\,\mathbf{y}_{1:n-1}\right)$ as marginal distribution.

A recursive expression for the importance weights can then be derived by substituting Equations (3.49) and (3.5-3.6) into Equation (3.42):

$$\begin{aligned}w_n &= \frac{p\left(\mathbf{x}_{0:n-1}\middle|\,\mathbf{y}_{1:n-1}\right)p\left(\mathbf{y}_{n}\middle|\,\mathbf{x}_{n}\right)p\left(\mathbf{x}_{n}\middle|\,\mathbf{x}_{n-1}\right)}{\pi\left(\mathbf{x}_{0:n-1}\middle|\,\mathbf{y}_{1:n-1}\right)\pi\left(\mathbf{x}_{n}\middle|\,\mathbf{x}_{0:n-1},\mathbf{y}_{1:n}\right)}\\ &= w_{n-1}\frac{p\left(\mathbf{y}_{n}\middle|\,\mathbf{x}_{n}\right)p\left(\mathbf{x}_{n}\middle|\,\mathbf{x}_{n-1}\right)}{\pi\left(\mathbf{x}_{n}\middle|\,\mathbf{x}_{0:n-1},\mathbf{y}_{1:n}\right)}. \qquad (3.50)\end{aligned}$$

The algorithm could be initialized by sampling $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0^{(i)})$ and setting w_0 $= 1$ for $i = 1, \ldots, N$. For $n \geq 1$, it proceeds as follows:

Sequential Importance Sampling

- For $i = 1, \ldots, N$, sample $\mathbf{x}_n^{(i)} \sim \pi(\mathbf{x}_n|\,\mathbf{x}_{0:n-1}^{(i)}, \mathbf{y}_{1:n})$ and set $\mathbf{x}_{0:n}^{(i)} = (\mathbf{x}_{0:n-1}^{(i)}, \mathbf{x}_n^{(i)})$.

- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$w_n^{(i)} \propto w_{n-1}^{(i)}\frac{p\left(\mathbf{y}_n\middle|\,\mathbf{x}_n^{(i)}\right)p\left(\mathbf{x}_n^{(i)}\middle|\,\mathbf{x}_{n-1}^{(i)}\right)}{\pi\left(\mathbf{x}_n^{(i)}\middle|\,\mathbf{x}_{0:n-1}^{(i)},\mathbf{y}_{1:n}\right)}.$$

- For $i = 1, \ldots, N$, normalize the importance weights:

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

The numerical complexity of this algorithm is $O(N)$, with its greatest advantage being parallelisability. The memory requirements are $O((n+1)N)$, as it is necessary to keep all the simulated trajectories. If, however, one is interested in the estimation of filtering distribution $p\left(\mathbf{x}_n \middle| \mathbf{y}_{1:n}\right)$ only, and the chosen importance distribution does not depend on the previous values of particles $\left\{\mathbf{x}_{0:n-2}^{(i)}\right\}_{i=1}^{N}$, the memory requirements are $O(N)$.

3.3.4 Selection procedure

3.3.4.1 Degeneracy of the algorithm

Unfortunately, the algorithm presented above has a serious limitation. What happens is that, after running it for a few iterations, one typically finds that one of the normalized importance weights tends to 1, while the remaining weights are negligible. A large computational effort, thus, is directed into updating trajectories which virtually do not contribute to the final estimate.

This occurs due to increase of the variance of the importance weights over time which has a negative effect on the accuracy of the algorithm. Indeed, in the ideal case, when we are able to sample directly from the distribution of interest $\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}) = p(\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$, the mean and the variance of the importance weights are equal to 1 and 0 correspondingly:

$$\mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[w(\mathbf{x}_{0:n})\right] = \mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[\frac{p\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}{\pi\left(\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)}\right] = 1, \tag{3.51}$$

$$var_{\pi(\cdot|\mathbf{y}_{1:n})}\left[w(\mathbf{x}_{0:n})\right] = \mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[\left(\frac{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}{\pi(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})} - \mathbb{E}_{\pi(\cdot|\mathbf{y}_{1:n})}\left[\frac{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}{\pi(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\right]\right)^2\right] = 0, \tag{3.52}$$

and this is the situation we would like to be as close to as possible. It is, however, proven that the unconditional variance (that is, when the measurements are regarded as random) of the importance weights increases over time [Doucet *et al.*, 2000; Kong *et al.*, 1994], and, therefore, it is impossible to avoid a degeneracy of the algorithm.

3.3.4.2 Selection step

One way of limiting the degeneracy would be to choose an appropriate importance function which minimizes the conditional variance of the importance weights (see more details in Subsection 3.4.1.1.2). Unfortunately, this is not always possible. Another approach would be to introduce a forgetting factor on the weights as in [Moral & Salut, 1995]. Although this technique does slow down the degeneracy, the problem is not eliminated completely. A solution might be to introduce a selection procedure in the algorithm which discards the low weighted trajectories, and replicates samples with the high normalized importance weights as in Sequential Importance Resampling (SIR) first suggested in [Rubin, 1988].

In general, a selection scheme introduces an additional sampling step based on the obtained discrete distribution, which places a probability mass $\tilde{w}_n^{(i)}$ at each of the original points, say $\tilde{\mathbf{x}}_{0:n}^{(i)}$, referred to as *particles*:

$$\hat{p}_N\left(d\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(d\mathbf{x}_{0:n} - \tilde{\mathbf{x}}_{0:n}^{(i)}\right). \tag{3.53}$$

At each step, a certain number of offspring (replicas), say N_i, is assigned to each trajectory $\tilde{\mathbf{x}}_{0:n}^{(i)}$ according to its weight $\tilde{w}_n^{(i)}$, with N_i being equal to 0 for particles with negligible weights that should be discarded. These values are chosen so that a total number of samples in the scheme stays the same, i.e. $\sum_{i=1}^{N} N_i = N$. The resulting samples $\left\{\mathbf{x}_{0:n}^{(i)}\right\}_{i=1}^{N}$ form an approximate sample from $p\left(d\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)$ and the approximating distribution follows as

$$\hat{p}_N\left(d\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(d\mathbf{x}_{0:n} - \mathbf{x}_{0:n}^{(i)}\right), \tag{3.54}$$

that is the weights are reset to N^{-1}, and the "surviving" particles $\left\{\mathbf{x}_{0:n}^{(i)}\right\}_{i=1}^{N}$ are distributed approximately according to $p\left(d\mathbf{x}_{0:n}\middle|\,\mathbf{y}_{1:n}\right)$.

3.3.4.3 Particle filtering algorithm

The resulting algorithm for the nth step, with $n \geq 1$, $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$ $w_0 = 1$ for $i = 1, \ldots, N$, is described below.

Particle Filtering Algorithm

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $\tilde{\mathbf{x}}_n^{(i)} \sim \pi(\mathbf{x}_n | \mathbf{x}_{0:n-1}^{(i)}, \mathbf{y}_{1:n})$ and set $\tilde{\mathbf{x}}_{0:n}^{(i)} = (\mathbf{x}_{0:n-1}^{(i)}, \tilde{\mathbf{x}}_n^{(i)})$.

- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$w_n^{(i)} \propto w_{n-1}^{(i)} \frac{p\left(\mathbf{y}_n | \tilde{\mathbf{x}}_n^{(i)}\right) p\left(\tilde{\mathbf{x}}_n^{(i)} \middle| \tilde{\mathbf{x}}_{n-1}^{(i)}\right)}{\pi\left(\tilde{\mathbf{x}}_n^{(i)} \middle| \tilde{\mathbf{x}}_{0:n-1}^{(i)}, \mathbf{y}_{1:n}\right)}.$$

- For $i = 1, \ldots, N$, normalize the importance weights:

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

Selection Step

- Multiply / discard particles $\left\{\tilde{\mathbf{x}}_{0:n}^{(i)}\right\}_{i=1}^{N}$ with respect to high/low normalized importance weights $\tilde{w}_n^{(i)}$ to obtain N particles $\left\{\mathbf{x}_{0:n}^{(i)}\right\}_{i=1}^{N}$ with $\tilde{w}_n^{(i)} = \frac{1}{N}$.

Numerous algorithms proposed in the literature are, in fact, special cases of this general (and simple) algorithm.

3.3.4.4 Selection Schemes

A number of different selection schemes have been previously proposed in the literature. These include sequential importance resampling, residual resampling and stratified sampling briefly presented below. All these schemes are unbiased, *i.e.* insure that $\mathbb{E}[N_i] = N\widetilde{w}_n^{(i)}$ at each step n, while they differ in terms of $var[N_i]$ and computational load.

While the most popular *Sequential Importance Resampling* (SIR) / *Multinomial Resampling*, based on drawing uniformly from the discrete set $\left\{\tilde{\mathbf{x}}_{0:n}^{(i)}\right\}_{i=1}^{N}$ with probabilities $\left\{\tilde{w}_n^{(i)}\right\}_{i=1}^{N}$, can be implemented in $O(N)$ operations [Ripley, 1987; Doucet *et al.*, 2000] instead of original $O(N \log N)$ [Gordon *et al.*, 1993], the variance for this algorithm is $N\widetilde{w}_n^{(i)}\left(1 - \widetilde{w}_n^{(i)}\right)$. This variance could be reduced by setting $\widetilde{N}_i$ equal to the integer part of $N\widetilde{w}_n^{(i)}$, and then selecting the remaining $\overline{N} = N - \sum_{i=1}^{N} \widetilde{N}_i$ samples with the new weights $\widetilde{w}_n'^{(i)} = \overline{N}^{-1}\left(N\widetilde{w}_n^{(i)} - \widetilde{N}_i\right)$ using SIR. The result is *Residual Resampling* with the lower variance of $\overline{N}\widetilde{w}_n'^{(i)}\left(1 - \widetilde{w}_n'^{(i)}\right)$ (see [Liu & Chen, 1998], for example).

However, one could do even better by employing *Stratified Sampling* [Kitagawa, 1996; Carpenter *et al.*, 1999]. The idea is to generate N points equally spaced in the interval $[0, 1]$, and to set the number of offspring N_i for each particle to be equal to the number of points lying between the partial sums of weights q_{i-1} and q_i, where $q_i = \sum_{j=1}^{i} \widetilde{w}_t^{(j)}$. This scheme can still be implemented in $O(N)$ operations and has the minimum variance one can achieve in the class of unbiased schemes [Crisan, 2001], namely,

$$var\left[N_i\right] = \left\{N\widetilde{w}_n^{(i)}\right\}\left(1 - \left\{N\widetilde{w}_n^{(i)}\right\}\right),$$

where, for any α, $\lfloor \alpha \rfloor$ is the integer part of α, and $\{\alpha\} \triangleq \alpha - \lfloor \alpha \rfloor$.

Recent theoretical results [Crisan *et al.*, 1999; Crisan, 2001] suggest that it is not necessary for the selection schemes to be unbiased, *i.e.* it is possible to have $N_i \neq N\widetilde{w}_t^{(i)}$, or to be randomized [Kitagawa, 1994].

3.3.4.5 Whether to resample

It has been argued in [Liu & Chen, 1995, 1998] that when all the importance weights are nearly equal, it is not beneficial to introduce the selection step in the algorithm, and a measure of degeneracy in a form of the *effective sample size* N_{eff} is proposed. It is suggested that one should resample only if N_{eff} defined as

$$N_{eff} = \frac{1}{\sum_{i=1}^{n} \left(\tilde{w}_n^{(i)}\right)^2}, \tag{3.55}$$

is below a fixed threshold. This is an intuitively reasonable result, however, the performance largely depends on the threshold employed. In addition, although resampling increases the variance of the estimate $\mathbb{E}_{p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})}\left[\Upsilon(\mathbf{x}_{0:n})\right]$ at time n, it usually leads to a decrease of variance of the future estimates [Liu & Chen, 1995; Doucet *et al.*, 2001b].

3.3.4.6 Problems with selection procedure

The selection procedure helps to solve the problems associated with the degeneracy of the algorithm. However, it introduces some practical and theoretical limitations. In particular, the statistical independence of the simulated trajectories no longer holds, and the convergence results of the algorithm should be re-established (see [Berzuini *et al.*, 1997]). The parallelisability of the algorithm employing the resampling procedure is also limited, and there is loss of diversity due to numerous copies of the same particles in the approximating sample [Gordon *et al.*, 1993; Higuchi, 1995a,b]. Despite these limitations, the algorithm is the basis for numerous works, and many strategies have been developed to increase its efficiency.

3.4 Efficient implementation of particle filters

The particle filter forms an approximate *weighted* sample from the required posterior using convenient proposal distribution, and, after resampling, produces a point-mass estimate of this posterior given by:

$$\hat{p}_N\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(d\mathbf{x}_{0:n} - \mathbf{x}_{0:n}^{(i)}\right). \tag{3.56}$$

The success of this operation depends on how accurate the employed sampling scheme is (i.e. how close the proposal distribution is to the posterior) and whether the resulting point-mass approximation is an adequate representation of the distribution of interest. The former is connected to the choice of the proposal distribution, or, indeed, of the sampling scheme itself, whereas the latter is related to the possible loss of diversity in the samples as a result of the selection. Not surprisingly, a lot of work on particle filtering methods was concentrated in these two areas as well as general reduction of the variance of the Monte Carlo integration.

3.4.1 Improvements on the accuracy of the sampling scheme

Sequential importance sampling produces an approximate sample from the distribution of interest, and, thus, a lot depends on how well the proposal distribution is chosen, or, in some cases, whether an alternative sampling scheme should be employed so that a true, independent sample from the target distribution could be obtained. The issues associated with this question are briefly discussed in this subsection.

3.4.1.1 Choice of the importance density

Although sequential importance sampling poses only one restriction on the importance density, Equation (3.49), with the number of choices, otherwise, being unlimited, the design of the appropriate proposal function is, in fact, one of the most critical issues in importance sampling algorithms. Poor choice leads to poor approximation in (3.56), and to poor algorithm performance in general. The problem, therefore, has received a lot of interest in the literature, with different importance distributions being advocated by different researchers. The most popular choices are described below.

3.4.1.1.1 Prior distribution

This by all means is the most popular and most widely used [Handschin & Mayne, 1969; Gordon *et al.*, 1993] proposal distribution, which is largely due to the simplicity of its implementation. If one takes

$$\pi\left(\mathbf{x}_n | \mathbf{x}_{0:n-1}, \mathbf{y}_{1:n}\right) = p\left(\mathbf{x}_n | \mathbf{x}_{n-1}\right), \tag{3.57}$$

the importance weights could be evaluated straightforwardly:

$$w_n^{(i)} \propto w_{n-1}^{(i)} p\left(\mathbf{y}_n | \mathbf{x}_n^{(i)}\right), \tag{3.58}$$

often resulting in a reduced computational complexity of the designed algorithm (see Chapter 4 for an illustration). This distribution, however, does not incorporate the information contained in the most recent observations, and, therefore, may be inefficient and especially sensitive to outliers.

3.4.1.1.2 Optimal distribution

This choice of the proposal distribution was introduced in [Zaritskii *et al.*, 1975] and later used for a particular case in [Akashi & Kumamoto, 1977], in [Liu & Chen, 1995], and in [Doucet *et al.*, 2000], where list of references could also be found. The distribution of the form

$$\pi\left(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{y}_{1:n}\right) = p\left(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{y}_n\right), \tag{3.59}$$

is, indeed, optimal in a sense that it minimizes the variance of the importance weights $w(\mathbf{x}_{1:n})$ conditional upon the simulated trajectory $\mathbf{x}_{1:n-1}^{(i)}$ and the observations $\mathbf{y}_{1:n}$ [Doucet *et al.*, 2000], thus, limiting the problem of degeneracy of the algorithm (Subsection 3.3.4.1). Interestingly, the weights in this case do not depend on the current value of the state, $\mathbf{x}_n^{(i)}$:

$$\begin{aligned} w_n^{(i)} &\propto w_{n-1}^{(i)} \frac{p\left(\mathbf{y}_n | \mathbf{x}_n^{(i)}\right) p\left(\mathbf{x}_n^{(i)} \middle| \mathbf{x}_{n-1}^{(i)}\right)}{p\left(\mathbf{x}_n^{(i)} \middle| \mathbf{x}_{n-1}^{(i)}, \mathbf{y}_n\right)} && (3.60) \\ &\propto w_{n-1}^{(i)} p\left(\mathbf{y}_n | \mathbf{x}_{n-1}^{(i)}\right), && (3.61) \end{aligned}$$

which facilitates parallelisation of the simulation of $\left\{\mathbf{x}_n^{(i)}\right\}_{i=1}^{N}$ and evaluation of $w_n^{(i)}$ for $i = 1, \ldots, N$. The selection step in this case can be done prior to the sampling step. Unfortunately, sampling from $p\left(\mathbf{x}_n|\,\mathbf{x}_{n-1}^{(i)},\mathbf{y}_n\right)$, and evaluating $p\left(\mathbf{y}_n|\,\mathbf{x}_{n-1}^{(i)}\right)$

$$p\left(\mathbf{y}_n|\,\mathbf{x}_{n-1}\right) = \int p\left(\mathbf{y}_n|\,\mathbf{x}_n\right) p\left(\mathbf{x}_n|\,\mathbf{x}_{n-1}\right) d\mathbf{x}_n, \tag{3.62}$$

for many models is impossible, in which case other, suboptimal, importance distributions should be designed.

3.4.1.1.3 Suboptimal distributions

Since any probability density fulfilling the condition specified in (3.49) could be used as an importance density, a great number of "clever" suboptimal proposals could be derived. Unfortunately, there is still no general strategy specifying how to perform this procedure, and the design of each suboptimal function should be considered on the case by case basis.

One of the examples is the technique recently suggested in the literature and based on building a suboptimal distribution by local linearisation of the state space model similar to the Extended Kalman Filter. This approach is described in great detail in [Doucet *et al.*, 2000], where a number of other strategies for the importance distribution design are also discussed.

3.4.1.2 Auxiliary particle filter

Another method of reducing the variability of the importance weights was introduced in [Pitt & Shephard, 1999, 2001], resulting in an algorithm known as an *auxiliary* particle filter (APF). At $(n-1)$th step, the filter resamples the particles with probability close to $p\left(\mathbf{x}_{n-1}|\,\mathbf{y}_{1:n}\right)$, hence, taking into account the information carried by the new measurement $\mathbf{y}_n$. Such a "look ahead" allows us to sample more efficiently in comparison with the standard methods, and ,as a result, the algorithm is more robust to outliers. In practice, $p\left(\mathbf{x}_{n-1}|\,\mathbf{y}_{1:n}\right)$ is, of course, unavailable analytically and must be approximated. The approximation can be based, for instant, on the predictions $\hat{\mathbf{x}}_n$ of $\mathbf{x}_n$ from each particle $\mathbf{x}_n^{(i)}$, the details can be found in [Godsill & Clapp, 2001; Pitt & Shephard, 2001].

3.4.1.3 Rejection sampling

Importance sampling produces an approximate sample from the distribution of interest and might, therefore, be inaccurate. A solution would be to employ a related but slightly different method called rejection sampling [Hurzeler & Kunsch, 1998]. The technique is based on generating a proposal and then accepting or rejecting it with certain acceptance probability. As a result, a true, independent sample from the required distribution is obtained. Thus, if the proposal density is quite different from the target one, the method naturally compensates by sampling more points from the required distribution. This results, however, in an unpredictable number of iteration required to complete each step, and proves to be extremely computationally expensive in high-dimensional spaces. In general, the comparison of both methods is given in [Liu, 1996], where the conclusion in favour of importance sampling is drawn. In [Pitt & Shephard, 1999] some directions on increasing the efficiency of the approach by adopting an auxiliary filtering technique are presented.

A similar idea is actually employed by [Gordon *et al.*, 1993] when he introduces *prior editing* to the SIR filter in a form of acceptance/rejection test based on the sample weight. This is similar to increasing the number of particles in order to approximate the required distribution better yet without the increased memory capacity and computational cost of resampling.

3.4.1.4 Markov chain Monte Carlo

A third approach to sampling from the target distribution is to use the Markov chain Monte Carlo (MCMC) techniques [Berzuini *et al.*, 1997; Gordon & Whitby, 1995]. The idea is to construct a Markov chain that allows one to sample from $p\left(\mathbf{x}_n \middle| \mathbf{x}_{0:n-1}, \mathbf{y}_{1:n}\right)$. However, it may take a lot of time to converge, particularly if the amount of system noise is small, and can generally be too computationally demanding for on-line analyses. The MCMC methods may help, however, with solution of other problem associated with particle filtering, in particular, samples depletion.

3.4.2 Problems with sample depletion

The problem with sample depletion arises at the resampling stage, when the particles with high importance weights are duplicated multiple times, resulting in numerous identical samples. As a result, the discrete approximation of the distribution of interest would be inaccurate simply due to the number of samples being inadequate.

Of course, the total number of particles could be increased, however, this would be a rather brutal strategy. Alternatively, *prior boosting* [Gordon *et al.*, 1993] could be employed, increasing the number of the sample size, say, to $N_{\mathrm{PB}} > N$, with subsequent selection of N particles only. Although a standard SIR with N_{PB} particles has a lower variance, a computational gain is apparent.

More elegant strategies include an introduction of the MCMC step and kernel smoothing. They will now be briefly described.

3.4.2.1 MCMC step

The idea behind the use of MCMC step is based on the fact that, by applying a Markov transition kernel $\Xi\left(d\mathbf{x}_{0:n}|\,\mathring{\mathrm{x}}_{0:n}^{(i)}\right)$ of invariant distribution $p(\,d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$ to each particle $\mathring{\mathrm{x}}_{0:n}^{(i)}$ distributed marginally according to $p(\,d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n})$, we obtain new particles that are still distributed according to this target distribution. In this case, the total variation of the current distribution with respect to the invariant distribution can only decrease [Doucet *et al.*, 2001a; Gilks & Berzuini, 1998], and any standard MCMC method [Robert & Casella, 1999] can be used, employing a Markov kernel that does not even have to be ergodic [Gilks & Berzuini, 1998].

3.4.2.2 Kernel density smoothing

Kernel methods introduce a different approach to the problem of approximation discreteness [Gordon, 1994; Hurzeler & Kunsch, 1998; Liu & West, 2001]. They replace the point-mass estimator of the distribution

$$\hat{p}_N\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N} \delta\left(d\mathbf{x}_{0:n} - \breve{\mathbf{x}}_{0:n}^{(i)}\right). \tag{3.63}$$

by the estimator of the following form:

$$\hat{p}_\Omega\left(d\mathbf{x}_{0:n}|\,\mathbf{y}_{1:n}\right)=\sum_{i=1}^{N}\tilde{w}_n^{(i)}\frac{1}{\varkappa}\Omega\left(\frac{d\mathbf{x}_{0:n}-\breve{\mathbf{x}}_{0:n}^{(i)}}{\varkappa}\right), \tag{3.64}$$

where $\Omega(\cdot)$ is usually a unimodal symmetric density function, centered at $\breve{\mathbf{x}}_{0:n}^{(i)}$ with smoothing parameter $\varkappa$. The sample depletion problem is mitigated by sampling with replacement from $\left\{\tilde{\mathbf{x}}_{0:n}^{(i)}\right\}_{i=1}^{N}$ to obtain $\left\{\breve{\mathbf{x}}_{0:n}^{(i)}\right\}_{i=1}^{N}$, and adding a smoothing term $\xi^{(i)}$ consequently to obtain new particles $\left\{\mathbf{x}_{0:n}^{(i)}\right\}_{i=1}^{N}$:

$$\mathbf{x}_{0:n}^{(i)}=\breve{\mathbf{x}}_{0:n}^{(i)}+\xi^{(i)}, \tag{3.65}$$

Here $p\left(\xi^{(1)},\dots\xi^{(N)}\right)=\prod_{i=1}^{N}\frac{1}{\varkappa}K\left(\frac{\xi^{(i)}}{\varkappa}\right)$, and the estimation of the parameter $\varkappa$ is discussed, for example, in [Silverman, 1986; Liu & West, 2001].

The *jittering* approach proposed in [Gordon *et al.*, 1993], based on adding some random noise to each particle, is nothing but a simple Kernel smoothing technique.

3.4.3 Variance reduction techniques

In general, particle filtering may be viewed as both a sampling and Monte Carlo integration problem, and, therefore, a number of useful methods from Monte Carlo theory can be applied to increase its efficiency. A stratified sampling scheme [Kitagawa, 1996; Carpenter *et al.*, 1999], Subsection 3.3.4.4, is one of the results of such treatment based on the ideas from survey sampling theory. The other strategies include Rao-Blackwellisation, quasi-Monte Carlo methods and stratification, and will be briefly reviewed below.

3.4.3.1 Rao-Blackwellisation and Dimension reduction

Rao-Blackwellisation is a well known technique in mathematical statistics [Casella & Robert, 1996], which, as was noted by [Doucet *et al.*, 2000; Liu & Chen, 1998], can be successfully applied to particle filtering. It is shown there, that for certain important classes of models, a Rao-Blackwellised particle filter could be designed, which samples from a low-dimensional distribution with the rest of the integration being performed analytically. The result is a reduced variance of the estimates of interest [Doucet *et al.*, 2000], and, generally, more efficient algorithm performance. Rao-Blackwellised filter is used throughout this thesis, and described in great details in Chapter 4, and 7.

3.4.3.2 Quasi-Monte Carlo

Another method commonly used for increasing the accuracy of Monte Carlo integration is using quasi-random numbers to generate the points which are well spaced in the state space. The extent of the improvement would depend on the dimension of the space and the size of the sample, however, the results obtained recently are very promising [Paskov & Traub, 1995]. Regularised sampling in [Fearnhead, 1998] is an example of an application of a similar method to particle filtering problem.

3.4.3.3 Stratification

Stratification is yet another example of variance reduction technique aimed at "exploring" the state space in a better way by using systematic sampling. They do this by ensuring that only the right number of particles are generated from each component, and, although the idea has received much attention in a resampling context, systematic sampling procedures can also be used for proposal, and even in rejection sampling framework, see [Künsch, 2003] for details and the list of references.

Out of all non-linear non-Gaussian filters, the Sequential Monte Carlo approach seems to be the most promising one, and we are now in a position to apply it to the problems arising in digital communications. For clarity of presentation, we begin our treatment with a simpler, but nevertheless very important, problem of demodulation under conditions of flat fading channel, and we will try to develop an efficient particle filtering algorithm taking into account the guidelines specified above.

4 Particle Filtering for Demodulation in Flat Fading Channels

In the transmission of digital information over a communication channel, a sequence of binary digits is mapped into a set of corresponding analog waveforms that match the characteristics of the channel. Whatever the physical medium used for transmission of the information, the transmitted signal is corrupted in a random manner by a variety of possible mechanisms, such as thermal noise or interferences of different kinds. The function of the demodulator is to make a decision on the transmitted signal based on the observations. This is a challenging non-linear filtering problem, and it has proved to be especially difficult under conditions of noisy fading transmission channels.

In general, multipath fading occurs when the transmitted signal arrives at the receiver via multiple propagation paths at different delays. The signal components in this case may add destructively, resulting in random amplitude and phase variations on the signal. Thus, demodulation of the signal requires recursive estimation of both the signal and channel distortions given the currently available data. The problem can be formulated in a linear Gaussian state space form (conditional upon the symbols), and within a sequential framework, general recursive expressions may be derived for the filtering and fixed-lag smoothing distributions, from which estimates of the states can be obtained. However, the exact computation of these estimates involves a prohibitive computational cost exponential in the growing number of observations, and thus approximate methods must be employed.

Several classical approximate schemes have, in the past, been proposed to tackle the problem of demodulation, including extended Kalman filtering (EKF) [Haeb & Meyr, 1989; Lodge & Moher, 1990], or coupled EKF and hidden Markov model (HMM) approaches [Collings & Moore, 1994, 1995]. In [Georghiades & Han, 1997] the expectation-maximization (EM) algorithm is applied to demodulation of signals in fading conditions; in [Gertsman & Lodge, 1997] joint demodulation and decoding using iterative processing techniques is developed. Other more recent techniques involve the pilot symbol-aided schemes (PSAM) [Cavers, 1991; Sampei & Sunaga, 1993; Torrance & Hanzo, 1995], decision-feedback [Kam & Ching, 1992; Liu & Blostein, 1995], and per-survivor processing [Vitetta & Taylor, 1995].

In this chapter, we concentrate on the general case of M-ary modulated signals[7] under conditions of noisy fading channels and base our approach on particle filtering techniques. Particle filters are efficient simulation-based algorithms combining sequential importance sampling, a selection scheme and Markov chain Monte Carlo (MCMC) methods in order to perform optimal estimation. The key idea is to use an adaptive stochastic grid approximation of the conditional probability of the state vector with particles (values of the grid) evolving randomly in time according to a simulation-based rule. Depending on their ability to represent the different zones of interest of the state space which is dictated by the observation process and the dynamics of the underlying system, the particles can either give birth to offspring particles or die.

This chapter develops particle filtering techniques to compute optimal

[7] Both coherent (PSK, QAM) and non-coherent (DPSK) demodulation is considered in this chapter.

estimates of the symbols under conditions of frequency-nonselective (flat) Rayleigh fading channels. The tracking of the channel is naturally incorporated into the estimation scheme developed here and the case of a possibly non-Gaussian additive channel noise can be easily treated. A straightforward extension of the algorithm allows us to easily address the fixed-lag smoothing problem and, thus, more accurate delayed symbol and channel estimates can be obtained. Although particle filtering (in the absence of differential encoding) also requires the transmission of pilot symbols in order to prevent cycle slipping, the pilot symbol spacing is not restricted by the Doppler spread of the channel as with PSAM and can be made much longer. The method proposed here is also designed to make use of the structure of the model, and incorporates efficient variance reduction strategies based on Kalman filtering techniques. At each iteration the algorithm has a computational complexity that is linear in the number of particles, and can easily be implemented on parallel processors.

The rest of the chapter is organized as follows. The model specification and estimation objectives are stated in Section 4.1. Section 4.2, develops a particle filtering method to solve the problem of demodulation in the general case of M-ary modulated signals. The bit-error-rate (BER) performance of the M-ary phase shift keyed (PSK), differential phase shift keyed (DPSK) and quadrature amplitude modulated (QAM) signals is examined in Section 4.3 by means of an extensive simulation study. Some conclusions are drawn at the end of the chapter.

4.1 Model Specification and Estimation Objectives

We shall begin our treatment of digital signaling over flat Rayleigh fading channels by formulating the model specified in the first chapter for the particular case of uncoded M-ary modulated signals (see Figure 4.1). For clarity of presentation, since the number of sources K and the number of antennas L are equal to 1, and the channel is flat, the indices k, l and g will be suppressed throughout this chapter. Once the model is specified, we

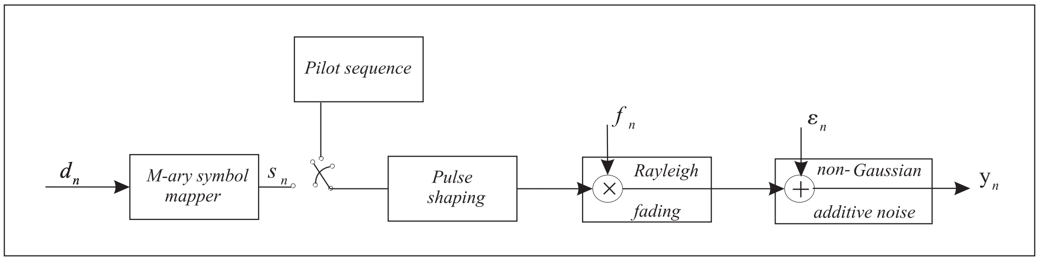

Figure 4.1: Transmission of M-ary modulated signals in Rayleigh fading channels.

motivate the use of the pilot symbols periodically inserted into the symbol stream, and recall the model for the flat Rayleigh fading channel and the equations for the output of the channel corrupted by possibly non-Gaussian noise. Eventually, the problem of demodulation is formulated in a state space form; and, finally, the estimation objectives are stated.

4.1.1 Model Specification

We denote for any sequence α_n, $\alpha_{i:j} \triangleq (\alpha_i, \alpha_{i+1}, \ldots, \alpha_j)^{\mathsf{T}}$ if $i < j$, and $\alpha_{i:j} \triangleq (\alpha_i, \alpha_{i-1}, \ldots, \alpha_j)^{\mathsf{T}}$ otherwise.

4.1.1.1 Representation of digitally modulated signals

In the general case of M-ary modulation, the information sequence is subdivided into blocks of κ binary digits, which are mapped into $M = 2^{\kappa}$ deterministic, finite energy waveforms $\{s_{\text{trans},m}(n),\ m = 1, 2, \ldots, M\}$. Recall that d_n is an indicator variable associated with one of M possible κ-bit sequences

$$d_n \in \mathcal{R} = \{1, 2, \ldots, M\}, \quad n = 1, 2, \ldots \tag{4.1}$$

n being a discrete time index. Since no channel coding is employed in the system, $d_n = r_n$ for $n = 1, 2, \ldots$, with q being equal to n (Chapter 2). The signal waveform transmitted in the signaling interval of duration T (we suppress index q for convenience) may be represented as

$$s_{\text{trans}}(\tau) = \text{Re}[s_n(d_{1:n})\eta(\tau)\exp(j2\pi \mathrm{f}_{\text{car}}\tau)], \quad (n-1)T \leq \tau \leq nT, \tag{4.2}$$

where $\mathrm{f_{car}}$ is a carrier frequency, $\eta(\tau)$ is a real-valued signal pulse which shape influences the spectrum of the transmitted signal (the absence of DS spectrum spreading leads also to $H = 1$, $c_1 = 1$ and $T_{ch} = T$), $d_{1:n}$. $s_n(\cdot)$ performs the mapping from the digital sequence to waveforms and, in a general case of a modulator with memory, it depends on one or more previously transmitted symbols.

In the absence of an encoder, the message symbols are assumed to be *independent identically distributed* (i.i.d.). However, if error correcting coding, such as convolutional coding [Proakis, 1995] or trellis coding [Du & Vucetic, 1991], has been employed, the signals produced contain symbols which are not i.i.d. and as such display Markov properties. Thus, (to a first approximation) it is reasonable to assume that d_n is a first order, time-homogeneous, M-state, Markov process with known transition probabilities

$$p_{ij} = \Pr\left\{ d_{n+1} = j | \, d_n = i \right\}, i, j \in \mathcal{R}, \tag{4.3}$$

such that $p_{ij} \geq 0$, $\sum_{j=1}^{M} p_{ij} = 1$ for each i, and initial probability distribution $p_i = \Pr\{d_1 = i\}$, $p_i \geq 0$, $\sum_{i=1}^{M} p_i = 1$ for $i \in \mathcal{R}$ (see [Collings & Moore, 1994, 1995], for the same approach). Taking into account that $\Pr\left\{ d_{n+1} | \, d_n \right\} = \Pr\left\{ d_{1:n+1} | \, d_{1:n} \right\}$, it is easy to show that $d_{1:n}$ is also a first order Markov process with known transition probability matrix.

4.1.1.2 Pilot symbols

The analog waveforms are then passed to a noisy fading channel which can cause the severe amplitude and phase fluctuations. This may lead to incorrect demodulation even in the absence of noise. If the information sequence has not been differentially encoded, one error might result in the whole cycle slipping (especially under conditions of deep fading and low signal to noise ratio). In order to prevent this, known pilot symbols can be periodically inserted into the transmitted symbol stream (see [Georghiades & Han, 1997; Gertsman & Lodge, 1997; Seymour & Fitz, 1995] for the same approach). However, in our case unlike with pilot-symbol-assisted modulation (PSAM) [Cavers, 1991; Sampei & Sunaga, 1993; Torrance & Hanzo, 1995], the pilot symbol rate is not restricted by the Doppler spread of the channel and the pilot spacing can be made much longer thus reducing the bandwidth penalty incurred.

4.1.1.3 Channel model

In this thesis we concentrate on a frequency-nonselective (flat) fading channel, which, providing that the output is sampled at T^{-1} rate, can be described by a multiplicative discrete time disturbance f_n, $t = n$ in this case. In the case of Rayleigh fading, f_n is a complex low-pass filtered Gaussian process with zero-mean real and imaginary components. (The time variations of f_n are realistically assumed to be slow in comparison to the message rate.)

Ideally, in order to approximate the power spectral density of the fading process, a high order low-pass pole-zero filter is required [Stuber, 1996]. Thus, f_n can be modelled as the following ARMA(ρ, ρ) process (Butterworth filter of order q)

$$f_n = \mathbf{a}^{\mathrm{T}} f_{n-1:n-\rho} + \mathbf{b}^{\mathrm{T}} v_{n:n-\rho}, \tag{4.4}$$

where v_n is a complex white Gaussian noise with i.i.d. zero-mean real and imaginary components of variance 1, $v_n \overset{i.i.d.}{\sim} \mathcal{N}_c(0, 1)$. The coefficients of the filter $\mathbf{a} \triangleq (a_{1,}a_2, \ldots, a_q)^{\mathrm{T}}$, $\mathbf{b} \triangleq (b_{0,}b_1, \ldots, b_q)^{\mathrm{T}}$ are chosen so that the cut-off frequency of the filter matches the normalized channel Doppler frequency $\mathrm{f_d}T$ (T is the symbol rate). We assume here that $\mathrm{f_d}T$ and, thus, $\mathbf{a}$ and $\mathbf{b}$ are known.

4.1.1.4 Observations

At the demodulator, the received signal is passed through a filter whose impulse response is matched to the waveform $\eta(\tau)$ and, as was already mentioned, is sampled at a rate T^{-1}, resulting in one sample corresponding to each symbol. The input of the filter is assumed to be corrupted by additive noise with mutually independent real and imaginary parts, which are i.i.d. random sequences distributed as a mixture of zero-mean, complex, Gaussians with a known number of components Z. Thus, non-Gaussian additive noise could be approximated. A latent allocation variable z_n, $z_n \in$

$\mathcal{Z} = \{1, 2, \ldots, Z\}$, $n = 1, 2, \ldots$, is introduced in order to identify the variance, $\sigma^2_{z_n}$, of the distribution from which one samples. It is chosen such that $\Pr(z_n = \xi) = \lambda_\xi$, for $\xi = 1, \ldots, Z$, $\sum_{\xi=1}^{Z} \lambda_\xi = 1$. Thus, one obtains

$$y_n = s_n(d_{1:n}) f_n + \sigma_{z_n} \epsilon_n, \qquad \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0, 1\right), \tag{4.5}$$

where y_n and ϵ_n are respectively the complex output of the matched filter and the additive complex zero-mean Gaussian noise. Without loss of generality f_n and s_n are normalized to have unity power. This results in the average signal to noise ratio (SNR) per bit equal to $-10 \log\left(N_0 \log_2 M\right)$, where $N_0 = 2\sigma_0^2$ and σ_0^2 is the overall variance of the noise, $\sigma_0^2 = \sum_{\xi=1}^{Z} \lambda_\xi \sigma_\xi^2$.

4.1.2 State Space Signal Model

In order to express the system in a familiar form of a state space model (see for example [West & Harrison, 1997]), we define a state x_n such that

$$f_n = \mathbf{b}^{\mathrm{T}} x_{n:n-\rho+1}. \tag{4.6}$$

Then, from Equation (4.4) one obtains

$$x_n = \mathbf{a}^{\mathrm{T}} x_{n-1:n-\rho} + v_n, \tag{4.7}$$

and, thus, conditional upon the symbols and allocation variables, the problem can be formulated in the following linear state space form

$$\begin{aligned} x_{n:n-\rho+1} &= \mathbf{A} x_{n-1:n-\rho} + \mathbf{B} v_n, \\ y_n &= \mathbf{C}(d_{1:n}) x_{n:n-\rho+1} + \mathbf{D}(z_n) \epsilon_n, \end{aligned} \tag{4.8}$$

where $\mathbf{B} = (1, 0, \ldots, 0)^{\mathrm{T}}$, $\mathbf{C}(d_{1:n}) = \mathbf{s}_n(d_{1:n})\mathbf{b}^{\mathrm{T}}$, $\mathbf{D} = \sigma_{z_n}$ and

$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 & \ldots & a_{\rho-1} & a_\rho \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}. \tag{4.9}$$

We assume $x_{0:1-\rho} \sim \mathcal{N}_c(\hat{\mathbf{x}}_0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0 > 0$, and let v_n, ϵ_n be mutually independent for all $n > 0$. The message symbols d_n, the channel characteristics x_n and the latent allocation variable z_n are unknown for $n > 0$, whereas the model parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}(d_{1:n})$, $\mathbf{D}(z_n)$, $\hat{\mathbf{x}}_0$, $\mathbf{\Sigma}_0$ are known for each $d_n \in \mathcal{R}$, $z_n \in \mathcal{Z}$.

4.1.3 Estimation Objectives

Filtering objectives: obtain the MMAP (marginal maximum *a posteriori*) estimates of the symbols

$$\hat{d}_n = \arg\max_{d_n} p\left(d_n \mid y_{1:n}\right). \tag{4.10}$$

The problem does not admit any analytical solution as computing $p\left(d_n \mid y_{1:n}\right)$ involves a prohibitive computational cost exponential in the (growing) number of observations.

Remark 1 *Although the tracking of the channel is naturally incorporated into the proposed algorithm, in principle, one may also be interested in obtaining the MMSE (conditional mean) estimates of the fading coefficients* x_n*, given by* $\mathbb{E}\left(x_n \mid y_{1:n}\right)$*. Due to the phase ambiguity problem, this cannot be done straightforwardly. However, the estimates of* x_n *could be obtained conditional upon the estimated symbols* $\hat{d}_n$*.*

Remark 2 *In order to obtain more accurate results, if it is possible, one may want to wait for a fixed delay* L_s *before performing the estimation (**fixed-lag***

smoothing*). The delayed MMAP estimates of the symbols in this case are given by* $\arg\max_{d_{n-L_s}} p\left(d_{n-L_s}\middle| y_{1:n}\right)$ *for* $n > L_s$ *where* $L_s \in \mathbb{N}^*$.

4.2 Particle Filtering for Demodulation

In this section, we first describe the advantages of using Monte Carlo (MC) simulation techniques, and show how variance reduction can be achieved by integrating out the states $x_{0:n}$ using the Kalman filter. Then, we present a Rao-Blackwellised version of the sequential importance sampling method, and discuss the choice of the importance distribution and the selection scheme. Finally, we propose a generic particle filtering algorithm to obtain an on-line estimate of the time-varying posterior distribution of the symbols.

4.2.1 Monte Carlo Simulation for the Optimal Filter

Given the observations $y_{1:n}$, all Bayesian inference for the signal model specified in Section 4.1.2 relies on the joint posterior distribution $p\left(d_{1:n}, z_{1:n}, dx_{0:n}\middle| y_{1:n}\right)$. If it is possible to sample N *i.i.d.* samples, called particles, $\left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}, x_{0:n}^{(i)}\right\}_{i=1}^{N}$ according to $p\left(d_{1:n}, z_{1:n}, dx_{0:n}\middle| y_{1:n}\right)$, an empirical estimate of this distribution may be obtained by making use of the Monte Carlo (MC) approximation:

$$\hat{p}_N\left(d_{1:n}, z_{1:n}, dx_{0:n}\middle| y_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(\{d_{1:n}, z_{1:n}, dx_{0:n}\} - \left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}, x_{0:n}^{(i)}\right\}\right). \tag{4.11}$$

As a corollary, an estimate of filtering distribution $p\left(d_n, z_n, dx_n\middle| y_{1:n}\right)$ follows as

$$\hat{p}_N\left(d_n, z_n, dx_n\middle| y_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(\{d_n, z_n, dx_n\} - \left\{d_n^{(i)}, z_n^{(i)}, x_n^{(i)}\right\}\right), \tag{4.12}$$

and the estimates of the marginal posterior distributions $p\left(d_n\middle| y_{1:n}\right)$ and $p\left(x_n\middle| y_{1:n}\right)$ can be easily deduced.

Once these filtering distributions are computed, symbol detection can be performed using MMAP criterion, Equation (4.10), and, consequently, the channel coefficients can be estimated if necessary.

The advantage of using the MC approximation scheme is clear. First of all, the particles are automatically selected in the region of high probability, which is a major improvement over a direct numerical approximation method. Second, the estimates of interest are easily obtained. Moreover, the rate of convergence of these estimates does not depend on n or the dimension of the state space, but only on the number of particles N. Finally, in order to obtain more accurate estimates, the algorithm can be straightforwardly extended to fixed-lag smoothing with a fixed delay L_s. The estimate of the marginal posterior distribution in this case is

$$\hat{p}_N\left(d_{1:n-L_s}, z_{1:n-L_s}\right| y_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(\{d_{1:n-L_s}, z_{n-L_s}\} - \left\{d_{1:n-L_s}^{(i)}, z_{1:n-L_s}^{(i)}\right\}\right). \tag{4.13}$$

However, it is not possible to sample directly from the distribution $p\left(d_{1:n}, z_{1:n}, dx_{0:n}\right| y_{1:n})$ at any n, and, as an alternative strategy, based on particle filtering methods described in the previous chapter is proposed in Subsection 4.2.3. But first we will make some simplifications to the algorithm, which will furthermore lead to a reduction in the variance of the estimates.

4.2.2 Variance Reduction by Rao-Blackwellisation

In fact, the joint posterior density $p\left(d_{1:n}, z_{1:n}, x_{0:n}\right| y_{1:n})$ can be factorized as:

$$p\left(d_{1:n}, z_{1:n}, x_{0:n}\right| y_{1:n}) = p\left(x_{0:n}\right| d_{1:n}, z_{1:n}, y_{1:n})\, p\left(d_{1:n}, z_{1:n}\right| y_{1:n}). \tag{4.14}$$

Given $d_{1:n}, z_{1:n}$, the probability density $p\left(x_{0:n}\right| d_{1:n}, z_{1:n}, y_{1:n})$ is a Gaussian distribution whose parameters may be computed using the Kalman filter. Thus, it is possible to reduce the problem of estimating $p\left(d_{1:n}, z_{1:n}, x_{0:n}\right| y_{1:n})$ to one of sampling from a lower-dimensional distribution $p\left(d_{1:n}, z_{1:n}\right| y_{1:n})$, which intuitively requires a reduced number of samples N in order to reach a given precision. This is proved in [Doucet *et al.*, 2000] where it is shown that the variance of the estimates when $x_{0:n}$ can be integrated out analytically is lower.

The marginal posterior distribution $p\left(x_{0:n}\right|y_{1:n})$ in this case is given by

$$p\left(x_{0:n}\right|y_{1:n}) = \sum_{d_{1:n}}\sum_{z_{1:n}} p\left(x_{0:n}\right|d_{1:n}, z_{1:n}, y_{1:n})\, p\left(d_{1:n}, z_{1:n}\right|y_{1:n}), \tag{4.15}$$

and can be approximated via:

$$\hat{p}_N\left(d_{1:n}, z_{1:n}\right|y_{1:n}) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(\{d_{1:n}, z_{1:n}\} - \left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}\right), \tag{4.16}$$

resulting in:

$$\hat{p}_N\left(x_{0:n}\right|y_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} p\left(x_{0:n}\right|d_{1:n}^{(i)}, z_{1:n}^{(i)}, y_{1:n}\Big). \tag{4.17}$$

Then, the estimate of x_n is obtained as

$$\mathbb{E}_{\hat{p}_N(x_{0:n}|y_{1:n})}\left[x_n\right] = \int x_n \hat{p}_N\left(x_{0:n}\right|y_{1:n})\, dx_{0:n} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(x_n\right|d_{1:n}^{(i)}, z_{1:n}^{(i)}, y_{1:n}\Big), \tag{4.18}$$

where $\mathbb{E}\left(x_n\right|d_{1:n}^{(i)}, z_{1:n}^{(i)}, y_{1:n}\Big)$ is computed through the Kalman filter associated with Equation (4.8) (see Chapter 3).

Thus, one should focus now on obtaining the approximation of $p\left(d_{1:n}, z_{1:n}\right|y_{1:n})$, and as was already mentioned, one solution to estimate $p\left(d_{1:n}, z_{1:n}\right|y_{1:n})$ on-line is to use particle filtering techniques.

4.2.3 Sequential Importance Sampling

According to importance sampling, N particles $\left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}_{i=1}^{N}$ are easily simulated according to an arbitrary convenient importance distribution $\pi(d_{1:n}, z_{1:n}|\, y_{1:n})$ (such that $p\left(d_{1:n}, z_{1:n}\right|y_{1:n}) > 0$ implies $\pi\left(d_{1:n}, z_{1:n}\right|y_{1:n}) > 0$), which is easy to sample from. Then, using the importance sampling identity

$$p\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right) = \frac{w(d_{1:n}, z_{1:n})\pi\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)}{\sum_{(d_{1:n}, z_{1:n})} w(d_{1:n}, z_{1:n})\pi\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)}, \tag{4.19}$$

where the importance weight $w\left(d_{1:n}, z_{1:n}\right)$ is defined as

$$w(d_{1:n}, z_{1:n}) \propto \frac{p\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)}{\pi\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)}, \tag{4.20}$$

one can obtain the estimate of $p\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)$

$$\hat{p}_N\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right) = \textstyle\sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(\left\{d_{1:n}, z_{1:n}\right\} - \left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}\right). \tag{4.21}$$

Here $\tilde{w}_n^{(i)}$ is the normalized version of the importance weight $w\left(d_{1:n}^{(i)}, z_{1:n}^{(i)}\right)$

$$\tilde{w}_n^{(i)} = \frac{w(d_{1:n}^{(i)}, z_{1:n}^{(i)})}{\sum_{j=1}^{N} w(d_{1:n}^{(j)}, z_{1:n}^{(j)})}. \tag{4.22}$$

Analogously,

$$\hat{p}_N\left(d_n, z_n \middle| y_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(\left\{d_n, z_n\right\} - \left\{d_n^{(i)}, z_n^{(i)}\right\}\right). \tag{4.23}$$

If now we want to obtain the estimate of $p\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right)$ on-line, we have to be able to propagate this estimate in time without subsequently modifying the past simulated trajectories $\left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}_{i=1}^{N}$. This means that $\pi(d_{1:n}, z_{1:n} | y_{1:n})$ should admit $\pi\left(d_{1:n-1}, z_{1:n-1} \middle| y_{1:n-1}\right)$ as marginal distribution:

$$\begin{aligned}\pi\left(d_{1:n}, z_{1:n} \middle| y_{1:n}\right) = {} & \pi\left(d_{1:n-1}, z_{1:n-1} \middle| y_{1:n-1}\right) \\ & \times \pi\left(d_n, z_n \middle| d_{1:n-1}, z_{1:n-1}, y_{1:n}\right),\end{aligned}$$

in which case the importance weights w_n can be evaluated recursively, i.e.

$$\begin{aligned} w_n &= w_{n-1} \frac{p\left(y_n \mid y_{1:n-1}, d_{1:n}, z_{1:n}\right) p(d_n \mid d_{n-1}) p(z_n)}{p(y_n \mid y_{1:n-1}) \pi\left(d_n, z_n \mid y_{1:n}, d_{1:n-1}, z_{1:n-1}\right)} \\ &\propto \frac{p\left(y_n \mid y_{1:n-1}, d_{1:n}, z_{1:n}\right) p(d_n \mid d_{n-1}) p(z_n)}{\pi\left(d_n, z_n \mid y_{1:n}, d_{1:n-1}, z_{1:n-1}\right)}. \end{aligned} \tag{4.24}$$

Of course, there is an unlimited number of choices for the importance distribution of such a form, the only restriction being that its support includes that of $p\left(d_{1:n}, z_{1:n} \mid y_{1:n}\right)$. However, in practice, one obtains quite poor performance of the algorithm when $\pi\left(d_{1:n}, z_{1:n} \mid y_{1:n}\right)$ is not well-chosen. Therefore, selection of the importance function is the topic of the following section.

4.2.4 The Choice of the Importance Distribution

A sensible criterion for selection of the importance distribution would be to choose a proposal that minimizes the conditional variance of the importance weights given $d_{1:n-1}, z_{1:n-1}$ and $y_{1:n}$. The importance distribution that satisfies this condition is

$$\pi\left(d_n, z_n \mid d_{1:n-1}, z_{1:n-1}, y_{1:n}\right) = p\left(d_n, z_n \mid d_{1:n-1}, z_{1:n-1}, y_{1:n}\right), \tag{4.25}$$

(see Chapter 3 for details), and this "optimal" importance distribution is employed throughout the chapter.

It can be implemented in the following way. From Bayes' rule $p\left(d_n, z_n \mid d_{1:n-1}, z_{1:n-1}, y_{1:n}\right)$ may be expressed as

$$\begin{aligned} & p\left(d_n, z_n \mid d_{1:n-1}, z_{1:n-1}, y_{1:n}\right) = \\ & \qquad \frac{p\left(y_n \mid y_{1:n-1}, d_{1:n-1}, z_{1:n-1}, d_n, z_n\right) p(d_n \mid d_{n-1}) p(z_n)}{p(y_n \mid y_{1:n-1}, d_{1:n-1}, z_{1:n-1})}, \end{aligned} \tag{4.26}$$

leading to w_n in (4.24) being

$$w_n \propto w_n p\left(y_n \mid y_{1:n-1}, d_{1:n-1}, z_{1:n-1}\right), \tag{4.27}$$

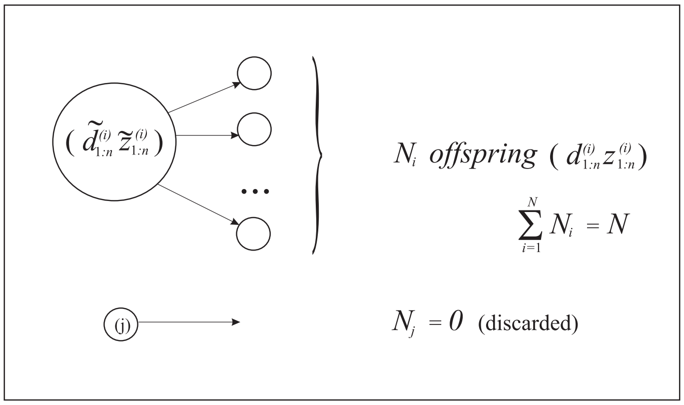

Figure 4.2: Selection step.

with $p(y_n|\, y_{1:n-1}, d_{1:n-1}, z_{1:n-1})$ being equal to

$$\begin{aligned} p(y_n|\, y_{1:n-1}, d_{1:n-1}, z_{1:n-1}) \quad = \quad & \sum_{m=1}^{M} \sum_{\xi=1}^{Z} [p(d_n = m|\, d_{n-1}) p(z_n = \xi) \\ & \times p(y_n|\, y_{1:n-1}, d_{1:n-1}, z_{1:n-1}, d_n = m, z_n = \xi)] \,, \end{aligned}$$

where $p(y_n|\, y_{1:n-1}, d_{1:n-1}, z_{1:n-1}, d_n = m, z_n = \xi)$ is given by the Kalman filter (see Chapter 3).

Thus, sampling from the optimal distribution requires evaluation of the $M \times Z$ one-step ahead Kalman filter steps, which may be computationally intensive if $M \times Z$ is large. In this case, the *prior* distribution $p(d_n, z_n|\, d_{n-1}, z_{n-1})$ could be used as the importance distribution. Then the associated importance weights are proportional to $p(y_n|\, y_{1:n-1}, d_{1:n}, z_{1:n})$, and only one step of the Kalman filter for each particle has to be evaluated.

4.2.5 Selection Step

As discussed in the previous chapter, the variance of the importance weights in the described method can only increase (stochastically) over time, thus resulting in a degeneracy phenomenon. Practically, after a few iterations of the algorithm, all but one of the normalized importance weights are very close to zero. It is, therefore, of crucial importance to include a selection step in the proposed filtering algorithm.

The purpose of the selection step is to discard particles with low normalized importance weights and multiply those with high normalized importance weights. A selection procedure associates with each particle, say $\left\{\widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(i)}\right\}$, a number of children $N_i \in \mathbb{N}$, such that $\sum_{i=1}^{N} N_i = N$, to obtain N new particles $\left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}_{i=1}^{N}$. When $N_i = 0$, $\left\{\widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(i)}\right\}$ is discarded, otherwise it has N_i offspring at time n (Figure 4.2). If the selection procedure is performed at each time step then, before the selection step, we have a weighted distribution $\widetilde{p}_N\left(d_{1:n}\middle| y_{1:n}\right) = \sum_{i=1}^{N} \widetilde{w}_n^{(i)} \delta\left(d_{1:n} - \widetilde{d}_{1:n}^{(i)}\right)$, and afterwards, the approximating distribution follows as $\hat{p}_N\left(d_{1:n}\middle| y_{1:n}\right) = N^{-1} \sum_{i=1}^{N} \delta\left(d_{1:n} - d_{1:n}^{(i)}\right)$.

A number of different selection schemes have been previously proposed in the literature (the list of references is given in Chapter 3). In this work, the selection step is done according to stratified sampling (see Chapter 3 for details), which has the minimum variance one can achieve in the class of unbiased schemes, and can be implemented in $O(N)$ operations.

4.2.6 Particle Filtering Algorithm

Thus, given at time $n-1$, $\left\{d_{1:n}^{(i)}, z_{1:n}^{(i)}\right\}_{i=1}^{N}$ distributed approximately according to $p\left(d_{1:n-1}, z_{1:n-1}\middle| y_{1:n-1}\right)$, at time n the particle filter proceeds as follows.

Particle Filtering Algorithm

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $(\widetilde{d}_n^{(i)}, \widetilde{z}_n^{(i)}) \sim \pi(d_n, z_n | d_{1:n-1}^{(i)}, z_{1:n-1}^{(i)}, y_{1:n})$ and set $\widetilde{d}_{1:n}^{(i)} = (d_{1:n-1}^{(i)}, \widetilde{d}_n^{(i)})$, $\widetilde{z}_{1:n}^{(i)} = (z_{1:n-1}^{(i)}, \widetilde{z}_n^{(i)})$.

- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$w_n^{(i)} \propto \frac{p\left(y_n\middle| y_{1:n-1}, \widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(i)}\right) p(\widetilde{d}_n^{(i)} \middle| \widetilde{d}_{n-1}^{(i)}) p(\widetilde{z}_n^{(i)})}{\pi\left(\widetilde{d}_n^{(i)}, \widetilde{z}_n^{(i)} \middle| \widetilde{d}_{1:n-1}^{(i)}, \widetilde{z}_{1:n-1}^{(i)}, y_{1:n}\right)}.$$

- For $i = 1, \ldots, N$, normalize the importance weights:

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

Selection Step

- Multiply / discard particles $\left\{ \widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(i)} \right\}_{i=1}^{N}$ with respect to high/low normalized importance weights $\tilde{w}_n^{(i)}$ to obtain N particles $\left\{ d_{1:n}^{(i)}, z_{1:n}^{(i)} \right\}_{i=1}^{N}$.

Remark 3 *In the case where the optimal importance distribution is employed, the importance weight $w_n^{(i)}$ does not actually depend on $\widetilde{d}_n^{(i)}, \widetilde{z}_n^{(i)}$. Thus, the selection step should be done prior to the sampling step.*

The computational complexity of this algorithm at each iteration is $O(N)$. Moreover, since both the optimal and prior importance distributions $\pi\left(d_n, z_n \middle| d_{1:n-1}, z_{1:n-1}, y_{1:n} \right)$ and the associated importance weights depend on $d_{1:n-1}, z_{1:n-1}$ via a set of low-dimensional sufficient statistics $\left\{ x_{n|n-1}\left(d_{1:n}^{(i)}, z_{1:n}^{(i)} \right), P_{n|n-1}\left(d_{1:n}^{(i)}, z_{1:n}^{(i)} \right) \right\}_{i=1}^{N}$, only these values need to be kept in memory and, thus, the storage requirements for the proposed algorithm are also $O(N)$ and do not increase over time.

4.2.7 Convergence Results

Let

$$p_N\left(d_{1:n}, z_{1:n} \middle| y_{1:n} \right) = \frac{1}{N} \sum_{i=1}^{N} \delta\left(\left\{ d_{1:n}, z_{1:n} \right\} - \left\{ d_{1:n}^{(i)}, z_{1:n}^{(i)} \right\} \right) \tag{4.28}$$

be the empirical measure of the posterior distribution $p\left(d_{1:n}, z_{1:n} \middle| y_{1:n} \right)$ generated by the particle filtering algorithm. An application of Theorem 1 in [Crisan & Doucet, 2000] proves the convergence (and the rate of convergence) of the average mean square error

$$\mathbb{E}\left[\left(\sum_{d_{1:n}, z_{1:n}} \Upsilon\left(d_{1:n}, z_{1:n} \right) \left(\hat{p}_N\left(d_{1:n}, z_{1:n} \middle| y_{1:n} \right) - p\left(d_{1:n}, z_{1:n} \middle| y_{1:n} \right) \right) \right)^2 \right]$$

to 0 for any bounded function $\Upsilon(d_{1:n}, z_{1:n})$, and, as a result, the point-wise convergence of $p_N(d_{1:n}, z_{1:n}| y_{1:n})$ towards the true posterior distribution $p(d_{1:n}, z_{1:n}| y_{1:n})$. This theorem is an extension of previous results obtained in [Crisan *et al.*, 1999].

Theorem 1 *For all $n \geq 0$, there exists c_n independent of N such that, for any fixed value $(d_{1:n}, z_{1:n}) \in \{1, \ldots, M\}^n \times \{1, \ldots, Z\}^n$, one has*

$$\mathbb{E}\left[\left(\hat{p}_N(d_{1:n}, z_{1:n}| y_{1:n}) - p(d_{1:n}, z_{1:n}| y_{1:n})\right)^2\right] \leq \frac{c_n}{N}$$

where the expectation operator is over all the realizations of the random particle filtering method algorithm. This result proves that, though the particles are interacting, one still keeps the convergence rate of classical Monte Carlo methods.

4.3 Simulations

In the following experiments the bit-error-rate (BER) was evaluated by means of computer simulations[8]. We assumed that the fading does not decorrelate the pulse after matched filtering at the receiver, thus pulse shaping and the matched filter were omitted. The channel was generated from the low pass filtered (3rd order Butterworth filter) zero-mean complex Gaussian noise with the bandwidth of the filter $\mathrm{f_d}T$ times the bit rate different in each example. The number of particles used was $N = 50$ unless otherwise stated. In a number of simulations known pilot symbols were inserted into the data stream at a rate of 1 in $P + 1$ (i.e. $1 : P$).

4.3.1 BER Performance for PSK signals

4.3.1.1 Gaussian additive noise

First, we applied the algorithm proposed above to the case of demodulation of 4-PSK signals ($s_n = A_c \exp(j\theta_n)$, $\theta_n = \frac{2\pi d_n}{M}$) with Gray coding. Computer simulations were performed for different fading bandwidth $\mathrm{f_d}T$,

[8]In each simulation, a symbol sequence was transmitted until at least 100 errors were collected. The results were averaged over 20 runs.

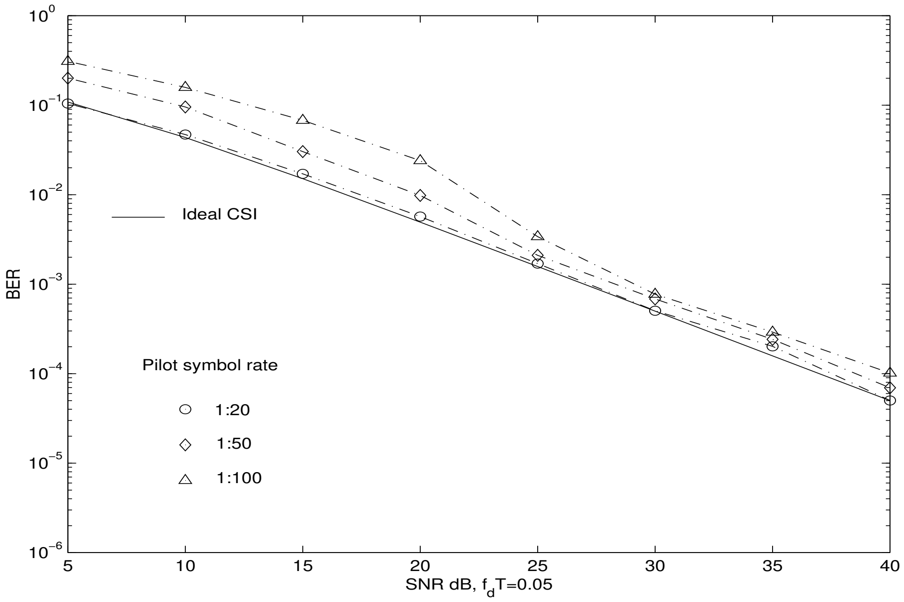

Figure 4.3: Bit error rate for 4PSK signals for different pilot symbol rate (additive Gaussian noise).

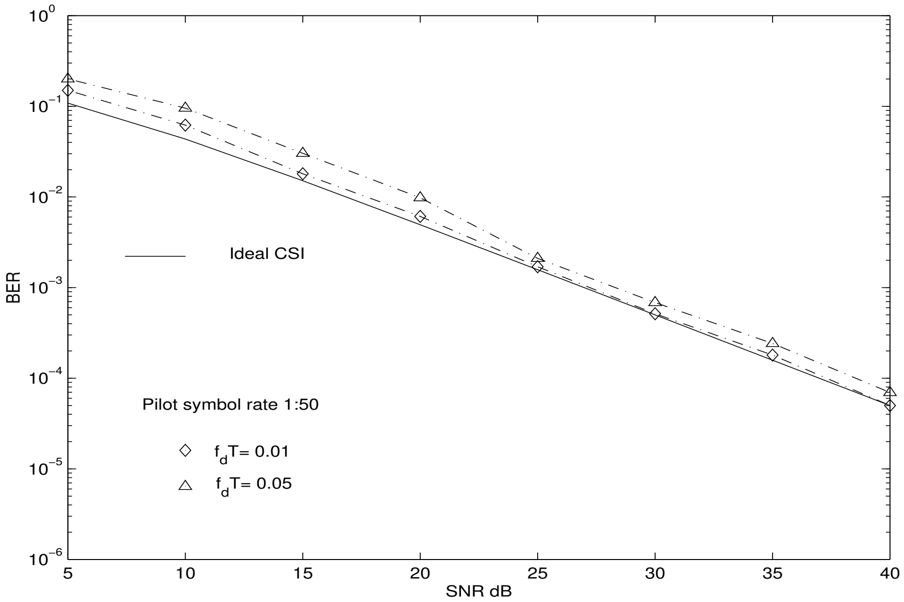

Figure 4.4: Bit error rate for 4PSK signals for different fading characteristics (additive Gaussian noise).

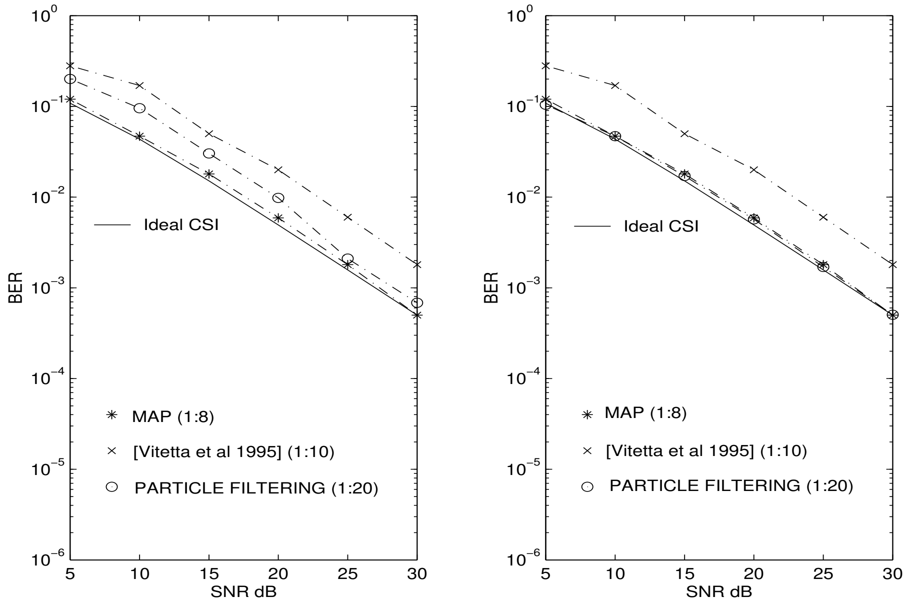

Figure 4.5: Bit error rate for 4PSK signals (additive Gaussian noise).

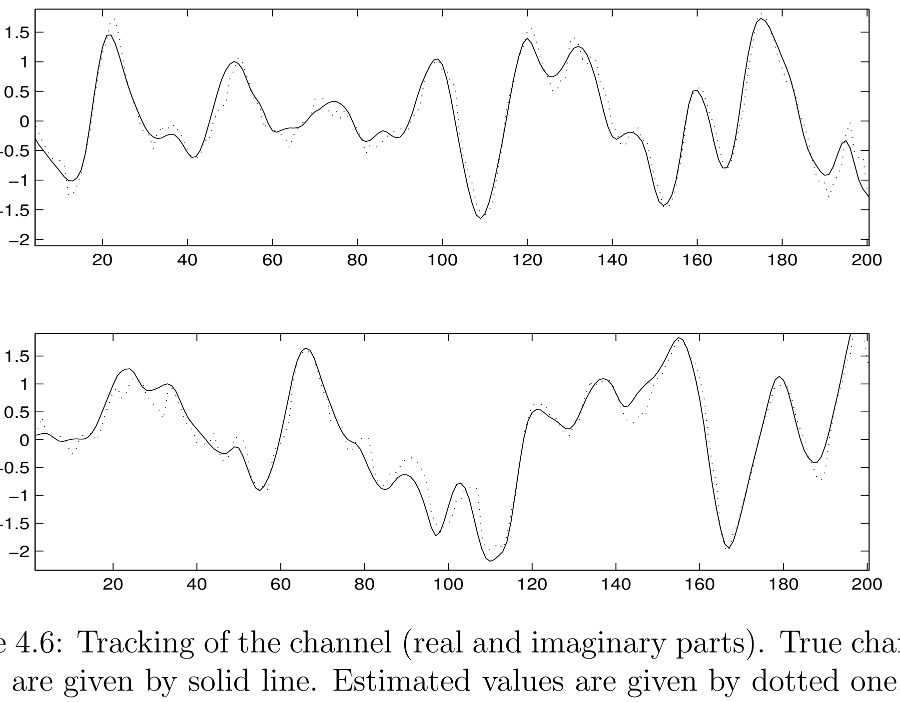

Figure 4.6: Tracking of the channel (real and imaginary parts). True channel values are given by solid line. Estimated values are given by dotted one.

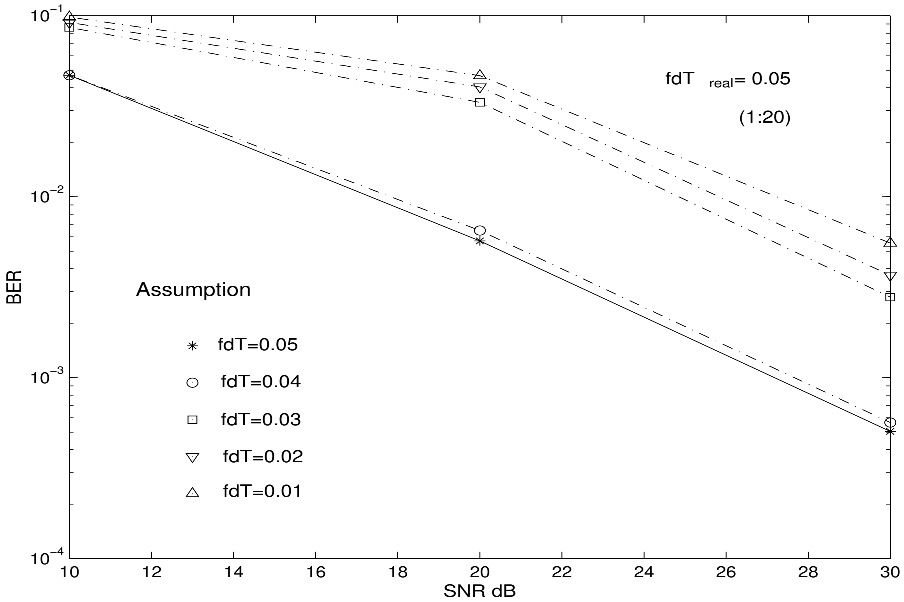

Figure 4.7: Bit error rate for 4PSK signals. $f_dT = 0.05$ with $f_dT = 0.01, 0.02, \ldots, 0.05$ used in the experiments (additive Gaussian noise).

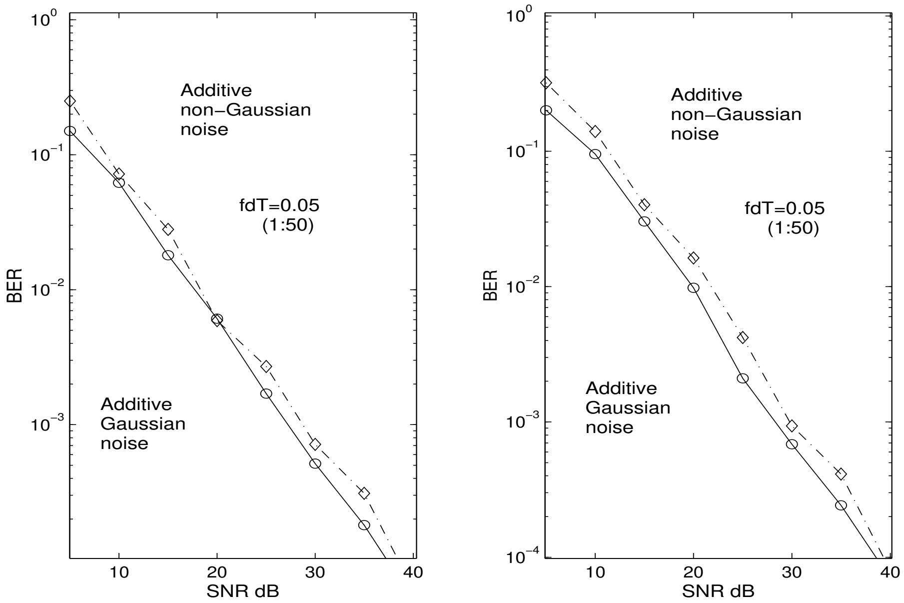

Figure 4.8: Bit error rate for 4PSK signals (additive non-Gaussian noise).

pilot symbol rate $1:P$ and average signal to noise ratio (SNR). Figure 4.3, 4.4 show the obtained BER for the number of particles $N = 50$. A number of experiments have shown that increasing the number of particles does not modify the results significantly (see Table 4.1). The results for $\mathrm{f_d}T = 0.05$ and the pilot symbol rate of $1 : P$ ($P = 20, 50$) compared to those obtained by a Maximum Likelihood receiver employing Per-Survivor Processing (MLPSP) ($P = 10$) [Vitetta & Taylor, 1995], symbol-by-symbol maximum *a posteriori* probability (MAP) receiver ($P = 8$) [Gertsman & Lodge, 1997], and the one with ideal channel state information (CSI), are shown in Figure 4.5. As can be seen, even in the case of fast fading ($\mathrm{f_d}T = 0.05$) and low pilot symbol rate (1 : 50) the proposed algorithm performs well, and it outperforms the existing methods when $P \leq 20$ (compared to $P = 10$ for MLPSP receiver and $P = 8$ for the MAP one). In addition, Figure 4.6 illustrates the tracking abilities of the algorithm for $\mathrm{f_d}T = 0.05$, $P = 20$ and SNR=20 dB by showing the true channel values and estimated channel values in real and imaginary format. Finally, the robustness of the algorithm was studied by assuming the fading rate $\mathrm{f_d}T$ different from the one which was in reality. As one can see in Figure 4.7, there is no appreciable difference in the BER performance when the fading rate $\mathrm{f_d}T = 0.04$ is used (for $\mathrm{f_d}T_{\mathrm{real}} = 0.05$). However, the degradation in the case of $\mathrm{f_d}T \leq 0.03$ is quite significant since the channel coefficient estimates are not able to track the actual channel states as accurately as in the case of faster fading assumed.

	SNR, dB	$N = 50$	$N = 200$	$N = 500$
BER	10	4.69×10^{-2}	4.69×10^{-2}	4.67×10^{-2}
BER	20	5.70×10^{-3}	5.68×10^{-3}	5.61×10^{-3}
BER	30	5.05×10^{-4}	4.93×10^{-4}	4.89×10^{-4}

Table 4.1: Bit error rate for 4PSK signals for different number of particles N. Pilot symbol rate is 1:20 and f_dT=0.05 (additive Gaussian noise).

4.3.1.2 Non-Gaussian additive noise

In the second experiment we applied the proposed algorithm to the case when the additive noise is distributed as a two-component mixture of zero-mean Gaussians ($Z = 2$). The overall variance of the noise in this case is $\lambda_1\sigma_1^2+(1-\lambda_1)\sigma_2^2$ with $\lambda_1 = 0.2$. σ_1^2 was chosen so that the average SNR would be equal to 5 dB if $\lambda_1 = 1$. The characteristics of the signal and channel are the same as for the first experiment, and the results for the pilot symbol rate 1:50 and $\mathrm{f_d}T = 0.05$ are shown in Figure 4.8.

4.3.2 BER Performance for DPSK signals

An important consideration in dealing with PSK signals is the problems associated with an M-ary phase ambiguity. One way to overcome these problems is to introduce pilot symbols into the data stream (see Subsection 4.3.1). Another common method to solve them is to differentially encode the information before transmission, in which case the information is carried in the difference between adjacent received phases. Differential detection can be then applied. Such differentially encoded PSK (DPSK) signals are the subject of this subsection.

As a practical example we investigate 4-DPSK signals: $s_n = \exp(j\theta_n)$, $\theta_n = \sum_{j=1}^{n}\sum_{m=1}^{M}\frac{2\pi m}{M}\delta\left(d_j - d_m\right)$ (Gray coding has been employed). The channel fading rate was $\mathrm{f_d}T = 0.05$, and Figure 4.9 shows the results for different signal to noise ratio, compared to those obtained by *a posteriori* probability (APP) demodulator [Hoeher & Lodge, 1999] ($L_s = 2$ for the fixed-lag smoothing here and in the following experiments).

Then we considered 8-DPSK signal examined previously in [Collings & Moore, 1995]. It is assumed that a coding scheme was employed in transmission, which has the following signal properties: $p_{ii} = 0.95$, $p_{ij} = (1 - p_{ii})/(M-1)$ for $i \neq j$, $M = 8$ (such strong dependence from one message symbol to the next one may occur in some convolutional codes or if oversampling is used). In Figure 4.10, 4.11, the results for different $\mathrm{f_d}T$ are compared to those obtained by a hidden Markov model (HMM) and a matched filter/automatic gain controller/phase locked loop (MF/AGC/PLL) schemes [Collings & Moore, 1995].

We also applied the proposed algorithm to 8-DPSK signals when the additive noise is distributed as a two-component mixture of Gaussians with

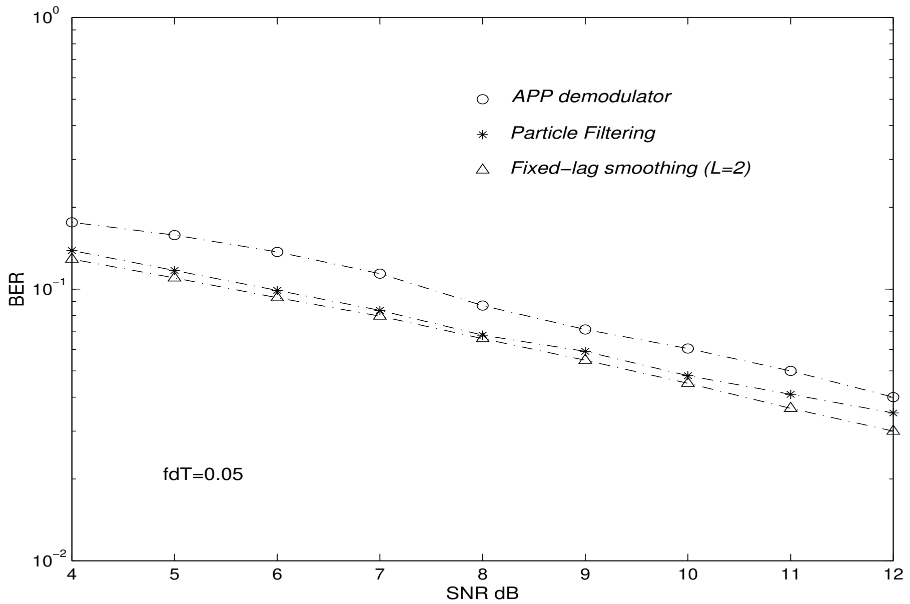

Figure 4.9: Bit error rate for 4-DPSK signals (i.i.d. message symbols).

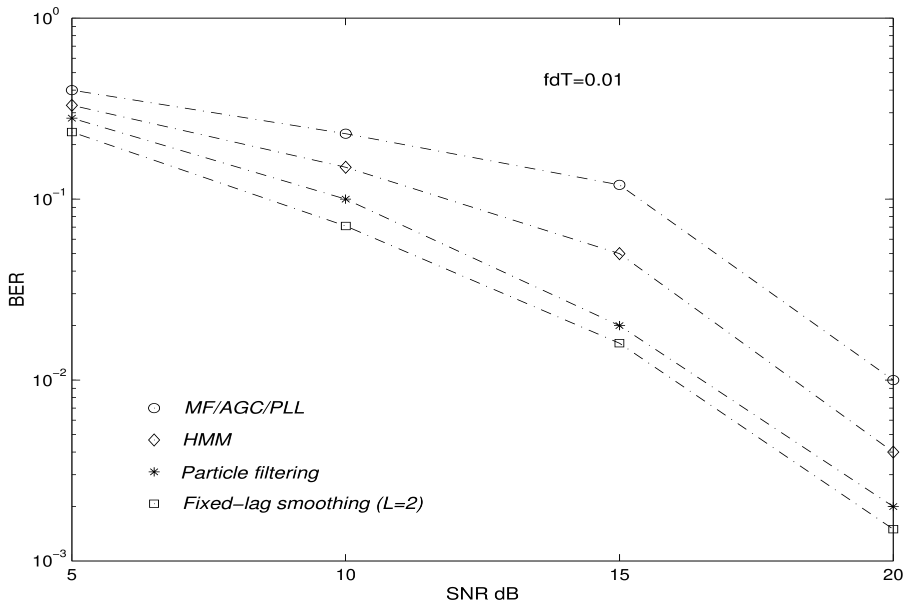

Figure 4.10: Bit error rate for 8-DPSK signals (non-i.i.d. message symbols).

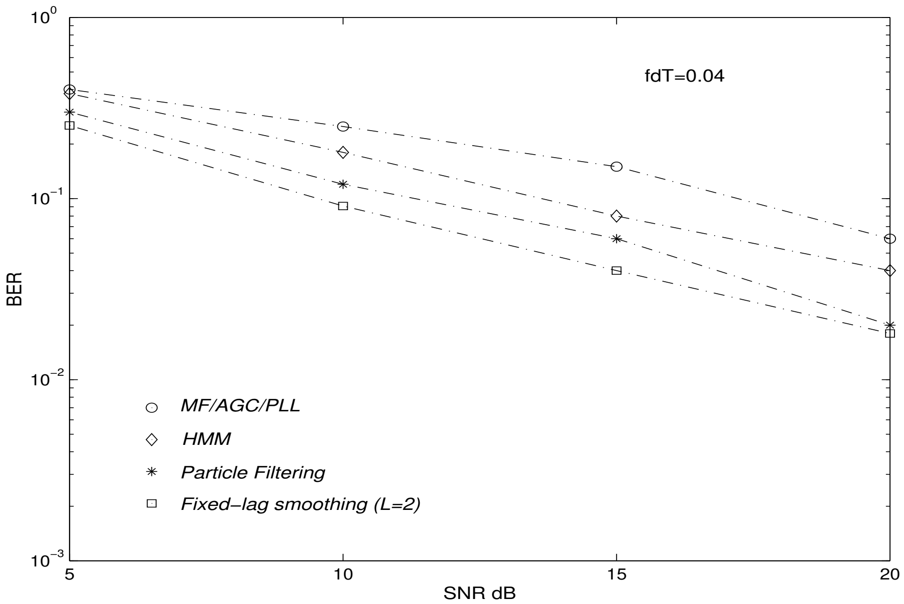

Figure 4.11: Bit error rate for 8-DPSK signals (non-i.i.d. message symbols).

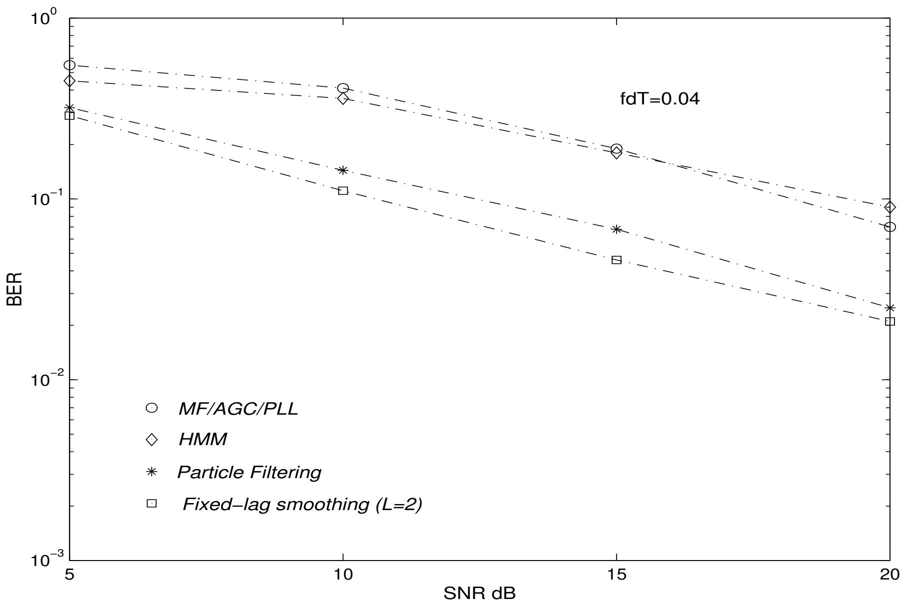

Figure 4.12: Bit error rate for 8-DPSK signals (additive non-Gaussian noise).

the same parameters as in Subsection 4.3.1. The results for $\mathrm{f_d}T = 0.04$ are presented in Figure 4.12.

4.3.3 BER Performance for QAM signals

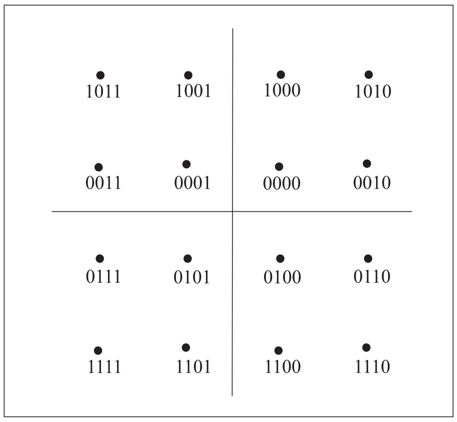

Figure 4.13: 16QAM with Gray coding.

Finally, the performance of the particle filtering demodulation of QAM signals is studied. In the case of QAM signal, mapping function $s_n = \mu_n$, where at any time n, the complex elements $\mu_n = \mu_n^R + j\mu_n^I$, associated with different strings of bits, are chosen so as to generate a rectangular grid of equally spaced points in the complex plane space $\mathbb{C}$, which is illustrated in Figure 4.13 for $M = 16$. The message is then modulated and transmitted on two quadrature carriers $\cos(2\pi \mathrm{f_{car}}\tau)$ and $\sin(2\pi \mathrm{f_{car}}\tau)$ as a bandpass signal, thus resulting in a higher spectral efficiency.

In the following simulations $M = 16$ was assumed, i.e. 16 QAM (with Gray coding) was studied. A pilot sequence was used in all simulations, and the results for different pilot symbol rates for the case of fast fading ($\mathrm{f_d}T = 0.05$) are shown in Figure 4.14. In Figure 4.15 the results for different fading bandwidths ($\mathrm{f_d}T = 0, 01, 0.02$ and 0.05) for $P = 20$ are presented. The performance of the proposed method is compared to the one of the ideal pilot-symbol-assisted modulation (PSAM) and MAP Λ-lag decision feedback estimator (DFE) ($\Lambda = 7$, pilot symbol spacing is 5) in Figure 4.16. Comparison with the ideal CSI case is also presented. Similarly to PSK demodulation,

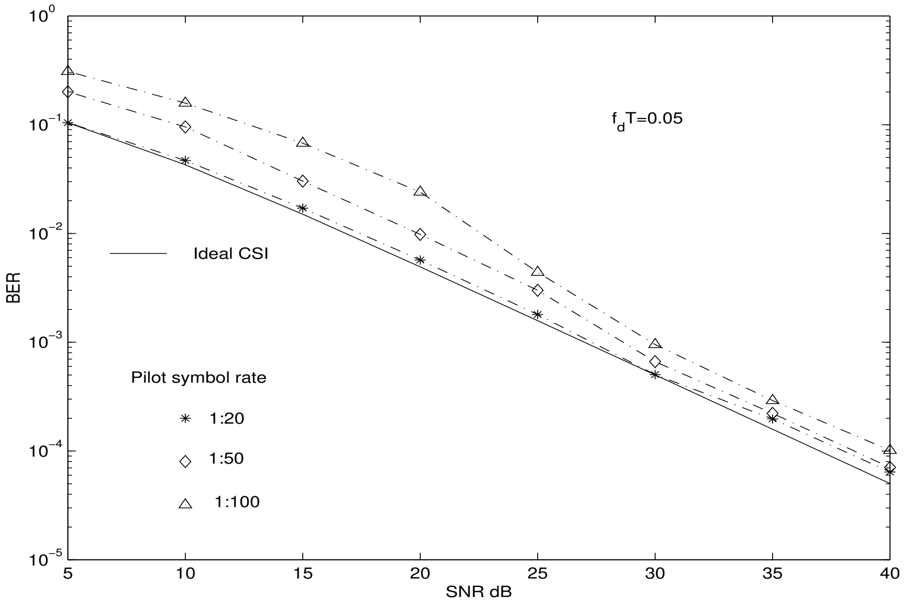

Figure 4.14: Bit error rate for 16QAM signals for different pilot symbol rate (additive Gaussian noise).

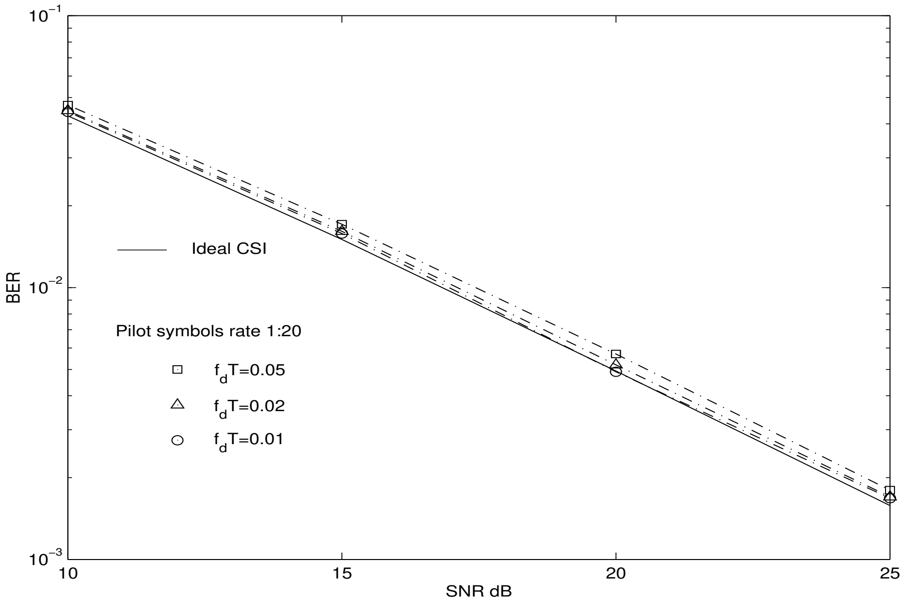

Figure 4.15: Bit error rate for 16QAM signals for different fading characteristics (additive Gaussian noise).

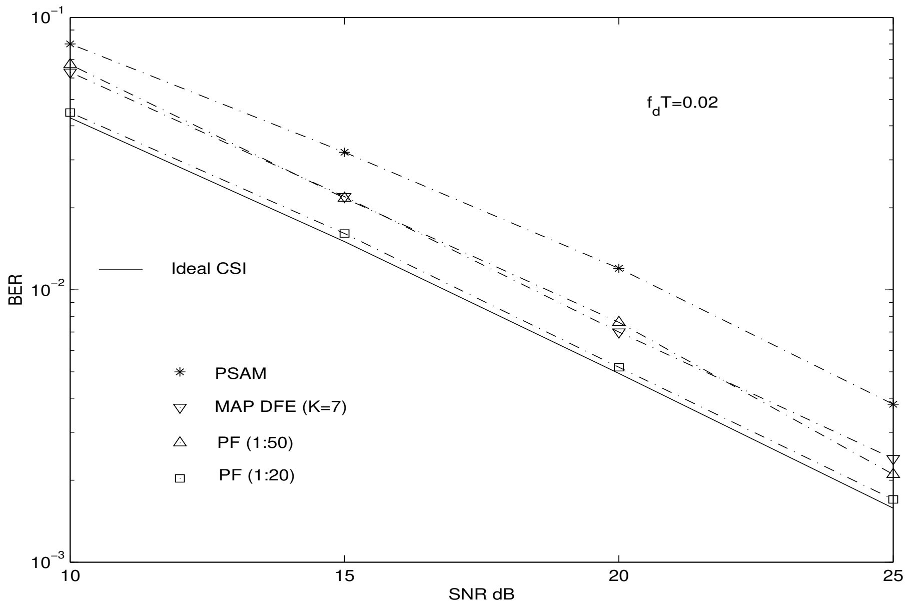

Figure 4.16: Bit error rate for 16QAM signals (additive Gaussian noise).

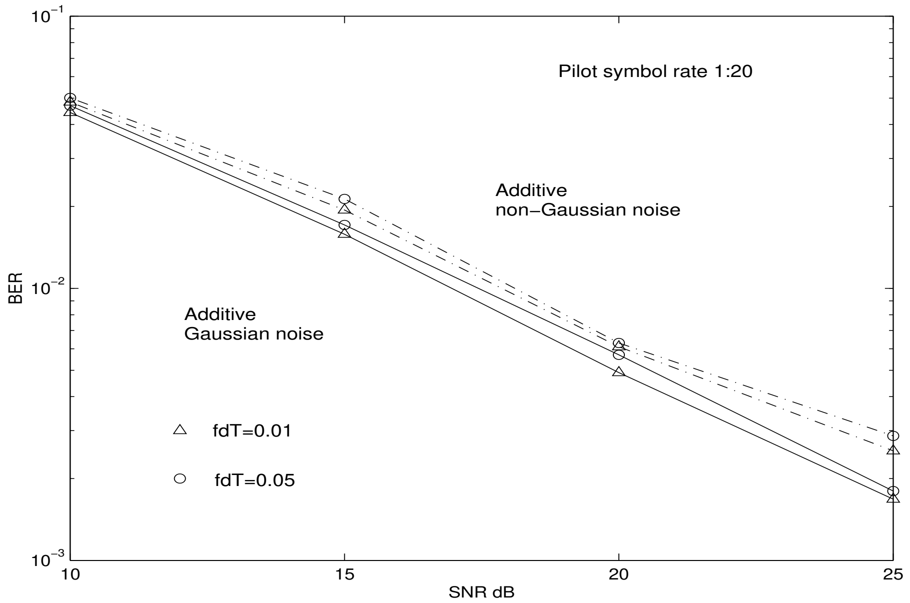

Figure 4.17: Bit error rate for 16QAM signals (additive non-Gaussian noise).

the number of particles was $N = 50$, and little or no improvement was gained by increasing this number. Finally, Figure 4.17 shows the results for non-Gaussian additive noise with the same characteristics as in Subsection 4.3.1, for $\mathrm{f_d}T = 0.02$ and $P = 20$.

4.4 Discussion

It can be seen that the proposed algorithm outperforms the existing methods simulated in the comparison in the case of additive Gaussian noise for all modulation schemes examined above. In particular, a number of points can be noted here:

- There is a really small (or even virtually no) performance degradation compared to that of the receiver with ideal CSI even in the case of fast fading (for $P \leq 20$).
- The number of pilot symbols inserted (in the absence of differential encoding) is significantly reduced in comparison with other methods. Furthermore, additional simulations have indicated little if any improvement is gained by decreasing a pilot symbol spacing.
- The number of particles used is reasonably small. Computer simulations for different modulation schemes also showed that increasing this number does not modify the results sensibly.
- Extensive study of PSK signals demonstrated, among others, good tracking ability of the algorithm.
- Simulations for 16 QAM showed that demodulation of higher order modulations (with high spectral efficiency) can be successfully performed.
- Lastly, the BER results for DPSK proved that differentially encoded signals can be treated efficiently as well.

In addition, the algorithm exhibits good performance in the case of non-i.i.d. message symbols, and the degradation of the performance in the non-

Gaussian noise case is quite small whereas other standard methods are not actually designed to treat this case.

The efficient particle filtering algorithm that we have developed for demodulation under flat fading conditions have proved to perform well in comparison with currently existing methods. We will now try to extend it to address more complicated problems of joint demodulation and decoding, space diversity and multiuser detection.

5 Extended particle filtering receiver

The demand for greater capacity and enhanced performance of the transmission in modern communication systems is becoming more and more prevalent with the ever growing number of users and the spectrum of services provided. One of the obstacles that must be removed in order to satisfy these requirements is signal degradation due to Rayleigh fading. To meet this goal, a modern digital communication system is equipped with a channel encoder and a diversity receiver. An increased system capacity is typically achieved by employing multiuser transmission techniques. The question is now, how to detect the transmitted information-bearing signals efficiently under such complicated conditions.

A particle filtering demodulator developed in the previous chapter shows great promise in coping with this task, and could be combined with other existing schemes to perform symbol demodulation and detection. It is, however, possible to address these problems jointly in the particle filtering framework, and, hence, design a generic particle filtering receiver. Similar to demodulation scenario, the transmission model in all these cases facilitates a state-space representation (conditional upon the symbols), and an efficient particle filtering algorithm can be proposed that makes use of this structure and incorporates variance reduction strategies. The algorithm is a straightforward

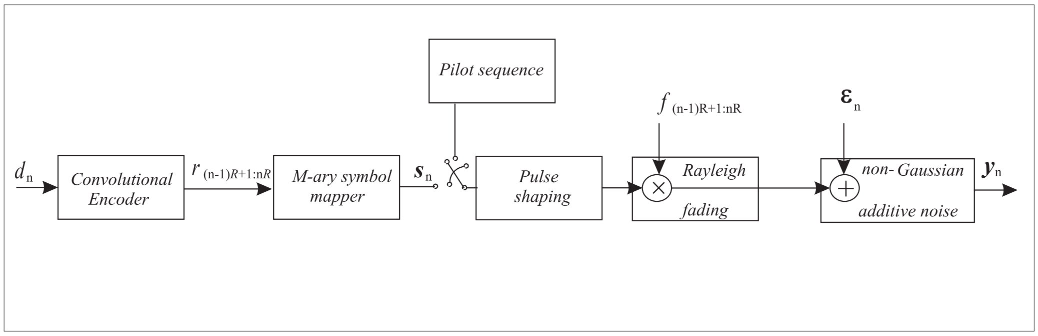

Figure 5.1: Transmission of convolutionally coded M-ary modulated signals in Rayleigh fading channels.

extension of the method described up to now, has a computational complexity that is linear in the number of particles, and can be easily implemented on parallel processors.

The remainder of the chapter is organized as follows. Section 5.1 extends the receiver of the previous chapter to consider convolutionally coded signals, a diversity particle filtering receiver is presented in Section 5.2, and in Section 5.3 multiuser detection for synchronous code division multiple access (CDMA) schemes is considered. Simulation results are presented at the end of each section, and are discussed at the end of the chapter in Section 5.4.

5.1 Joint Demodulation and Decoding

As already mentioned, the particle filtering algorithm presented in the previous chapter can be easily extended to the case when the forward error correcting (FEC) coding is employed. The purpose of FEC is to improve the capacity of a channel by adding some carefully designed redundant information to the data being transmitted through it so that the receiver is able to correct a certain number of errors. The process of adding this redundant information is also known as channel coding, and a distinction is made between two code families suitable for FEC, namely the linear block codes and convolutional codes. The later is the subject of this section[9].

[9]In principle, a decoder structure for an interleaved coded system can be designed. However, a direct extension of the proposed method will lead to quite a complicated model

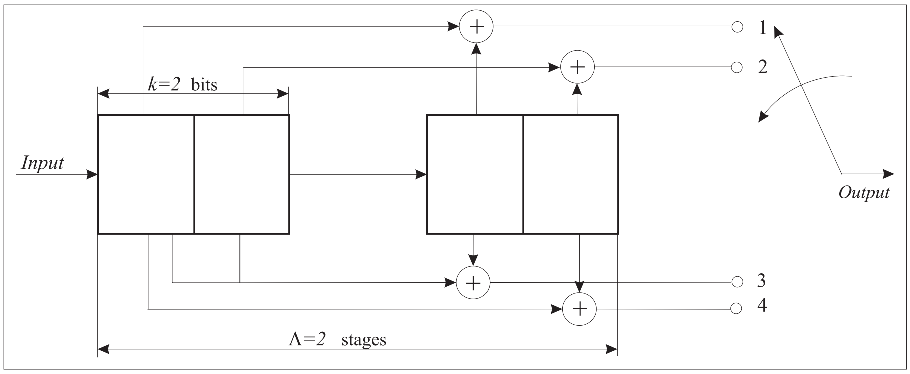

Figure 5.2: Nonbinary 4-state (4,2,2) convolutional encoder.

5.1.1 Model Specification and Estimation Objectives

Let us consider a nonbinary M-state convolutional code $(N_{code}, \kappa, \Lambda)$. The code has one of $M = 2^{\kappa}$ possible input and R output κ-bit symbols, resulting in a code ratio of $1/R$. It is generated by passing the original sequence through a linear shift register, which consists of Λ (κ-bit) stages and $N_{code} = kR$ linear algebraic function generators. (A schematic structure of a simple nonbinary 4-state, rate 1/2 convolutional encoder is illustrated in Figure 5.2.) The output of the encoder is then treated by the modulator and transmitted through the communication channel by means of some M-ary modulation technique (see Figure 5.1).

Let us recall from Chapter 2 that d_n is an input and $r_{(n-1)R+1:nR}$ is the corresponding output of the encoder[10], where d_n, r_q indicate one of the M possible κ-bit sequences ($d_n \in \{1, 2, \ldots, M\}$, $r_q \in \{1, 2, \ldots, M\}$ for $n = 1, 2, \ldots$ and $q = 1, 2, \ldots$). Each symbol r_q at the output of the encoder is mapped into a corresponding waveform $s_q(r_q)$ transmitted at times $q = (n-1)R + \chi$, $\chi = 1, \ldots, R$. (This is illustrated in Figure 5.3 for $\kappa = 2$ and $R = 2$.) The block of the code symbols $r_{(n-1)R+1:nR}$ is determined by the

and a large decoding delay when the size of the block interleaver matrix is relatively large. Thus, an alternative method should be proposed, which is the subject of our current research.

[10] Similar to the previous chapter, the indexes k, l and g are suppressed in this section for clarity of presentation.

input symbols $d_{n-\Lambda-1:n}$,

$$r_{(n-1)R+1:nR} = \Phi(d_{n-\Lambda-1:n}), \tag{5.1}$$

where $\Phi(\cdot)$ depends on the structure of the encoder. The block of corresponding waveforms $\mathbf{s}_n \triangleq \left(s_{(n-1)R+1}, s_{(n-1)R+2} \ldots, s_{nR}\right)^{\mathsf{T}}$ with the components $s_{(n-1)+\chi} = s(r_{(n-1)+\chi})$ is then a function of $d_{n-\Lambda-1:n}$, and the output of the channel is given by

$$y|_{t=(n-1)R+\chi} = s_{(n-1)R+\chi}\left(d_{n-\Lambda-1:n}\right) f_{(n-1)R+\chi} + \sigma_{z_t}\epsilon_t, \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0,1\right). \tag{5.2}$$

Here, the channel disturbance f_t is modelled as an ARMA (ρ, ρ) process with coefficients $\mathbf{a}$ (AR part) and $\mathbf{b}$ (MA part), and, as described in the previous chapter, is represented through the parameter x_t defined as

$$f_t = \mathbf{b}^{\mathsf{T}} x_{t:t-\rho+1}, \tag{5.3}$$

$\sigma^2_{z_t}$ in Equation (5.2) is the variance of the Gaussian distribution corresponding to the latent allocation variable $z_t \in \mathcal{Z} = \{1, 2, \ldots, Z\}$, $t = 1, 2, \ldots$, and assumed to be i.i.d.

The message symbols d_n which are also assumed to be i.i.d., the channel characteristics $x_{nR:(n-1)R-\rho+2}$ and the latent variables $\mathbf{z}_n \triangleq z_{(n-1)R+1:nR}$, corresponding to this nth symbol are unknown for $n > 0$, whereas the model parameters $\mathbf{a}$, $\mathbf{b}$, $\sigma^2_{z_t}$ are known for each d_n, $\mathbf{z}_n$.

The objective: to obtain the MMAP estimates of the symbols from the corresponding observations $\mathbf{y}_n \triangleq y_{(n-1)R+1:nR}$:

$$\hat{d}_n = \arg\max_{d_n} p\left(d_n | \, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\right). \tag{5.4}$$

5.1.2 Joint Demodulation and Decoding Algorithm

We will now extend the particle filtering algorithm presented in the previous chapter to address the problem of joint demodulation and decoding of information sequences. In this case, recalling that d_n and $\mathbf{z}_n$ are i.i.d.

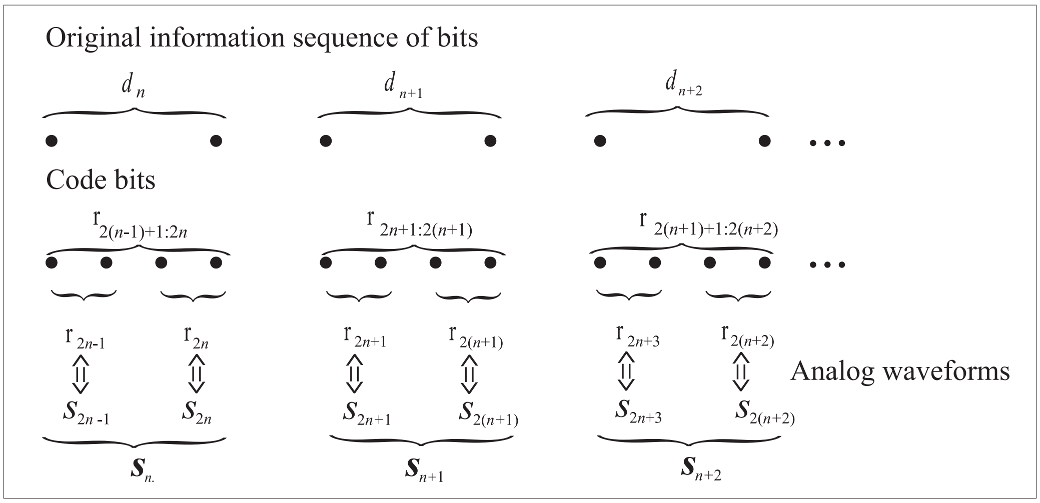

Figure 5.3: Nonbinary 4-state rate 1/2 coding and 4-ary modulation.

sequences, the optimal importance distribution $p\left(d_n, \mathbf{z}_n \middle| d_{1:n-1}, \mathbf{z}_{1:n-1}, \mathbf{y}_{1:n}\right)$, may be expressed in the following form[11]

$$p\left(d_n, \mathbf{z}_n \middle| d_{1:n-1}, \mathbf{z}_{1:n-1}, \mathbf{y}_{1:n}\right) = \frac{p\left(\mathbf{y}_n \middle| d_{1:n-1}, \mathbf{z}_{1:n-1}, \mathbf{y}_{1:n-1}, d_n, \mathbf{z}_n\right) p(d_n) p(\mathbf{z}_n)}{p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, d_{1:n-1}, \mathbf{z}_{1:n-1})} \tag{5.5}$$

leading to the weights w_n being equal to

$$w_n \propto p(\mathbf{y}_n | d_{1:n-1}, \mathbf{z}_{1:n-1}), \tag{5.6}$$

where $p(\mathbf{y}_n | d_{1:n-1}, \mathbf{z}_{1:n-1})$ could be calculated as follows:

$$p(\mathbf{y}_n | d_{1:n-1}, \mathbf{z}_{1:n-1}) = \sum_{m=1}^{M} \sum_{\xi_1=1}^{Z} \cdots \sum_{\xi_R=1}^{Z} \Big[p(d_n = m) p(\mathbf{z}_n = \xi_{1:R}) \times$$
$$p(\mathbf{y}_n | d_{1:n-1}, \mathbf{z}_{1:n-1}, d_n = m, \mathbf{z}_n = \xi_{1:R}) \Big]$$

with $\xi_{1:R} = \{\xi_1, \xi_2, \ldots \xi_R\}$.

Taking into account that, conditional on $d_{1:n}$, $\mathbf{z}_{1:n}$ (and, consequently, on $r_{1:nR}$ and $z_{1:nR}$), the measurement y_q (arranged into vector $\mathbf{y}_n$ for the nth

[11]We use the following notation $\mathbf{y}_{1:n} \triangleq (\mathbf{y}_1, \mathbf{y}_2 \ldots, \mathbf{y}_n)^{\mathrm{T}}$ and $\mathbf{z}_{1:n} \triangleq (\mathbf{z}_1, \mathbf{z}_2 \ldots, \mathbf{z}_n)^{\mathrm{T}}$.

symbol, i.e. $\mathbf{y}_n \triangleq y_{(n-1)R+1:nR}$) are independent of the states at all other times, one obtains for the ith particle

$$w_n^{(i)} \propto \sum_{m=1}^{M} \Bigg[p(d_n = m) \sum_{\xi_1=1}^{Z} \cdots \sum_{\xi_R=1}^{Z} \prod_{\chi=1}^{R} p(z_\chi = \xi_\chi) \times \\ p\left(y_{(n-1)R+\chi} \middle| \{r_{1:(n-1)R}^{(i)}, \rho_{n,1:\chi}^{(i,m)}\}, \{z_{1:(n-1)R}^{(i)}, \xi_{1:\chi}\} \right) \Bigg]$$

where the code sequence $\rho_{n,1:R}^{(i,m)}$ is equal to $\rho_{n,1:R}^{(i,m)} = \Phi(d_n = m, d_{n-1:n-\Lambda-1}^{(i)})$ (see Equation (5.1)) and $p\left(y_{(n-1)R+\chi} \middle| \{r_{1:(n-1)R}^{(i)}, \rho_{n,1:\chi}^{(i,m)}\}, \{z_{1:(n-1)R}^{(i)}, \xi_{1:\chi}\} \right)$ is given by the Kalman filter (see Chapter 3).

Thus, the importance weight $w_n^{(i)}$ can be evaluated up to a normalizing constant according to the following algorithm, where particles $\left\{ d_{1:n-1}^{(i)}, \mathbf{z}_{1:n-1}^{(i)} \right\}_{i=1}^{N}$ are distributed approximately according to $p\left(d_{1:n-1}, \mathbf{z}_{1:n-1} \middle| \mathbf{y}_{1:n-1} \right)$.

Evaluation of the importance weight

(optimal importance distribution)

- For $m = 1, \ldots, M$, obtain coded symbols $\rho_{n,1:R}^{(i,m)}$

$$\rho_{n,1:R}^{(i,m)} = \Phi(d_n = m, d_{n-1:n-\Lambda-1}^{(i)}).$$

- For $m = 1, \ldots, M$,

$$w_n^{(i,m,\xi_0)} = p(d_n = m),$$

  - For $\chi = 1, \ldots, R$,
    - For $\xi_1 = 1, \ldots, Z$,
      - ...
      - For $\xi_\chi = 1, \ldots, Z$,
        perform one-step Kalman update and evaluate

$$w_n^{(i,m,\xi_0,\xi_1,\ldots,\xi_\chi)} = w_n^{(i,m,\xi_0,\xi_1,\ldots,\xi_{\chi-1})} \times$$
$$p\left(y_{(n-1)R+\chi} \middle| \{d_{1:n-1}^{(i)}, \rho_{n,1:\chi}^{(i,m)}\}, \{\mathbf{z}_{1:n-1}^{(i)}, \xi_{1:\chi}\}\right) p(\xi_\chi)$$

- Evaluate the importance weight $w_n^{(i)}$ up to a normalizing constant

$$w_n^{(i)} = \sum_{m=1}^{M} \sum_{\xi_1=1}^{Z} \ldots \sum_{\xi_1=1}^{Z} w_n^{(i,m,\xi_0,\xi_1,\ldots,\xi_R)}.$$

Thus, sampling from the optimal distribution requires evaluation of the $M \times (Z + Z^2 + \ldots Z^R) = MZ(Z^R - 1)/(Z-1)$ one-step Kalman filter updates. As it was mentioned previously, if $MZ(Z^R - 1)/(Z - 1)$ is large, the use of the *prior* distribution as the importance distribution is recommended due to computational complexity. The associated importance weights in this case are proportional to $p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, d_{1:n}, \mathbf{z}_{1:n})$, and only R steps of the Kalman filter for each particle have to be evaluated.

In the general case, given symbol particles $\left\{ d_{1:n-1}^{(i)}, \mathbf{z}_{1:n-1}^{(i)} \right\}_{i=1}^{N}$ distributed approximately according to $p\left(d_{1:n-1}, \mathbf{z}_{1:n-1} | \mathbf{y}_{1:n-1}\right)$, the particle filter proceeds as follows.

Particle Filtering Algorithm for joint demodulation and decoding

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $(\tilde{d}_n^{(i)}, \tilde{\mathbf{z}}_n^{(i)}) \sim \pi(d_n, \mathbf{z}_n | d_{1:n-1}^{(i)}, \mathbf{z}_{1:n-1}^{(i)}, \mathbf{y}_{1:n})$ and set $\tilde{d}_{1:n}^{(i)} = (d_{1:n-1}^{(i)}, \tilde{d}_n^{(i)})$, $\tilde{\mathbf{z}}_{1:n}^{(i)} = (\mathbf{z}_{1:n-1}^{(i)}, \tilde{\mathbf{z}}_n^{(i)})$.

- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$w_n^{(i)} \propto \frac{p\left(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \tilde{d}_{1:n}^{(i)}, \tilde{\mathbf{z}}_{1:n}^{(i)}\right) p(\tilde{d}_n^{(i)}) p(\tilde{\mathbf{z}}_n^{(i)})}{\pi\left(\tilde{d}_n^{(i)}, \tilde{\mathbf{z}}_n^{(i)} \middle| \tilde{d}_{1:n-1}^{(i)}, \tilde{\mathbf{z}}_{1:n-1}^{(i)}, \mathbf{y}_{1:n}\right)}.$$

- For $i = 1, \ldots, N$, normalize the importance weights:

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

<u>*Selection Step*</u>

- Multiply / discard particles $\left\{\widetilde{d}_{1:n}^{(i)}, \widetilde{\mathbf{z}}_{1:n}^{(i)}\right\}_{i=1}^{N}$ with respect to high/low normalized importance weights $\tilde{w}_n^{(i)}$ to obtain N particles $\left\{d_{1:n}^{(i)}, \mathbf{z}_{1:n}^{(i)}\right\}_{i=1}^{N}$.

If the optimal importance distribution is employed, the importance weight $w_n^{(i)}$ does not depend on $\widetilde{d}_{1:n}^{(i)}, \widetilde{\mathbf{z}}_{1:n}^{(i)}$, and the weights evaluation and selection can be done prior to the sampling step. The computational complexity and memory requirements of the algorithm in this case are approximately equivalent to the ones of $N \times MZ(Z^R - 1)/(Z - 1)$ Kalman filters associated to the model equations. Indeed, sampling from $p\left(d_n, \mathbf{z}_n \middle| d_{1:n-1}, \mathbf{z}_{1:n-1}, \mathbf{y}_{1:n}\right)$ and evaluating w_n require $MZ(Z^R - 1)/(Z - 1)$ one step Kalman filters associated to the N particles $\left(d_{1:n-1}^{(i)}, \mathbf{z}_{1:n-1}^{(i)}\right)$. In the case of the prior distribution used as the importance distribution, only $N \times R$ Kalman filters are required. Moreover, in both cases $p\left(d_n, \mathbf{z}_n \middle| d_{1:n-1}, \mathbf{z}_{1:n-1}, \mathbf{y}_{1:n}\right)$ and w_n only depend on $(d_{1:n-1}, \mathbf{z}_{1:n-1})$ via the mean and covariance of the channel state $\mathbf{x}_n$ conditional upon $(d_{1:n-1}, \mathbf{z}_{1:n-1})$, therefore, one only needs to keep in memory these statistics and the storage requirements do not increase over time as might appear at first sight.

5.1.3 BER Performance for the coded PSK signals

Simulations were carried out to determine the performance of the proposed receiver with coded PSK signals[12]. The code utilized was the standard 8PSK (with Gray encoding) 4-state code described in [Chan & Bateman, 1992, p.191, Fig. 2]. Different channel Doppler frequencies and average signal-to-noise ratios were considered. Figure 5.4 and 5.5 show the results for $f_d T = 0.01/0.05$ (with and without pilot symbols) compared to those

[12] For a more detailed description of the simulation set-up see Section 4.3.

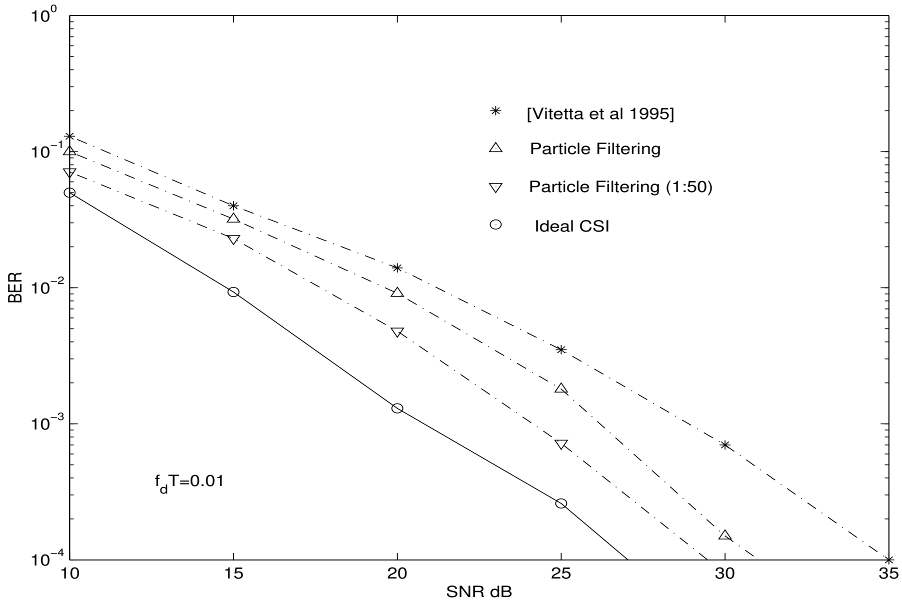

Figure 5.4: Bit error rate for coded 8PSK signals (additive Gaussian noise).

obtained by a maximum likelihood receiver employing per-survivor processing (MLPSP) (pilot sequence is not used) [Vitetta & Taylor, 1995], and to the ideal CSI case. Whenever the pilot symbols were employed, the rate was $1 : P$ with $P = 50$. The number of particles used in the algorithm was equal to $N = 50$. The results for non-Gaussian additive noise are presented in Figure 5.6.

5.2 Diversity Receiver for Fading Channels

The performance of the receiver in fading channels can sometimes be improved significantly through the use of diversity combining techniques. These techniques are based on the fact that the majority of errors under fading conditions occur when the channel attenuation is large. Hence, if the receiver can be supplied with several independently fading replicas of the same transmitted signal, the probability that all received signals will experience deep fade simultaneously is considerably reduced. The optimal combining of the received signals in this case will yield vastly improved performance.

There are several common ways in which such independently fading repli-

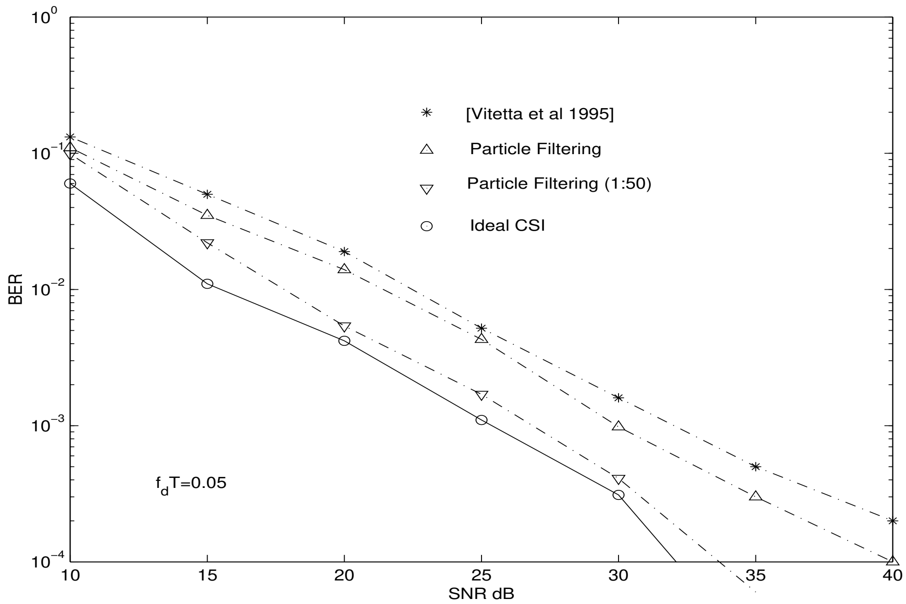

Figure 5.5: Bit error rate for coded 8PSK signals (additive Gaussian noise).

cas of the signal can be provided such as *frequency* (the signal is transmitted on several separated carriers), *time* (the signal is transmitted in several different time slots) and *space* (several receiving antennas are employed) *diversity.* This section considers the later form, which is discussed in details below.

5.2.1 Diversity Model and Estimation Objectives

Let us consider a diversity receiver with L antennas spaced sufficiently far apart to ensure different propagation paths of the transmitted signal (see Figure 5.7). In this case, it is reasonable to assume that the fading processes among the L diversity channels are mutually independent. Furthermore, the channels are assumed to be frequency non-selective (flat) slowly fading with Rayleigh-distributed envelop statistics modelled by a complex ARMA model of order (ρ, ρ) (Subsection 4.1.1.3). The transmitted information-bearing signal $s_{\text{trans}}(\tau)$ (see Subsection 4.1.1.1) is the same for all the channels ($K = 1$, k is subsequently suppressed) and in each channel is corrupted by an additive non-Gaussian noise with characteristics given in Subsection 4.1.1.4. The noise processes in the L channels are assumed to be mutually

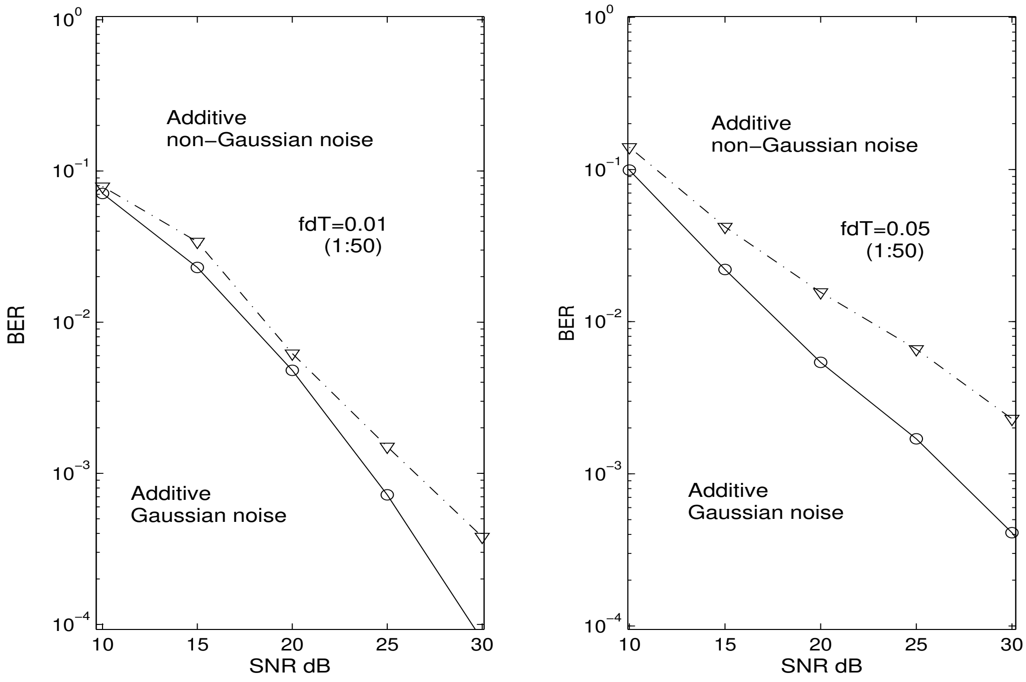

Figure 5.6: Bit error rate for coded 8PSK signals (additive non-Gaussian noise).

independent as well. Thus, the received signal for the lth channel can be expressed by the following expression, see Subsection 4.1.2 (no coding is employed in the system), i.e.

$$\begin{array}{lcl} x_{n:n-\rho+1}^{(l)} & = & \mathbf{A}x_{n-1:n-\rho}^{(l)} + \mathbf{B}v_n^{(l)}, \\ y_n^{(l)} & = & \mathbf{C}(d_{1:n})x_{n:n-\rho+1}^{(l)} + \mathbf{D}(z_n^{(l)})\epsilon_n^{(l)}, \end{array} \qquad l = 1, \ldots, L, \tag{5.7}$$

where $y_n^{(l)}$, $x_{n:n-\rho+1}^{(l)}$, and $z_n^{(l)}$ represent observations, channel states and allocation variables for the lth channel, and $v_n^{(l)} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$, $\epsilon_n^{(l)} \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$. We assume $x_{0:1-n}^{(l)} \sim \mathcal{N}_c\left(\hat{\mathbf{x}}_0^{(l)}, \mathbf{\Sigma}_0^{(l)}\right)$, where $\mathbf{\Sigma}_0^{(l)} > 0$, and let $v_n^{(l)}$, $\epsilon_n^{(l)}$ be mutually independent for all $l = 1, \ldots, L$, $n > 0$.

The message sequence d_n, the channel characteristics $x_n^{(l)}$ and the latent allocation variable $z_n^{(l)}$ are unknown for all $l = 1, \ldots, L$, $n > 0$, and the **aim** is to recover d_n from the observations $y_{1:n}^{(1)}, \ldots, y_{1:n}^{(L)}$, in particular, to obtain

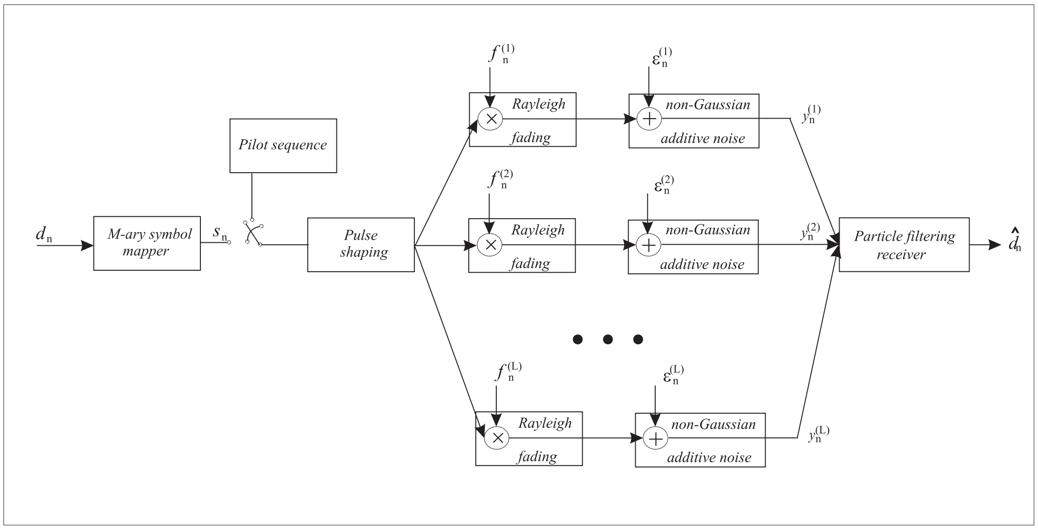

Figure 5.7: Diversity receiver

the MMAP estimates of the symbols:

$$\hat{d}_n = \arg\max_{d_n} p\left(d_n | \, y_{1:n}^{(1)}, \ldots, y_{1:n}^{(L)} \right). \tag{5.8}$$

For each $d_n \in \mathcal{R}$, $z_n^{(l)} \in \mathcal{Z}$, $l = 1, \ldots, L$, the model parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}(d_{1:n})$, $\mathbf{D}(z_n^{(l)})$, $\hat{\mathbf{x}}_0^{(l)}$, $\mathbf{\Sigma}_0^{(l)}$ are given *a priori*.

Again, the evaluation of $p\left(d_n | \, y_{1:n}^{(1)}, \ldots, y_{1:n}^{(L)} \right)$ does not admit any analytical solution, and, similarly to $L = 1$ case, particle filtering algorithm can be employed in order to obtain the approximation of this posterior distribution on-line.

5.2.2 Particle Filtering Diversity Receiver

In Chapter 4, particle filtering techniques were applied to the problem of demodulation under conditions of Rayleigh fading channels. The steps analogous to those taken in this section can yield the particle filtering diversity combining algorithm presented below.

Let us denote $y_n^{(1:L)} = \left(y_n^{(1)}, \ldots, y_n^{(L)} \right)^{\mathsf{T}}$, $x_n^{(1:L)} = \left(x_n^{(1)}, \ldots, x_n^{(L)} \right)^{\mathsf{T}}$, and $z_n^{(1:L)} = \left(z_n^{(1)}, \ldots, z_n^{(L)} \right)^{\mathsf{T}}$. Using the assumption of independent fading channels, the joint posterior density $p\left(d_{1:n}, z_{1:n}^{(1:L)}, x_{0:n}^{(1:L)} \middle| y_{1:n}^{(1:L)}\right)$ can be factorized as follows

$$\begin{aligned}
p&\left(d_{1:n}, z_{1:n}^{(1:L)}, x_{0:n}^{(1:L)} \middle| y_{1:n}^{(1:L)}\right) \\
&= p\left(x_{0:n}^{(1:L)} \middle| d_{1:n}, z_{1:n}^{(1:L)}, y_{1:n}^{(1:L)}\right) p\left(d_{1:n}, z_{1:n}^{(1:L)} \middle| y_{1:n}^{(1:L)}\right) \\
&= p\left(d_{1:n}, z_{1:n}^{(1:L)} \middle| y_{1:n}^{(1:L)}\right) \prod_{l=1}^{L} p\left(x_{0:n}^{(l)} \middle| d_{1:n}, z_{1:n}^{(l)}, y_{1:n}^{(l)}\right),
\end{aligned}$$

where, given $d_{1:n}, z_{1:n}^{(l)}$, $p\left(x_{0:n}^{(l)} \middle| d_{1:n}, z_{1:n}^{(l)}, y_{1:n}^{(l)}\right)$ may be computed using the Kalman filter. Alike in $L = 1$ case, the Sequential Importance Sampling method can be employed for the estimation of $p\left(d_{1:n}, z_{1:n}^{(1:L)} \middle| y_{1:n}^{(1:L)}\right)$. If the optimal importance distribution

$$\pi\left(d_n, z_n^{(1:L)} \middle| d_{1:n-1}, z_{1:n-1}^{(1:L)}, y_{1:n}^{(1:L)}\right) = p\left(d_n, z_n^{(1:L)} \middle| d_{1:n-1}, z_{1:n-1}^{(1:L)}, y_{1:n}^{(1:L)}\right)$$

is to be used, the importance weights can be evaluated according to the following procedure.

From Bayes' rule, using the same assumption on fading channels as above, one obtains

$$\begin{aligned}
p&\left(d_n, z_n^{(1:L)} \middle| d_{1:n-1}, z_{1:n-1}^{(1:L)}, y_{1:n}^{(1:L)}\right) \\
&= \frac{p\left(y_n^{(1:L)} \middle| y_{1:n-1}^{(1:L)}, d_{1:n-1}, z_{1:n-1}^{(1:L)}, d_n, z_n^{(1:L)}\right) p(d_n | d_{n-1}) p(z_n^{(1:L)})}{p(y_n^{(1:L)} | y_{1:n-1}^{(1:L)}, d_{1:n-1}, z_{1:n-1}^{(1:L)})} \\
&= \prod_{l=1}^{L} \left[\frac{p\left(y_n^{(l)} \middle| y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)}, d_n, z_n^{(l)}\right) p(z_n^{(l)})}{p(y_n^{(l)} | y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)})}\right] p(d_n | d_{n-1}).
\end{aligned} \tag{5.9}$$

Then, the importance weights w_n are given by

$$w_n \propto \prod_{l=1}^{L} p\left(y_n^{(l)} \middle| y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)}\right), \tag{5.10}$$

where

$$p\left(y_n^{(l)}\middle|\, y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)}\right) = \\ \sum_{m=1}^{M}\sum_{k=1}^{K}\left[p\left(y_n^{(l)}\middle|\, y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)}, d_n = m, z_n^{(l)} = k\right)\right. \\ \left.\times\, p(d_n = m|\, d_{n-1})p(z_n^{(l)} = k)\right], \tag{5.11}$$

and $p\left(y_n^{(l)}\middle|\, y_{1:n-1}^{(l)}, d_{1:n-1}, z_{1:n-1}^{(l)}, d_n = m, z_n^{(l)} = k\right)$ is given by the Kalman filter (see Chapter 3).

Thus, given $\left\{d_{1:n-1}^{(i)}, \left(z_{1:n-1}^{(1:L)}\right)^{(i)}\right\}_{i=1}^{N}$ distributed approximately according to $p\left(d_{1:n-1}, z_{1:n-1}^{(1:L)}\middle|\, y_{1:n-1}\right)$, the importance weights can be calculated (up to a normalizing constant), according to the following algorithm:

Evaluation of the importance weight

(optimal importance distribution)

- For $l = 1, \ldots, L$ evaluate $p\left(y_n^{(l)}\middle|\, y_{1:n-1}^{(l)}, d_{1:n-1}^{(i)}, \left(z_{1:n-1}^{(l)}\right)^{(i)}\right)$.
- Evaluate the importance weights up to a normalizing constant:

$$w_n^{(i)} \propto \prod_{l=1}^{L} p\left(y_n^{(l)}\middle|\, y_{1:n-1}^{(l)}, d_{1:n-1}^{(i)}, \left(z_{1:n-1}^{(l)}\right)^{(i)}\right),$$

This requires the evaluation of $M \times K \times L$ one-step ahead Kalman filter steps for each particle, which may be computationally intensive. Alternative solution is to employ the *prior* distribution $p(d_n, z_n^{(1:L)}\big|\, d_{n-1}, z_{n-1}^{(1:L)})$ as the importance distribution for which only L single steps of the Kalman filter are required.

In general, given at time $n-1$ a set of particles $\left\{d_{1:n-1}^{(i)}, \left(z_{1:n-1}^{(1:L)}\right)^{(i)}\right\}_{i=1}^{N}$ distributed approximately according to $p\left(d_{1:n-1}, z_{1:n-1}^{(1:L)}\middle|\, y_{1:n-1}\right)$, at time n the particle filtering diversity receiver proceeds as follows:

Particle Filtering Algorithm

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$,

  sample $\left(\widetilde{d}_n^{(i)}, \left(\widetilde{z}_n^{(1:L)}\right)^{(i)}\right) \sim \pi\left(d_n, z_n^{(1:L)} \middle| d_{1:n-1}^{(i)}, \left(z_{1:n-1}^{(1:L)}\right)^{(i)}, y_{1:n}^{(1:L)}\right)$

  and set $\widetilde{d}_{1:n}^{(i)} = (d_{1:n-1}^{(i)}, \widetilde{d}_n^{(i)})$, $\left(\widetilde{z}_{1:n}^{(1:L)}\right)^{(i)} = \left(\left(z_{1:n-1}^{(1:L)}\right)^{(i)}, \left(\widetilde{z}_n^{(1:L)}\right)^{(i)}\right)$.

- For $i = 1, \ldots, N$, evaluate the importance weights $w_n^{(i)}$.

- For $i = 1, \ldots, N$, normalize the importance weights $w_n^{(i)}$

$$\widetilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

Selection Step

- Multiply / discard particles $\left\{\widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(1:L),(i)}\right\}_{i=1}^{N}$ with respect to high/low normalized importance weights $\widetilde{w}_n^{(i)}$ to obtain N particles $\left\{d_{1:n}^{(i)}, z_{1:n}^{(1:L),(i)}\right\}_{i=1}^{N}$.

As was mentioned previously, in the case of the optimal importance distribution being used in the algorithm, the importance weight $w_n^{(i)}$ does not actually depend on $\widetilde{d}_{1:n}^{(i)}, \widetilde{z}_{1:n}^{(i)}$. Thus, the selection step can be done prior to the sampling step. In both cases of optimal and prior importance distributions the associated importance weights depend on $d_{1:n-1}$, $z_{1:n-1}^{(l)}$ via the mean and covariance of the state $x_{n:n-n+1}^{(l)}$ and only these values need to be kept in memory.

5.2.3 BER Performance for QAM diversity reception

We assessed the performance of the particle filtering diversity receiver for 16 QAM signals (with Gray coding) with the second order diversity ($L = 2$). Figure 5.8 and Figure 5.9 represent the results for $f_d T = 0.02$ for Gaussian

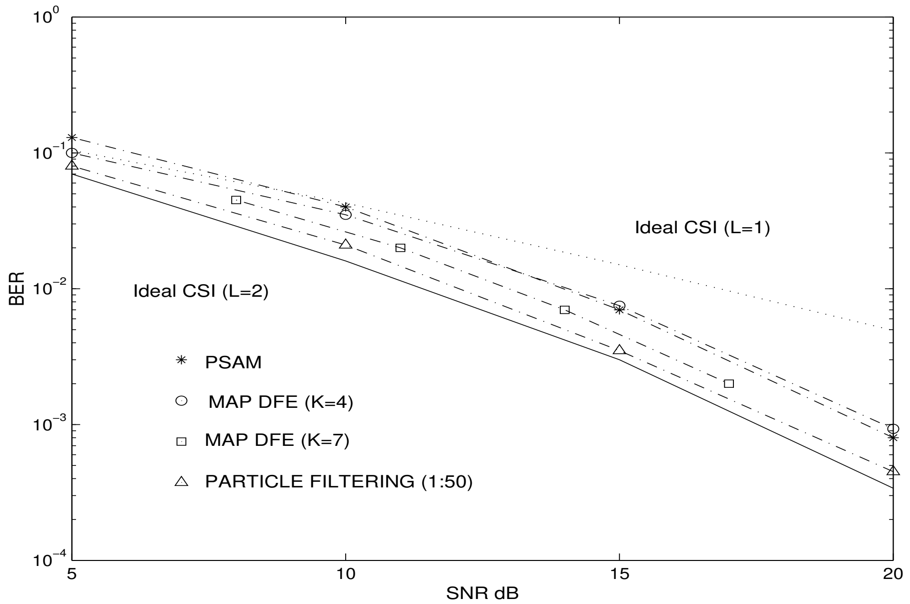

Figure 5.8: Bit error rate for 16QAM signals with second order diversity (additive Gaussian noise).

and non-Gaussian noise correspondingly (the characteristics of other parameters are the same as in the previous chapter[13]). The number of particles used was $N = 50$ with the pilot symbol rate of $1:50$. In Figure 5.8 the comparison with the ideal coherent detection, ideal pilot-symbol-assisted modulation (PSAM), and diversity combining MAP K-lag decision feedback estimator (DFE) with $K = 4{,}7$ and the pilot symbol spacing of 5 is also presented. The analytical results for the ideal coherent detection and ideal PSAM are derived in [Seymour, 1994].

5.3 CDMA

Code division multiple access (CDMA) systems have been recently under intensive research, a significant thrust of which has focused on the multiuser CDMA detection in fading environments [Davis & Collings, 1999; Hou & Chen, 2000; Raphaeli, 2000]. Fading results in a significant increase of both the intersymbol interference (ISI) among the data symbols of the same user,

[13]For a more detailed description of the simulation set-up see Section 4.3.

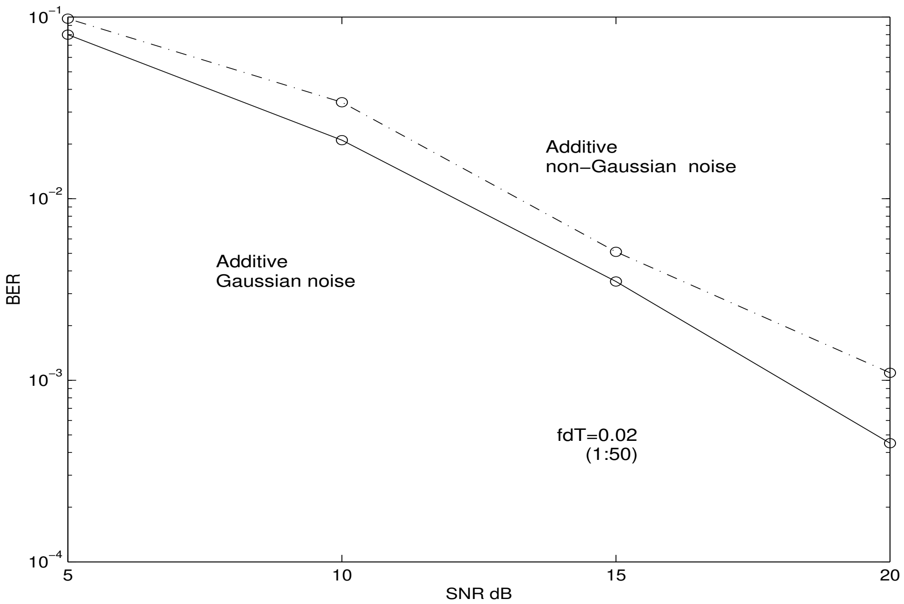

Figure 5.9: Bit error rate for 16QAM signals with second order diversity (additive non-Gaussian noise).

and the multiple-access interference (MAI) among the data symbols of different users. These, added to a possibly non-Gaussian, often impulsive, nature of the ambient noise in some physical channels (such as urban and indoor radio channels), make the problem of symbol detection extremely difficult. In this section, we address this problem using particle filtering techniques, and extend the algorithm developed in the previous chapter, and already proved useful for demodulation, decoding and detection of symbols received by multiple antennas, to consider this more complicated multiuser scenario.

5.3.1 Problem statement and estimation objectives

Let us consider the downlink of a synchronous CDMA system that is shared by K simultaneous users (see Figure 5.10), and let $d_{k,n}$ denote the nth information symbol from the kth user[14] and $s_{\text{trans},\,k}(\tau)$ be the corresponding equivalent lowpass signal waveform given by

[14]We assume that only $L = 1$ antenna and no channel coding is employed in the system, index l is, therefore, suppressed, and $r_n = d_n$, $q = n$.

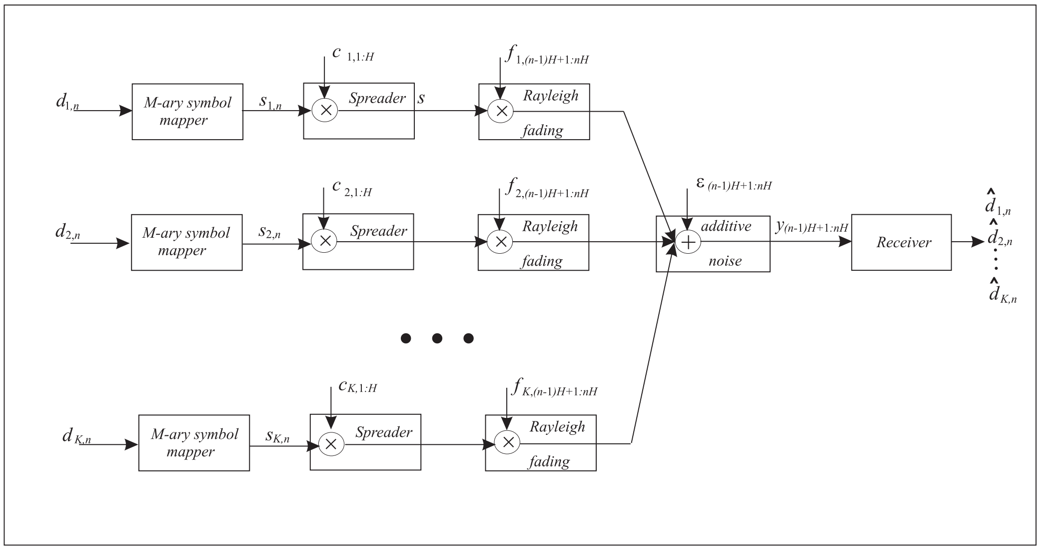

Figure 5.10: CDMA transmission system.

$$s_{\text{trans},\ k}(\tau) = s_{k,n}(d_{k,1:n})u_k(\tau), \quad (n-1)T_n < \tau \leq nT_n, \tag{5.12}$$

where $s_{k,n}(.)$ performs the mapping from the digital sequence to waveforms and corresponds to the modulation technique employed, and $u_k(\tau)$ is the signature waveform for the kth user,

$$u_k(\tau) = \sum_{h=1}^{H} c_{k,h}\eta(\tau - hT_{ch}).$$

Here, $c_{k,1:H}$ is a pseudo-noise (PN) code sequence consisting of H chips (with values $\{\pm 1\}$) per symbol, $\eta(\tau - hT_{ch})$ is a rectangular pulse of duration T_{ch} transmitted at $(h-1)T_{ch} < \tau \leq hT_{ch}$, and T_{ch} is the chip interval, $T_{ch} = T_n/H$.

The waveform goes through a flat[15] Rayleigh fading channel and is corrupted by additive complex noise which is assumed to be Gaussian[16]. Thus,

[15]Frequency-selective channels can be considered in the same framework.

[16]The case of non-Gaussian noise can be easily treated using the techniques presented in the previous Chapter.

after matched filtering and sampling at the rate T_{ch}^{-1}, the complex output of the channel at instant $t = (n-1)H + h$, $h = 1, \ldots, H$, corresponding to the transmission of the nth symbols can be expressed as

$$y|_{t=(n-1)H+h} = (c_{1,h}, \ldots, c_{K,h}) \underbrace{\begin{pmatrix} s_n(d_{1,1:n}) & 0 & \ldots & 0 \\ 0 & s_n(d_{2,1:n}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s_n(d_{K,1:n}) \end{pmatrix}}_{diag\left(s_{1,n},\ldots,s_{K,n}\right)} \underbrace{\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ \vdots \\ f_{K,t} \end{pmatrix}}_{f_{1:K,t}} + \sigma\epsilon_t,$$

where σ^2 is the variance of the additive complex noise $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$, and $f_{1:K,t}$ represents a multiplicative discrete time disturbance of the channels, which is at instant t modelled as an ARMA(ρ, ρ) process (Butterworth filter of order ρ). The ARMA coefficients $\mathbf{a}$ (AR part) and $\mathbf{b}$ (MA part) are chosen so that the cut-off frequency of the filter matches the normalized channel Doppler frequency $f_d T_{ch}$, which is known. Thus, the problem can be formulated in a linear Gaussian state space form (conditional upon the symbols), similar to Chapter 4.

The symbols $\mathbf{d}_n = d_{1:K,n}$, which are assumed i.i.d., and the channel characteristics $f_{1:K,t}$ corresponding to the transmission of the nth symbol are unknown for $n > 0$. Our aim is to estimate $\mathbf{d}_n$ given the currently available data $\mathbf{y}_{1:n}$, where $\mathbf{y}_n \triangleq y_{(n-1)H+1:nH}$, $\mathbf{y}_{1:n} \triangleq (\mathbf{y}_1, \mathbf{y}_2 \ldots, \mathbf{y}_n)^{\mathsf{T}}$. This can be done using the MAP (maximum *a posteriori*) criterion:

$$\hat{\mathbf{d}}_n = \arg\max_{\mathbf{d}_n} p\left(\mathbf{d}_n | \mathbf{y}_{1:n}\right).$$

As in all previously treated cases, the problem does not admit any analytical solution as computing $p\left(\mathbf{d}_n | \mathbf{y}_{1:n}\right)$ involves a prohibitive computational cost exponential in the (growing) number of observations and, we propose to use particle filtering techniques to estimate it.

5.3.2 CDMA particle filtering receiver

The particle filtering demodulator presented in Chapter 4 can be extended as follows. We simulate N particles $\left\{\mathbf{d}_{1:n}^{(i)}\right\}_{i=1}^{N}$ according to a convenient importance distribution $\pi(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n})$ (such that $p\left(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n}\right) > 0$ implies $\pi\left(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n}\right) > 0$, and $\pi(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n})$ admits $\pi\left(\mathbf{d}_{1:n-1}|\,\mathbf{y}_{1:n-1}\right)$ as marginal distribution), and, using the importance sampling identity, obtain an estimate of $p\left(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n}\right)$ given by the following point mass approximation:

$$\hat{p}_N\left(\mathbf{d}_{1:n}|\,\mathbf{y}_{1:n}\right) = \textstyle\sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta(d_{1:n} - \mathbf{d}_{1:n}^{(i)}), \tag{5.13}$$

The normalized importance weights $\tilde{w}_{1:n}^{(i)}$ in the above expression are given by,

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}, \tag{5.14}$$

$$w_n^{(i)} \propto \frac{p\left(\mathbf{d}_{1:n}^{(i)}\middle|\,\mathbf{y}_{1:n}\right)}{\pi\left(\mathbf{d}_{1:n}^{(i)}\middle|\,\mathbf{y}_{1:n}\right)}. \tag{5.15}$$

and a selection step is included in the algorithm at each time step in order to discard particles with low normalized importance weights and multiply those with high ones.

The resulting algorithm is similar to the one described in the previous section. If at stage $n-1$ one has N particles $\left\{\mathbf{d}_{1:n-1}^{(i)}\right\}_{i=1}^{N}$ distributed approximately according to $p\left(\mathbf{d}_{1:n-1}|\,\mathbf{y}_{1:n-1}\right)$, at time n one obtains:

CDMA Particle Filtering Receiver

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $\widetilde{\mathbf{d}}_n^{(i)} \sim \pi(\mathbf{d}_n|\,\mathbf{d}_{1:n-1}^{(i)}, \mathbf{y}_{1:n})$ and set $\widetilde{\mathbf{d}}_{1:n}^{(i)} = \left(\mathbf{d}_{1:n-1}^{(i)}, \widetilde{\mathbf{d}}_n^{(i)}\right)$.
- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant, Equation (5.14).

- For $i = 1, \ldots, N$, normalize the importance weights, Equation (5.15).

Selection Step

- Multiply / discard particles $\left\{ \widetilde{\mathbf{d}}_{1:n}^{(i)} \right\}_{i=1}^{N}$ with respect to high/low normalized importance weights $\tilde{w}_n^{(i)}$ to obtain N particles $\left\{ \mathbf{d}_{1:n}^{(i)} \right\}_{i=1}^{N}$.

The choice of the importance distribution and a selection scheme are discussed in Chapters 3 and 4; depending on those being employed, the computational complexity of the algorithm varies. As it is shown there, $\pi\left(\mathbf{d}_n \mid \mathbf{d}_{1:n-1}, \mathbf{y}_{1:n}\right) = p\left(\mathbf{d}_n \mid \mathbf{d}_{1:n-1}, \mathbf{y}_{1:n}\right)$ is an importance distribution that minimizes the conditional variance of $w\left(\mathbf{d}_{1:n}\right)$ and, therefore, is "*optimal*" in the framework considered (see Chapter 3 for details). However, for multiuser detection, for each particle it requires evaluation of the M^K H-step ahead Kalman filters for detection of the nth symbols since in this case

$$w_n \propto \sum_{m=1}^{M^K} p\left(\mathbf{y}_n \mid \mathbf{d}_{1:n-1}^{(i)}, \mathbf{d}_n = \boldsymbol{\rho}_m, \mathbf{y}_{1:n-1}\right) p(\mathbf{d}_n = \boldsymbol{\rho}_m), \tag{5.16}$$

where $\boldsymbol{\rho}_m$corresponds to the mth $(m = 1, \ldots, M^K)$ possible realization of d_n. Thus, sampling from the optimal distribution is computationally expensive if M^K is large. In this case, the *prior* distribution can be used alternatively as the importance distribution, i.e. $\pi\left(\mathbf{d}_n \mid \mathbf{y}_{1:n}, \mathbf{d}_{n-1}\right) = p\left(\mathbf{d}_n\right)$, so that, in total, for each particle at time t only one Kalman filter step is calculated (H one step Kalman filters per symbol). This could be inefficient, though, as it does not use the information carried by $\mathbf{y}_n$ to explore the state space.

Similar to the previous sections, a selection step is performed according to a stratified sampling [Kitagawa, 1996] scheme in our algorithm, which can be implemented in $O\left(N\right)$ operations.

5.3.3 Simulations

In order to demonstrate the bit-error-rate (BER) performance of our algorithm it was applied to the case of binary–phase-shift-keyed (BPSK) symbols transmitted over fast fading CDMA channels with $K = 3$, $H = 10$

and $f_d T_{ch} = 0.05$. The results for different average signal to noise ratio (SNR) compared to those obtained by maximum-a-posterior (MAP) decoder in [Davis & Collings, 1999] are given in Figure 5.11, where also the ideal channel state information (CSI) case is presented. Both, the particle receiver employing optimal (PFO) and prior (PFP) importance distributions, are considered[17]. Not surprisingly, PFO turns out to be more efficient than PFP. In order to achieve the same BER with PFP more than $N = 300$ particles were required. In general, even for just $N = 50$ particles and pilot symbol rate 1 : 20, both our algorithms outperform substantially that of [Davis & Collings, 1999], especially when the signal-to-noise ratio (SNR) is large. Additional simulations (with the number of users $K = 2$) have shown that the algorithm exhibits good performance in the case of non-Gaussian[18] additive noise (with $Z = 2$), whereas other standard methods are not actually designed to treat this case.

It should be emphasized, however, that if the number of users or processing gain in CDMA is large, a more complex modulation scheme is used and/or the additive noise is non-Gaussian (modelled as a mixture of Gaussians), PFP is of no use due to its computational complexity. Using the prior as an importance distribution combined with Markov chain Monte Carlo (MCMC) methods ([Doucet *et al.*, 2001b]) might be a good solution in this case.

5.4 Discussion

As one can see, the algorithm proposed in the previous chapter can be easily extended to consider

- joint symbol decoding and demodulation
- diversity reception
- multiuser detection

[17] For a more detailed description of the simulation set-up see Section 4.3.
[18] The extension of the algorithm is straightforward.

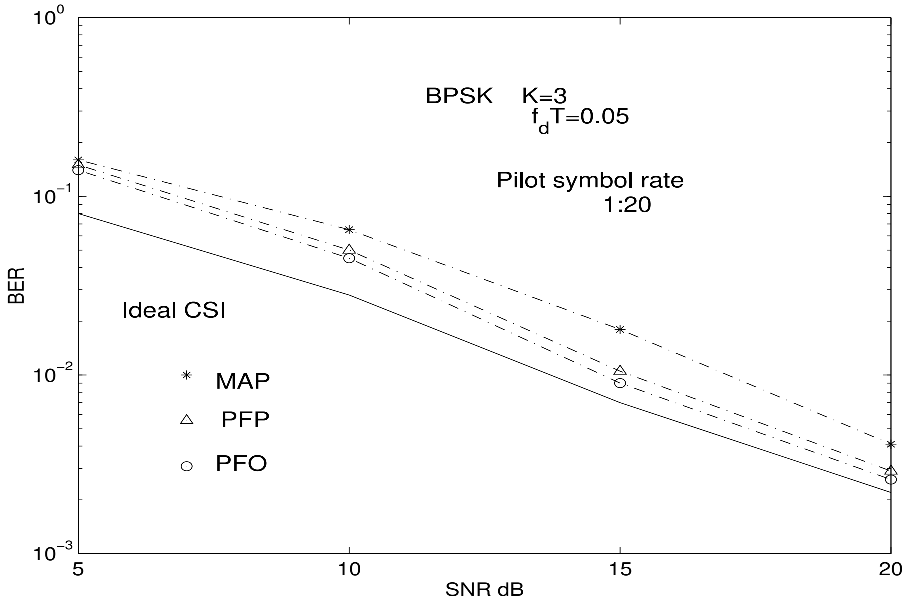

Figure 5.11: Bit error rate for CDMA transmission.

Performance advantages in comparison with other methods are demonstrated in all these cases for both Gaussian and non-Gaussian noise. The conclusions similar to the ones made at the end of the previous chapter could be reached. In addition, fixed-lag smoothing could be considered similar to Chapter 4.

The algorithm could be extended even further to consider more complicated scenarios (Chapter 7, for example). However, one of the drawbacks of the proposed methods for joint demodulation/decoding is that interleaving cannot be addressed in the same framework. In this case, the method is simply too computationally expensive to be applicable. In addition, in the case of multiuser detection with large number of users (or, indeed, in general case of the total number of states being large), a more efficient particle filtering approach involving the optimal importance distribution might be of no use due to its computational complexity. In this situation, PFP is relevant but requires the use of MCMC steps; see [Doucet *et al.*, 2001b] for details. Alternatively, a "cleverer" suboptimal importance distribution should be designed, or a different approach to multiuser detection (see [Iltis, 2001]) with consequent application of particle filtering could be used. This issues are the subject of our current research.

In this chapter, we presented an extension of the basic particle filtering demodulator to the problems of joint demodulation and decoding, space diversity and multiuser detection in flat fading conditions. The results show that the algorithm outperforms the existing methods in the case of both additive Gaussian and non-Gaussian noise for convolutionally coded information sequences, can be efficiently used with space diversity, and performs well in the difficult situation of multiuser detection. In the next chapter, we will try to investigate the proposed algorithm even further and compare it with other general approaches applicable to a similar class of models.

6 Detection Estimation Algorithm for Demodulation

The problem of demodulation under conditions of fading transmission channels (Chapter 4) has always received a lot of interest, especially with the development of particle filtering techniques capable of solving such challenging optimal filtering problems. The application of these powerful simulation-based methods to demodulation seems only to be a reasonable choice, and, while this thesis was in preparation, a number of approaches close to the work presented in Chapter 4 have been independently proposed in the literature [Chen *et al.*, 2000; Yang & Wang, 2002]. They all consider a special case with the unknown state of the model including the discrete parameters (symbols) only and the continuous-valued (channel) characteristics being integrated out. The problem one addresses hence belongs to a first class of digital communications problems (Chapter 2), and an efficient implementation of particle filtering techniques involves the use of the optimal (Chapters 3 and 4) importance sampling distribution. The results obtained by such a receiver are very promising (Chapter 4), and, in general, the method outperforms currently existing demodulation techniques.

In this chapter, however, we would like to review and compare this approach with alternative deterministic and stochastic algorithms previously presented in the literature for this class of problems. The ideas described here originate in [Tugnait & Haddad, 1979] (which is known as the detection estimation (DEA) algorithm), where preserving a limited number of most likely sequences is proposed as a way of avoiding exponentially increasing complexity of the problem. A more complicated selection scheme based on resampling low-weighted particles and keeping the ones with high weights [Fearnhead, 1998] is also considered. We discuss and test these approaches for demodulation and CDMA detection in flat Rayleigh fading channels; the performance of the proposed algorithms is demonstrated by means of computer simulations.

The chapter is organized as follows. In Section 6.1, we briefly recall the model specifications and estimation objectives for symbol demodulation in flat fading conditions. Section 6.2 introduces and reviews the deterministic and stochastic schemes to approximate the optimal filter. The simulation results comparing various approaches are presented in Section 6.3, and are discussed in Section 6.4. Finally, some conclusions are reached at the end of the chapter.

6.1 Model Specifications

First, let us briefly recall the model for the signal transmission in flat fading channels. Let d_n be the nth information symbol transmitted[19] by means of the analog waveform $s_{\text{trans}}(\tau) = \text{Re}[s_n(d_{1:n})\exp(j2\pi \text{f}_{\text{car}}\tau)]$, where f_{car} is a carrier frequency, and $s_n(\cdot)$ performs the mapping from the symbol to waveforms and depends on the modulation scheme employed. If the number of bits per symbol is κ, d_n indicates one of $M = 2^\kappa$ possible symbol sequences, $d_n \in \{1, 2, \ldots, M\}$, transmitted in the nth signalling interval; the sequence d_n is assumed to be independent identically distributed (i.i.d.).

The signal is passed through a flat Rayleigh fading channel which causes random amplitude and phase variations on the signal that can be described by a complex multiplicative discrete time disturbance f_n. We model f_n as an

[19]We denote for any generic sequence α_n, $\alpha_{i:j} \triangleq (\alpha_i, \alpha_{i+1}, \ldots, \alpha_j)^{\mathrm{T}}$.

ARMA(ρ, ρ) process (Butterworth filter of order ρ) as, for example, in [Turin & van Nobelen, 1998], with the ARMA coefficients $\mathbf{a}$ (AR part) and $\mathbf{b}$ (MA part) chosen so that the cut-off frequency of the filter matches the normalized channel Doppler frequency $\mathrm{f_d}T$ (T is the symbol rate), $\mathrm{f_d}T$ being known. Now, using a new channel parameter x_n, defined so that $f_n = \mathbf{b}^{\mathsf{T}} x_{n:n-\rho+1}$, we can formulate the problem in the linear state space form conditional upon the symbol sequence $d_{1:n}$:

$$\begin{aligned} x_{n:n-\rho+1} &= \mathbf{A} x_{n-1:n-\rho} + \mathbf{B} v_n, \\ y_n &= \mathbf{C}(d_{1:n}) x_{n:n-\rho+1} + \sigma \epsilon_n, \end{aligned} \tag{6.1}$$

where y_n is the complex output of the channel corrupted by additive complex noise ϵ_n with variance σ^2, $\mathbf{A}$ is a function of $\mathbf{a}$, $\mathbf{B} = (1, 0, \ldots, 0)^{\mathsf{T}}$ and $\mathbf{C}(d_{1:n}) = s_n(d_{1:n})\mathbf{b}^{\mathsf{T}}$. We assume $x_{0:1-\rho} \sim \mathcal{N}_c(\hat{\mathbf{x}}_0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0 > 0$, and let $v_n \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$, $\epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1)$ be mutually independent for all $n > 0$.

With the symbols $d_{1:n}$ and channel parameters $f_{1:n}$ (and correspondingly $x_{1:n}$) being unknown, our aim is to obtain the maximum *a posteriori* (MAP) estimate of $d_{1:n}$, $\underset{d_{1:n}}{\arg\max}\, p(d_{1:n}|\, y_{1:n})$, sequentially in time.

6.2 Particle filtering receiver

Given the observations $y_{1:n}$, all Bayesian inference on $d_{1:n}$ relies on the posterior probability distribution $p(d_{1:n}|\, y_{1:n})$, which satisfies the following recursion:

$$p(d_{1:n}|\, y_{1:n}) = p(d_{1:n-1}|\, y_{1:n-1}) \times \frac{p(y_n|\, d_{1:n}, y_{1:n-1}) p(d_n)}{p(y_n|\, y_{1:n-1})}. \tag{6.2}$$

For a given symbol sequence $d_{1:n}$, this posterior can be evaluated up to a normalizing constant $p(y_n|\, y_{1:n-1})$ with the help of the Kalman filter. In our case, at each stage n, there are M^n possible sequences that have to be considered resulting in exponentially increasing complexity of the scheme. Hence, the problem that has to be addressed here is how to avoid this prohibitive computational cost, i.e. limit the number of propagated terms to a finite predetermined integer.

6.2.1 Particle filter with optimal importance distribution

In Chapter 4, we propose to use particle filtering to address this problem. The idea is to approximate the posterior $p\left(d_{1:n} | y_{1:n}\right)$ by the following point mass approximation using the importance sampling identity:

$$\hat{p}_N\left(d_{1:n} | y_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(d_{1:n} - d_{1:n}^{(i)}\right), \tag{6.3}$$

The particles $\left\{d_{1:n}^{(i)}\right\}_{i=1}^{N}$ in the above expression are simulated according to an arbitrary convenient importance distribution $\pi(d_{1:n} | y_{1:n})$ (such that $p\left(d_{1:n} | y_{1:n}\right) > 0$ implies $\pi\left(d_{1:n} | y_{1:n}\right) > 0$, and $\pi(d_{1:n} | y_{1:n})$ admits $\pi\left(d_{1:n-1} | y_{1:n-1}\right)$ as marginal distribution), and an efficient implementation involves using $\pi(d_n | d_{1:n-1}, y_{1:n}) = p(d_n | d_{1:n-1}, y_{1:n})$. This distribution minimizes the conditional variance of the importance weights and, therefore, is "*optimal*" in the framework considered (see Chapter 3 and 4 for details). However, for each particle it requires evaluation of M Kalman filter steps for each symbol since in this case

$$w_n^{(i)} \propto \sum_{m=1}^{M} p\left(y_n | d_{1:n-1}^{(i)}, d_n = m, y_{1:n-1}\right) p(d_n = m), \tag{6.4}$$

where m corresponds to each possible realization of d_n. Thus, sampling from the optimal distribution is computationally expensive if M is large. Moreover, since all the calculations have to be performed anyway, it is better to base our approximation of $p\left(d_{1:n} | y_{1:n}\right)$ directly on :

$$\hat{p}_{N\times M}\left(d_n | y_{1:n}\right) = \sum_{i=1}^{N} \sum_{m=1}^{M} \tilde{w}_n^{(i,m)} \delta(d_{1:n} - \left\{d_{1:n-1}^{(i)}, d_n = m\right\}), \tag{6.5}$$

thus, considering all possible "extensions" of the existing state sequences for each particle at step n. We refer to this method as *deterministic* particle filter and describe it in details in the next section.

6.2.2 Deterministic particle filter

Suppose that, at the $(n-1)$th stage, N_{n-1} appropriately weighted symbol sequences $\left\{d_{1:n-1}^{(i)}\right\}_{i=1}^{N_{n-1}}$, hereafter referred to as *particles,* are available that approximate the posterior distribution $p\left(d_{1:n-1} \middle| y_{1:n-1}\right)$ as follows:

$$\hat{p}_{N_{n-1}}\left(d_{1:n-1} \middle| y_{1:n-1}\right) = \sum_{i=1}^{N_{n-1}} \tilde{w}_n^{(i)} \delta\left(d_{1:n-1} - d_{1:n-1}^{(i)}\right), \tag{6.6}$$

where $\tilde{w}_n^{(i)}$ are the normalized weights of the particles: $\sum_{i=1}^{N_{n-1}} \tilde{w}_n^{(i)} = 1$, $\tilde{w}_n^{(i)} > 0$ for all i.

At stage n, each of these particles $d_{1:n-1}^{(i)}$, $i = 1, \ldots, N_{n-1}$, could have M new "extensions" or so-called offspring $\left\{d_{1:n-1}^{(i)}, d_n = m : m = 1, \ldots, M\right\}$, where m corresponds to the possible realizations of d_n, thus resulting in a total number of $N_{n-1} \times M$ potential candidates and the following approximation of $p\left(d_{1:n} \middle| y_{1:n}\right)$:

$$\hat{p}_{N_{n-1}\times M}\left(d_{1:n} \middle| y_{1:n}\right) = \sum_{i=1}^{N_{n-1}} \sum_{m=1}^{M} \tilde{w}_n^{(i,m)} \delta\left(d_{1:n} - \left\{d_{1:n-1}^{(i)}, d_n = m\right\}\right). \tag{6.7}$$

From Equation (6.2), the corresponding weights of these particle offspring are proportional to the posterior $p\left(d_{1:n-1}^{(i)}, d_n = m \middle| y_{1:n}\right)$ and, therefore, depend on the weight of the parent at step $n-1$, the prior distribution and the likelihood term, which can be computed using the Kalman filter:

$$w_n^{(i,m)} \propto \tilde{w}_{n-1}^{(i)} p\left(y_n \middle| d_{1:n-1}^{(i)}, d_n = m, y_{1:n-1}\right) p(d_n = m). \tag{6.8}$$

The denominator in (6.2) is common for all $d_{1:n}$ and is eliminated in weight normalization:

$$\tilde{w}_n^{(i,m)} = \frac{w_n^{(i,m)}}{\sum_{i=1}^{N_{n-1}} \sum_{m=1}^{M} w_n^{(i,m)}}.$$

In terms of calculations, this approach is equivalent to the use of the optimal distribution in particle filtering algorithm. However, when performing inference on the symbol d_n, it is, of course, better to use $\hat{p}_{N_{n-1}\times M}\left(d_{1:n} \middle| y_{1:n}\right)$ than the standard particle filtering approximation (6.3) since one does not discard unnecessarily any information by selecting randomly one path out of the M available.

If now the total number of components with non-zero weights $N_{n-1} \times M - N_{\text{zero}}$ (where N_{zero} is the number of components with zero weight) exceeds the predetermined maximum allowable number N, a selection scheme has to be employed in the procedure in order to avoid the exponentially increasing complexity of the algorithm. The simplest way to perform such selection is to keep $N_n = N$ most likely particles with the highest weights $\tilde{w}_n^{(i,m)}$ (those selected are then renumbered), and discard the others (as, for example, in [Tugnait & Haddad, 1979]). This means that, at each stage, all possible "extensions" of the existing symbol sequences are considered, however, only a maximum allowable number of them are carried forward; the rest of the sequences are disregarded as unlikely ones.

A more complicated approach involves preserving the particles with high weights and resampling the ones with low weights (RLW), thus reducing their total number to N. An important condition for the design of the selection scheme in this specific context is to resample *without replacement*, i.e. with each particle appearing at most once in the resulting set, as, indeed, *there is no point in carrying along two particles evolving in exactly the same way.* An algorithm of this type is presented in [Fearnhead, 1998] but other selection schemes can be designed.

Thus, given at time $n-1$, $\left\{ d_{1:n}^{(i)} \right\}_{i=1}^{N_{n-1}}$ distributed approximately according to $p\left(d_{1:n-1} \middle| y_{1:n-1} \right)$, at time n the deterministic particle filter proceeds as follows:

Deterministic particle filter

- For $i = 1, \ldots, N_{n-1}$,

  for $m = 1, \ldots, M$, perform one-step Kalman filter update and evaluate the importance weight $w_n^{(i,m)}$ up to a normalizing constant:

$$w_n^{(i,m)} \propto \tilde{w}_{n-1}^{(i)} p\left(y_n \middle| d_{1:n-1}^{(i)}, d_n = m, y_{1:n-1} \right) p(d_n = m).$$

- For $i = 1, \ldots, N_{n-1}$, for $m = 1, \ldots, M$, normalize $w_n^{(i,m)}$ to obtain $\tilde{w}_n^{(i,m)}$:

$$\tilde{w}_n^{(i,m)} = \frac{w_n^{(i,m)}}{\sum_{i=1}^{N_{n-1}} \sum_{m=1}^{M} w_n^{(i,m)}}.$$

- If $N_{n-1} \times M - N_{\text{zero}} > N$,

  select $N_n = N$ trajectories out of $N_{n-1} \times M - N_{\text{zero}}$ possible "extensions" with respect to high/low normalized importance weights $\tilde{w}_n^{(i,m)}$ without replacement to obtain a set of new particles $\left\{ d_{1:n}^{(i)} \right\}_{i=1}^{N_n}$

Whether we choose to preserve the most likely particles or employ the selection scheme proposed in [Fearnhead, 1998], the computational load of the resulting algorithms at each step n is that of $N_{n-1} \times M$ Kalman filters with $N_{n-1} \leq N$. The selection step, if any, for both cases is implemented in $O(N_{n-1} \times M \log N_{n-1} \times M)$ operations compared to $O(N)$ when, for example, the stratified sampling [Kitagawa, 1996] in standard particle filtering is employed. Of course, if M is large, which is the case in many applications, both these methods are too computationally expensive to be used.

6.3 Simulations

Computer simulations were carried out in order to compare the performance of the methods presented in the previous section. The algorithms were applied to demodulation of $4DPSK$ signals

$$s_n = \exp(j\theta_n), \theta_n = \sum_{j=1}^{n} \sum_{m=1}^{4} \frac{2\pi m}{4} \left(d_j - d_m \right),$$

transmitted over reasonably fast (see [Proakis, 1995]) Rayleigh fading channels with normalized channel Doppler frequency $\mathrm{f_d}T = 0.05$. The fading chan-

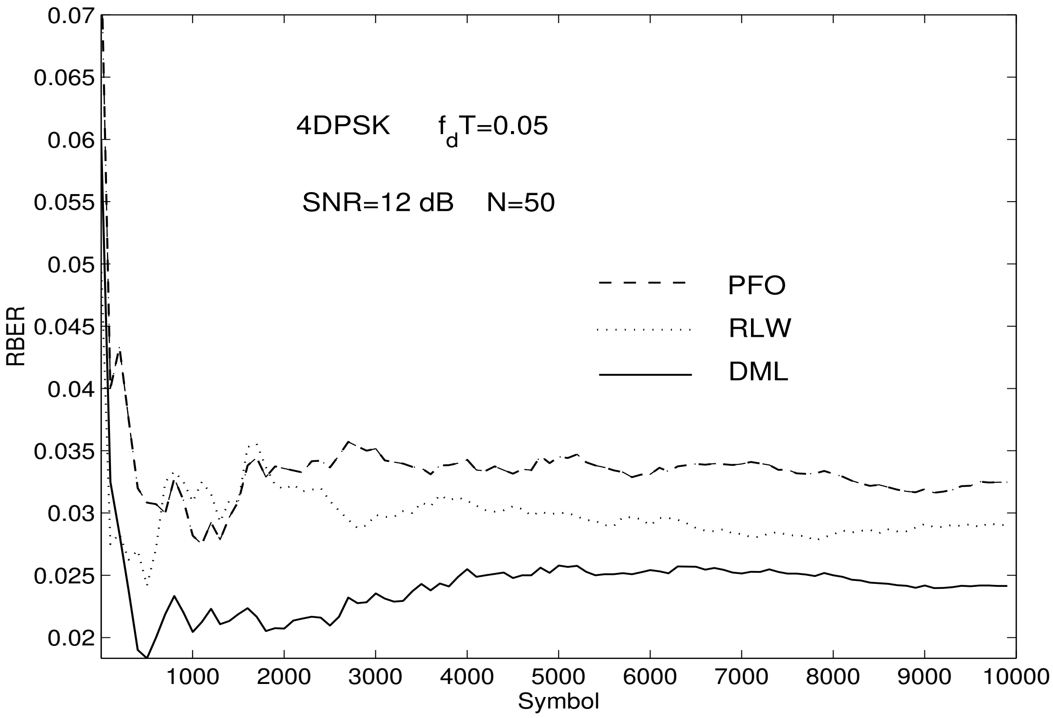

Figure 6.1: Bit error rate performance comparison.

nel was generated using Jakes' method with corresponding $\mathrm{f_d}T$. Figure 6.1 shows the bit-error-rate (BER) performance of the standard particle filtering algorithm employing optimal importance sampling distribution (PFO) compared to those obtained by deterministic approach preserving N most likely particles (DML), and the deterministic receiver incorporating the resampling low weights (RLW) selection procedure [Fearnhead, 1998]. The number of particles used in all algorithms was equal to $N = 50$ and the average signal to noise ratio (SNR) was 12 dB. The results for different average signal to noise ratio (SNR) and the number of particles N are presented in Figure 6.2 and 6.3 correspondingly[20]. They are interesting in the sense that, in this case, the simplest DML approach turned out to be the most effective one. In order to achieve the same BER using a more complicated RLW selection

[20]For a more detailed description of the simulation set-up see Section 4.3.

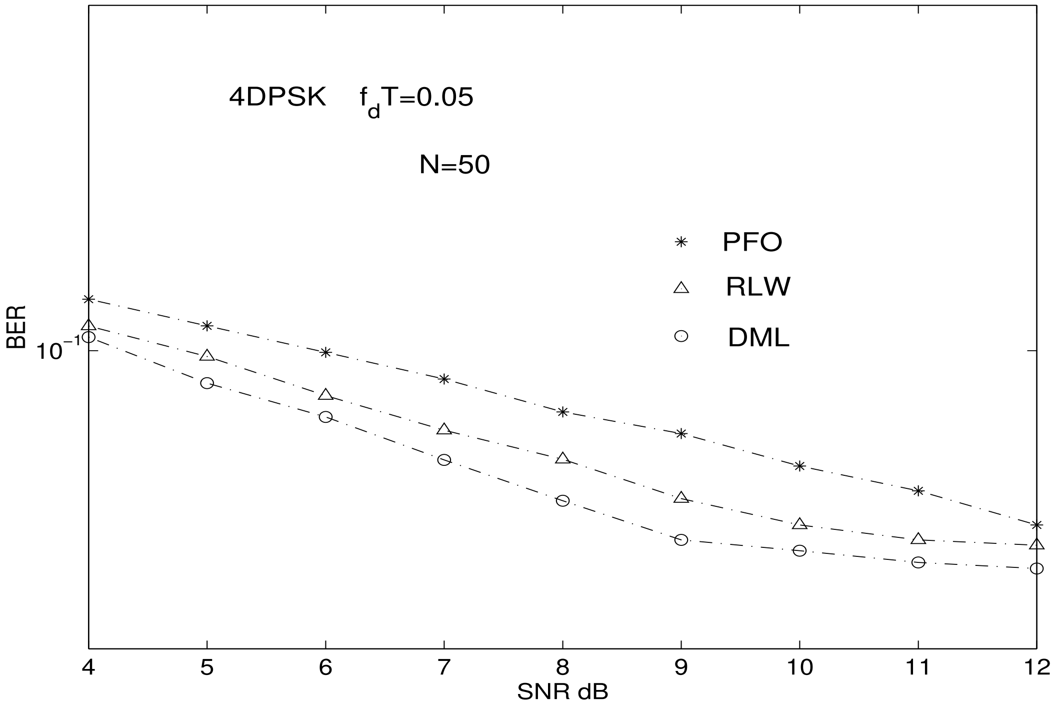

Figure 6.2: Bit error rate via SNR.

scheme more than $N = 600$ particles were required, and more than $N = 1000$ particles were needed with PFO. With other simulation parameters, however, one may find that the results between the different algorithms are much less pronounced, see BER for binary DPSK (BDPSK) signals in relatively slow fading, Figure 6.4, for example.

In the second series of experiments, we extended the proposed algorithms (analogously to Chapter 4 and 5) to the case of non-Gaussian additive noise. Similar to Chapter 4, the additive noise was distributed as a two-component mixture of zero-mean Gaussians ($Z = 2$) with the overall variance being equal to $\lambda_1\sigma_1^2 + (1-\lambda_1)\sigma_2^2$, $\lambda_1 = 0.2$. The results are shown in Figure 6.5, and

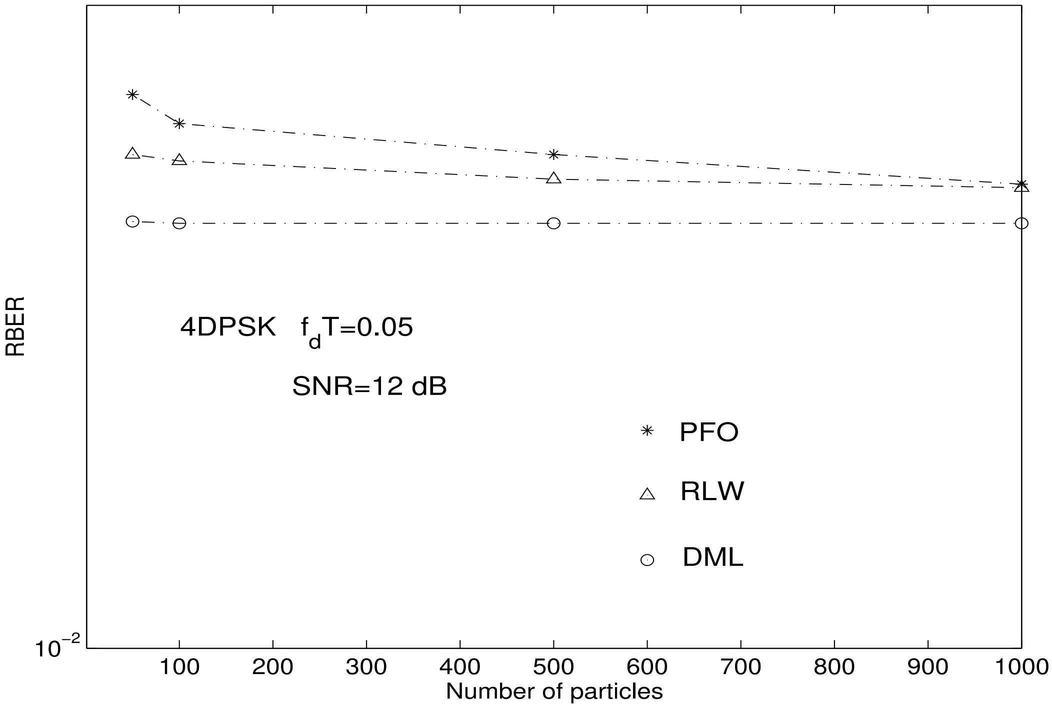

Figure 6.3: Bit error rate for different number of particles.

are similar to the ones obtained earlier in terms of algorithms performance.

Finally, a synchronous CDMA system with $K = 3$ users and $H = 10$ was considered. The algorithms were applied to the case of binary DPSK (BDPSK) signals with $\mathrm{f_d}T = 0.05$. The bit-error-rate performance for this scenario is illustrated in Figure 6.6, and the same conclusions can be made.

6.4 Discussion

To conclude, one could hope that randomization "helps" by allowing particles with a small weight to survive, but simulations presented in the previous section show that it is not necessarily the case. In this very specific but important context, particle filtering algorithms do not perform better than the

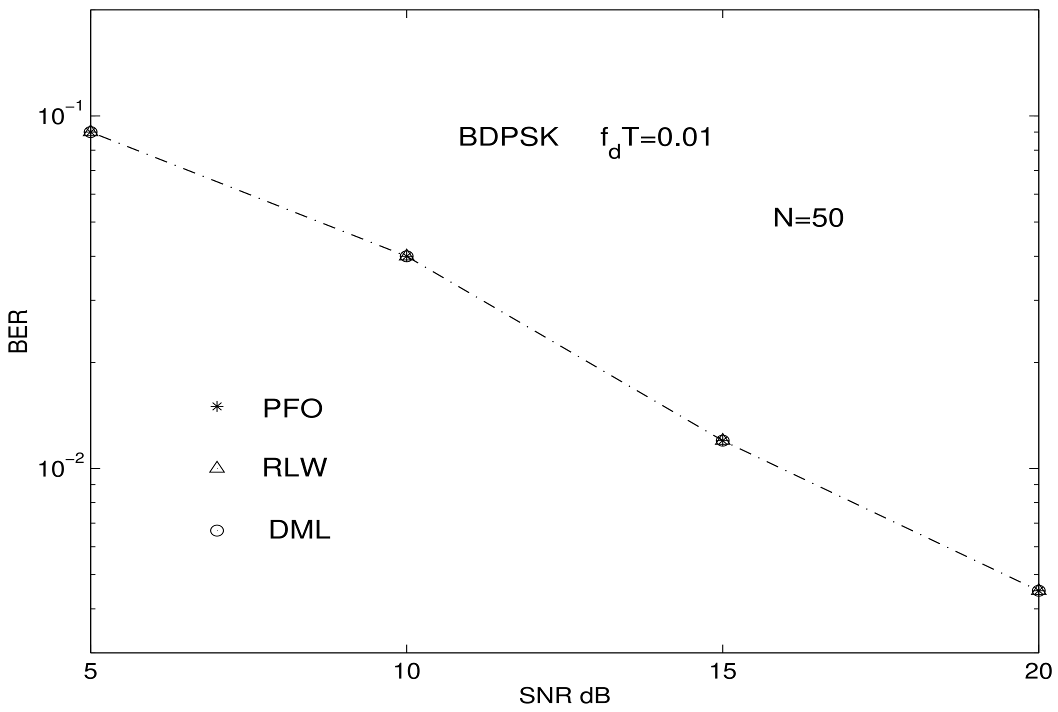

Figure 6.4: Bit error rate for BDPSK signals and relatively slow fading.

simplest deterministic method which consists of keeping at each time step the best N hypothesis. One should note, however, that

- These conclusions must indeed be interpreted cautiously. The results with other simulation parameters (such as demodulation of BDPSK signals in very slow multipath fading) were much less pronounced. These issues need further investigation.

- In principle, all schemes are capable of providing optimal performance given a large number of particles. The simplicity and efficiency is, however, one of the great advantages of DML.

- Although the DML approach outperforms many existing methods, it might suffer from the drawback of "never forgetting the past". As a result, one minor error made at the beginning of estimation might lead

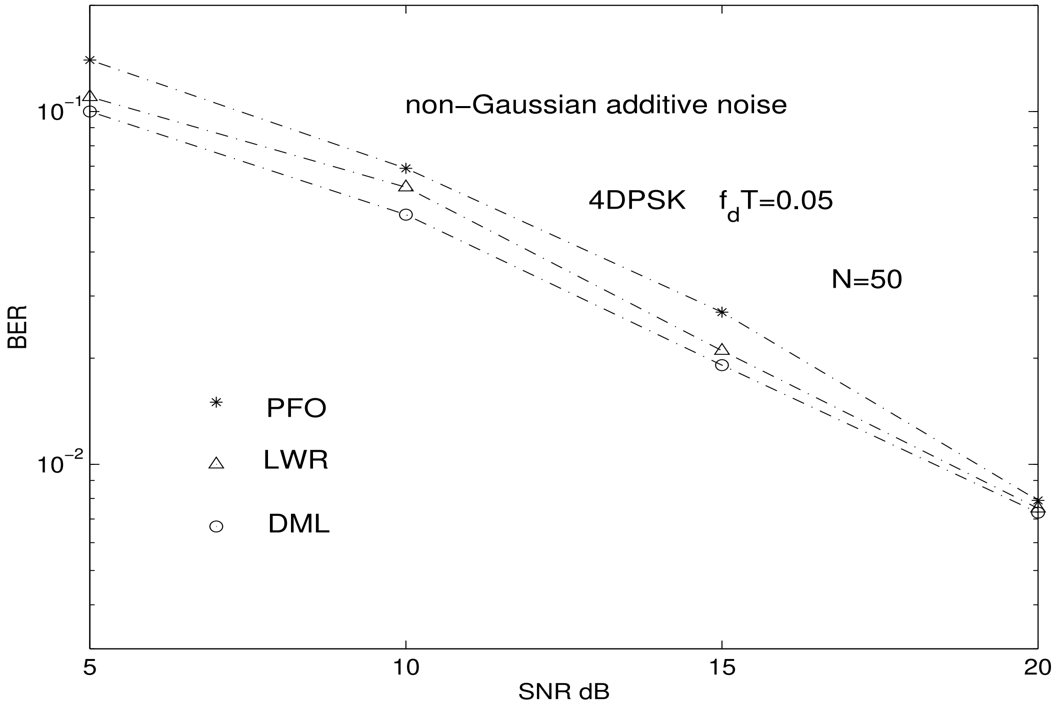

Figure 6.5: Bit error rate for non-Gaussian additive noise.

to a significant performance degradation at later stages. The algorithm, however, could be improved by introducing a discount factor β, $0 < \beta < 1$, in order to reduce the effect of the previous decisions on the future performance of the method. Following this approach, the discounted weights of the particles would be calculated as:

$$\tilde{w}_n^{(i,m)} \propto \left(\tilde{w}_{n-1}^{(i)}\right)^{\beta} p\left(y_n \middle| d_{1:n-1}^{(i)}, d_n = m, y_{1:n-1}\right) p(d_n = m), \qquad (6.9)$$

and the lower the value of the discount factor the less effect the past would have on the future. Such *discounted detection estimation* (DDE) algorithm is the subject of our current research. Further research should be directed into optimization of this scheme with respect to the parameter β.

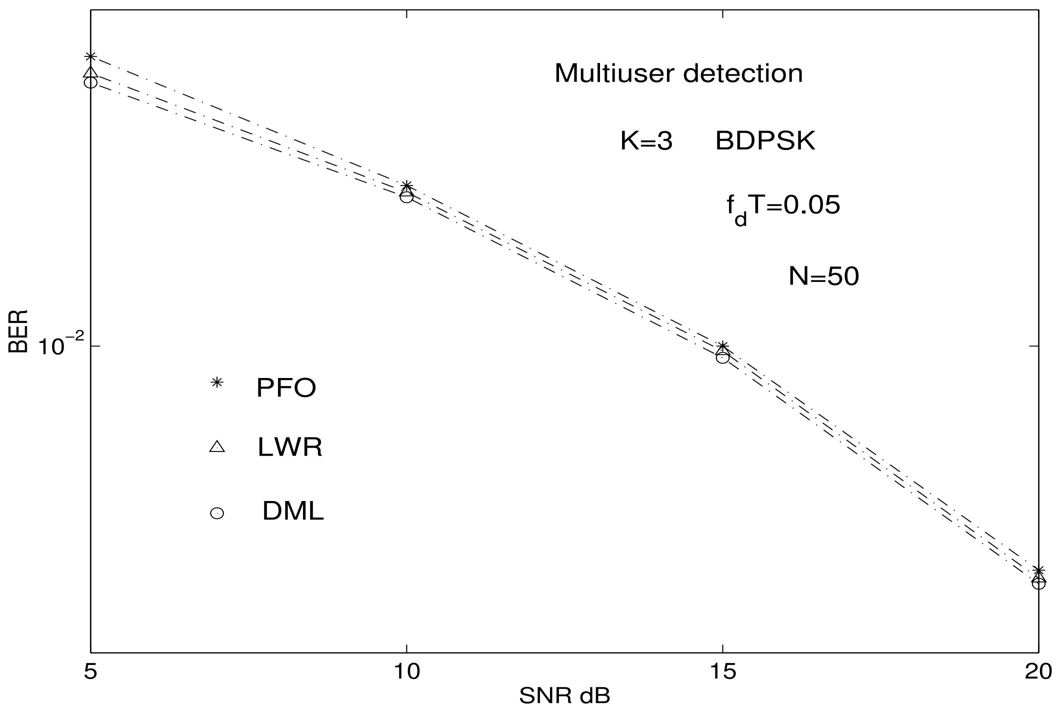

Figure 6.6: Bit error rate for multiuser detection.

- The problem considered in this chapter is rather specific in that the unknown state of the model involves the discrete parameters only. In more complex scenarios, the DML and RLW approaches are not applicable.

- In the context involving both the discrete and continuous-valued unknown parameters, DML could be combined with the particle filter to obtain a *deterministic particle filtering* receiver. Such algorithm is investigated in the next chapter for the problem of joint symbol detection and propagation delay estimation for DS spread spectrum systems in the multipath environment.

- Using the techniques presented in Chapter 4 and 5, the algorithm can be easily extended to consider non-Gaussian additive noise, decoding, space diversity and multiuser detection. It should be emphasized, however, that if the number of users is large, a more complex modulation scheme is used and/or the additive noise is non-Gaussian (modelled as a mixture of Gaussians), both DML and RLW are of no use due to their computational complexity.

- In the case where it is too costly to explore M hypothesis for each particle, standard particle filter employing the prior as an importance distribution and combined with Markov chain Monte Carlo (MCMC) methods (Chapter 2) leads to very good performance.

- The SMC algorithms could also prove useful if one could develop suboptimal importance distributions allowing more efficient sampling. This problem has to be addressed on a case by case basis. An interesting way to explore consists of randomizing standard deterministic algorithms such as successive interference cancellation or iterative least squares. However, advanced deterministic pruning strategies can also be developed using, for example, a coordinate ascent version of the algorithm proposed in [Doucet *et al.*, 2001b].

In this chapter, we have reviewed several approaches based on SMC methods to perform (approximate) optimal filtering for the *first class* of digital communications problems (Chapter 2). A simulation study has been carried out in order to compare these algorithms for symbols demodulation. Such a comparison has not been made before. In the context where only the discrete parameters are unknown, which is the case in demodulation, standard particle filtering methods, although quite capable of providing good performance, do not necessarily compare favorably with deterministic approaches: as simulations show, the most basic deterministic algorithm preserving the N most likely particles also turns out to be the most effective one.

This deserves further study and does not mean that particle filtering methods are of no use in communication systems. Indeed, for more complex problems involving continuous-valued unknown parameters, or in situations where DML and similar methods are of no use due to their computational complexity, these deterministic approaches do not apply and particle methods appears to be really useful. In this framework, particle filtering based on sampling importance resampling is relevant but requires the design of an efficient suboptimal importance distribution and/or the use of MCMC steps; see [Doucet *et al.*, 2001b] for details. For the cases including both discrete and continuous-valued parameter estimation, as in DS spread-spectrum system analyses, the particle filter could be combined with DML in order to explore the state space in a more efficient manner. This problem is addressed in the next chapter where the algorithms of this type are investigated.

7 Joint Symbol/Delay Estimation in Spread Spectrum Systems

Direct sequence (DS) spread spectrum systems are robust to many channel impairments, allow multiuser (CDMA) and low-detectability signal transmission, and, therefore, are widely used in different areas of digital communications. Unlike many other communication systems, however, spread spectrum receivers require additional code synchronization, which can be a rather challenging task under conditions of multipath fading, when severe amplitude and phase variations take place.

The problem of joint delay and multipath estimation has been addressed in the literature before (see [Iltis, 2001, 1990], for example), and has proved to be difficult due to its inherited nonlinearity. The previously proposed approaches are mainly based on the use of the Extended Kalman Filter (EKF). However, many of them concentrate on the channel parameters and delay estimation only; moreover, in a number of cases, when EKF methods are applied, the estimated parameters are divergent, [Iltis, 2001].

In this chapter, we propose to estimate the channel parameters, code delays and symbols jointly using particle filtering techniques. The methods have already been successfully applied to the problems arising in digital communications, in particular, demodulation in fading channels (Chapter 4) and detection in synchronous CDMA (Chapter 5). In this work, the unknown fading channel characteristics were integrated out and only the symbols needed to be imputed. The algorithm, thus, made use of the structure of the model, and the unknown state involved discrete parameters only. Later investigation (Chapter 6), however, revealed some concerns regarding the efficiency of standard particle filtering techniques in this context. It has been shown that, for a fixed computational complexity, more efficient deterministic schemes can be designed which lead to improved receiver performance.

We attempt here to study these results further, and compare various randomized and non-randomized approaches. The problem we are dealing with is more complex though, since it involves both discrete (symbols) and continuous-valued (delays) unknowns. Although the deterministic method is not applicable directly in this case, it can be combined with sequential importance sampling for the continuous-valued parameter, followed by an appropriate selection procedure. The resulting algorithm explores the state space in a more systematic way at little or no extra cost compared to standard particle filtering using a suboptimal importance distribution. We develop and test this approach against other deterministic and stochastic schemes, and demonstrate its performance by means of an extensive simulation study.

The remainder of the chapter is organized as follows. The model specifications and estimation objectives are stated in Section 7.1. In Section 7.2, a particle filtering method is developed for joint symbol/channel coefficients/code delay estimation. Several alternative deterministic and stochastic schemes are also introduced and reviewed. Simulation results and comparisons are presented in Section 7.3. Some conclusions are drawn at the end of the chapter.

7.1 Problem statement and estimation objectives

Let us begin with the model specification for signal transmission in DS spread spectrum systems in a multipath environment. For simplicity, we address a single user case with one receiving antenna, and set $K = 1$ and $L = 1$, suppressing the indices k and l hereafter. No channel coding is employed in the system, and the additive channel noise is assumed to be Gaussian. Extension to more complicated scenarios is straightforward.

Transmitted waveform. Let us denote for any generic sequence κ_t, $\kappa_{i:j} \triangleq (\kappa_i, \kappa_{i+1}, \ldots, \kappa_j)^{\mathrm{T}}$, and let d_n be the nth information symbol and $s_{\text{trans}}(\tau)$ be the corresponding analog bandpass spread-spectrum signal waveform transmitted in the symbol interval of duration T_n:

$$s_{\text{trans}}(\tau) = \mathrm{Re}[s_n(d_{1:n})u(\tau)\exp(j2\pi \mathrm{f}_{\text{car}}\tau)], \quad \text{for } (n-1)T_n < \tau \leq nT_n, \quad (7.1)$$

where $s_n(.)$ maps the digital sequence to waveforms and depends on the modulation technique employed, $d_n = r_n$, with $q = n$ since uncoded symbols are considered (see Chapter 2), f_{car} denotes the carrier frequency and $u(\tau)$ is a wide-band pseudo-noise (PN) waveform defined by

$$u(\tau) = \sum_{h=1}^{H} c_h \eta(\tau - hT_{ch}). \quad (7.2)$$

Here, $c_{1:H}$ is a spreading code sequence consisting of H chips (with values $\{\pm 1\}$) per symbol, $\eta(\tau - hT_{ch})$ is a rectangular pulse of unit height and duration T_{ch} transmitted at $(h-1)T_{ch} < \tau \leq hT_{ch}$, and T_{ch} is the chip interval satisfying the relation $T_{ch} = T_n/H$.

Channel model. The signal is passed through a noisy multipath fading channel which induces random amplitude and phase variations. The channel can be represented by a time-varying tapped-delayed line with taps spaced T_s seconds apart, where T_s is the Nyquist sampling rate for the transmitted waveform; $T_s = T_{ch}/2$ due to the PN bandwidth being approximately $1/T_{ch}$. The equivalent discrete-time impulse response of the channel is given by

$$h_{\text{channel},t} = \textstyle\sum_{g=0}^{G-1} f_t^{(g)} \delta(t-g), \quad (7.3)$$

where t is a discrete time index, G is the number of paths of the channel, $f_t^{(g)}$ are the complex-valued time-varying multipath coefficients (arranged into the vector $\mathbf{f}_t$), and δ denotes the Dirac delta function.

We assume here that the channel coefficients $\mathbf{f}_t$ and code delay θ_t propagate according to the first-order autoregressive (AR) model:

$$\begin{aligned} \mathbf{f}_t &= \mathbf{A}\mathbf{f}_{t-1} + \mathbf{B}\mathbf{v}_t, \mathbf{v}_t \overset{i.i.d.}{\sim} \mathcal{N}_c\left(\mathbf{0}, \mathbf{I}_G\right), & (7.4) \\ \theta_t &= \gamma\theta_{t-1} + \sigma_\theta \vartheta_t,\ \vartheta_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0,1\right), & (7.5) \end{aligned}$$

which corresponds to a Rayleigh uncorrelated scattering channel model; here $\mathbf{A} \triangleq diag(\alpha^{(0)}, \ldots, \alpha^{(G-1)})$, $\mathbf{B} \triangleq diag(\sigma_f^{(0)}, \ldots, \sigma_f^{(G-1)})$, where $\sigma_f^{(g)}$ is the standard deviation, and $\alpha^{(g)}$ accounts for the Doppler spread (see [Iltis, 1990; Komninakis *et al.*, 1999] for details and discussion of the use of higher order AR). In this work, matrices $\mathbf{A}$, $\mathbf{B}$, and parameters γ and σ_θ are assumed known. Directions on the choice of these parameters are given in [Iltis, 1990; Komninakis *et al.*, 1999]

Received signal. The complex output of the channel sampled at the Nyquist rate, (in which case samples $2H(n-1)+1, \ldots, 2Hn$ correspond to the nth symbol transmitted, i.e. $d_n \leftrightarrow y_{2H(n-1)+1:2Hn}$) can thus be expressed as

$$y_t = \mathbf{C}(d_{1:n}, \theta_{1:t}) + \sigma\epsilon_t,\ \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0,1\right), \qquad (7.6)$$

where $\mathbf{C}(d_{1:n}, \theta_{1:t}) = \sum_{g=0}^{G-1} f_t^{(g)} s_{\text{receive}}\left((t-g)\,T_s - \theta_t\right)$ and σ^2 is the noise variance[21]. The noise sequences ϑ_t, ϵ_t and $\mathbf{v}_t^{(g)}$, $n = 0, \ldots, G-1$ are assumed mutually independent and independent of the initial states $\mathbf{f}_0 \sim \mathcal{N}_c\left(\hat{\mathbf{f}}_0, \mathbf{\Sigma}_{\mathbf{f},0}\right)$, $\theta_0 \sim \mathcal{N}\left(\hat{\theta}_0, \Sigma_{\theta,0}\right)$. The received waveform $s_{\text{receive}}(\tau)$ is obtained after ideal low-pass filtering of rectangular pulses and is given by [Iltis, 1990]:

$$s_{\text{receive}}(\tau) = s_n(d_{1:n}) \sum_{h=1}^{H} c_h \frac{1}{\pi}[\text{Si}\left(2\pi\frac{\tau-(h-1)\,T_{ch}}{T_{ch}}\right) - \text{Si}\left(2\pi\frac{\tau - hT_{ch}}{T_{ch}}\right)],$$
$$\text{for } (n-1)T_n < \tau \leq nT_n,$$

[21] The case of non-Gaussian noise can be treated using the techniques presented in Chapter 4.

where

$$\mathrm{Si}(\phi) = \int_0^{\phi} \frac{\sin(\varphi)}{\varphi} d\varphi. \tag{7.7}$$

Estimation objectives. The symbols d_n, which are assumed i.i.d., the channel characteristics $\mathbf{f}_t$ and the code delay θ_t are unknown for $n, t > 0$. Our aim is to obtain sequentially in time an estimate of the joint posterior probability density of these parameters $p\left(d_{1:n}, \mathbf{f}_{0:2Hn}, \theta_{0:2Hn} \middle| y_{1:2Hn}\right)$, and some of its characteristics, such as the MMAP (marginal maximum *a posteriori*) estimates of the symbols

$$\hat{d}_{1:n} = \arg\max_{d_{1:n}} p\left(d_{1:n} \middle| y_{1:2Hn}\right), \tag{7.8}$$

and the minimum mean square error (MMSE) estimates of the channel characteristics $\mathbb{E}\left(\mathbf{f}_{0:2Hn} \middle| y_{1:2Hn}\right)$ and the delays $\mathbb{E}\left(\theta_{0:2Hn} \middle| y_{1:2Hn}\right)$. This problem, unfortunately, does not admit any analytical solution and, thus, approximate methods must be employed. One of the methods that has proved to be useful in practice is particle filtering, and in the next section we propose a receiver based on the use of this technique.

7.2 Particle filtering receiver

The particle filtering receiver has already been designed in Chapters 4 and 5, although for the simpler case of symbol estimation alone. The problem considered here is more complicated since an additional continuous parameter is involved, and, in this section, the particle filtering algorithm for the joint estimation of all unknown parameters is detailed. We begin our treatment by incorporating a variance reduction technique, particularly, Rao-Blackwellisation, and then proceed with derivation of the particle filtering equations for the estimation of the required posterior distribution. Alternative deterministic and stochastic approaches are considered at the end of the section.

7.2.1 Rao-Blackwellisation

We follow a Bayesian approach here, and, given the measurements $y_{1:2Hn}$, base our inference on the joint posterior distribution

$$p\left(d_{1:n}, d\mathbf{f}_{0:2Hn}, d\theta_{0:2Hn}\right| y_{1:2Hn}) = p\left(d_{1:n}, \mathbf{f}_{0:2Hn}, \theta_{0:2Hn}\right| y_{1:2Hn}) \, d\mathbf{f}_{0:2Hn} d\theta_{0:2Hn}.$$

A straightforward application of particle filtering would focus on the estimation of this joint probability distribution, and, consequently, obtaining estimates of $d_{1:n}$, $\mathbf{f}_{0:2Hn}$, and $\theta_{0:2Hn}$ sequentially in time. It is beneficial, however, to improve the standard approach by making the most of the structure of the model and applying variance reduction techniques.

Indeed, similar to Chapter 4, the problem of estimating $p\left(d_{1:n}, d\mathbf{f}_{0:2Hn}, d\theta_{0:2Hn}\right| y_{1:2Hn})$ can be reduced to one of sampling from a lower-dimensional posterior $p\left(d_{1:n}, d\theta_{0:2Hn}\right| y_{1:2Hn})$. If the approximation of $p\left(d_{1:n}, d\theta_{0:2Hn}\right| y_{1:2Hn})$ can be obtained, say, via particle filtering:

$$\hat{p}_N\left(d_{1:n}, d\theta_{0:2Hn}\right| y_{1:2Hn}) = \sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(\{d_{1:n}, d\theta_{0:2Hn}\} - \left\{d_{1:n}^{(i)}, \theta_{0:2nH}^{(i)}\right\}\right), \tag{7.9}$$

one can compute the probability density $p\left(\mathbf{f}_{0:2Hn}\right| y_{1:2Hn}, d_{1:n}, \theta_{0:2Hn})$ using the Kalman filter associated with Equation (7.4, 7.6). As a result, the posterior $p\left(\mathbf{f}_{0:2Hn}\right| y_{1:2Hn})$ can be approximated by a random mixture of Gaussians

$$\begin{aligned}
&\hat{p}_N\left(\mathbf{f}_{0:2Hn}\right| y_{1:2Hn}) = \\
&\quad \int_{\theta_{0:2Hn}} \sum_{d_{1:n}} p\left(\mathbf{f}_{0:2Hn}\right| y_{1:2Hn}, d_{1:n}, \theta_{0:2Hn}) \, \hat{p}_N\left(d_{1:n}, \theta_{0:2Hn}\right| y_{1:2Hn}) \, d\theta_{0:2Hn} \\
&\quad = \textstyle\sum_{i=1}^{N} \tilde{w}_n^{(i)} p(\mathbf{f}_{0:2Hn}| \, y_{1:2Hn}, d_{1:n}^{(i)}, \theta_{0:2Hn}^{(i)})
\end{aligned}$$

leading to lower variance of the estimates and, therefore, increased algorithm efficiency [Doucet *et al.*, 2000].

Strictly speaking, we are interested in estimating the information symbols only with the tracking of the channel being naturally incorporated into the proposed algorithm. However, the MMSE (conditional mean) estimates of fading coefficients can, of course, be obtained if necessary as follows:

$$\begin{aligned}\mathbb{E}_{\hat{p}_N}\left[\mathbf{f}_{2H(n-1)+1:2Hn}\middle| y_{1:2Hn}\right] &= \int \mathbf{f}_{2H(n-1)+1:2Hn}\hat{p}_N\left(\mathbf{f}_{0:2Hn}\middle| y_{1:2Hn}\right) d\mathbf{f}_{0:2Hn} \\ &= \sum_{i=1}^{N} \tilde{w}_n^{(i)} \mathbb{E}\left[\mathbf{f}_{2H(n-1)+1:2Hn}\middle| y_{1:2Hn}, d_{1:n}^{(i)}, \theta_{0:2Hn}^{(i)}\right],\end{aligned}$$

with $\mathbb{E}\left[\mathbf{f}_{2H(n-1)+1:2Hn}\middle| y_{1:2Hn}, d_{1:n}^{(i)}, \theta_{0:2Hn}^{(i)}\right]$ being computed by the Kalman filter, with $2H$ steps required for each symbol transmitted.

7.2.2 Particle filtering algorithm

We can now proceed with the estimation of $p\left(d_{1:n}, d\theta_{0:2Hn}\middle| y_{1:2Hn}\right)$ using particle filtering techniques. The method is based on the following remark. Suppose N particles, $\left\{d_{1:n}^{(i)}, \boldsymbol{\theta}_{0:n}^{(i)}\right\}_{i=1}^{N}$, with $\boldsymbol{\theta}_n$ denoting

$$\boldsymbol{\theta}_n = \theta_{2H(n-1)+1:2Hn}, \text{ for } n = 1, 2, \ldots, \tag{7.10}$$

can be easily simulated according to a convenient importance distribution $\pi(d_{1:n}, d\boldsymbol{\theta}_{0:n}| \mathbf{y}_{1:n})$ (such that $p(d_{1:n}, d\boldsymbol{\theta}_{0:n}| \mathbf{y}_{1:n}) > 0$ implies $\pi(d_{1:n}, d\boldsymbol{\theta}_{0:n}| \mathbf{y}_{1:n}) > 0$).

Then, using the importance sampling identity, an estimate of $p\left(d_{1:n}, d\boldsymbol{\theta}_{0:n}\middle| \mathbf{y}_{1:n}\right)$ is given by the following point mass approximation:

$$\hat{p}_N\left(d_{1:n}, d\boldsymbol{\theta}_{0:n}\middle| \mathbf{y}_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta\left(\left\{d_{1:n}, d\boldsymbol{\theta}_{0:n}\right\} - \left\{d_{1:n}^{(i)}, \boldsymbol{\theta}_{0:n}^{(i)}\right\}\right), \tag{7.11}$$

where $\tilde{w}_n^{(i)}$ are the so-called normalized *importance weights*

$$\tilde{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_{1:n}^{(j)}}, \; w_n^{(i)} \propto \frac{p\left(d_{1:n}^{(i)}, \boldsymbol{\theta}_{0:n}^{(i)}\middle| \mathbf{y}_{1:n}\right)}{\pi\left(d_{1:n}^{(i)}, \boldsymbol{\theta}_{0:n}^{(i)}\middle| \mathbf{y}_{1:n}\right)}, \tag{7.12}$$

and $\mathbf{y}_n$ denotes similarly

$$\mathbf{y}_n = y_{2H(n-1)+1:2Hn}, \text{ for } n = 1, 2, \ldots \tag{7.13}$$

The distribution $\pi\left(d_{1:n}^{(i)}, \boldsymbol{\theta}_{0:n}^{(i)} \middle| \mathbf{y}_{1:n}\right)$ has to admit $\pi\left(d_{1:n-1}^{(i)}, \boldsymbol{\theta}_{0:n-1}^{(i)} \middle| \mathbf{y}_{1:n-1}\right)$ as a marginal distribution so that one can propagate this estimate sequentially in time without subsequently modifying the past simulated trajectories. The weights $w_n^{(i)}$ can also be updated on-line in this case:

$$w_n^{(i)} \propto w_{n-1}^{(i)} p\left(\mathbf{y}_n \middle| d_{1:n}^{(i)}, \boldsymbol{\theta}_{1:n}^{(i)}, \mathbf{y}_{1:n-1}\right) \times \frac{p\left(d_n^{(i)}, \boldsymbol{\theta}_n^{(i)} \middle| d_{n-1}^{(i)}, \boldsymbol{\theta}_{n-1}^{(i)}\right)}{\pi\left(d_n^{(i)}, \boldsymbol{\theta}_n^{(i)} \middle| d_{1:n-1}^{(i)}, \boldsymbol{\theta}_{0:n-1}^{(i)}, \mathbf{y}_{1:n}\right)}. \tag{7.14}$$

The sequential importance sampling described above is combined with a selection procedure introduced at each time step. This helps to avoid degeneracy of the algorithm by discarding particles with low normalized importance weights and multiplying those with high ones.

Given for the $(n-1)$th symbol N particles $\left\{d_{1:n-1}^{(i)}, \boldsymbol{\theta}_{0:n-1}^{(i)}\right\}_{i=1}^{N}$ distributed approximately according to $p\left(d_{1:n-1}, d\boldsymbol{\theta}_{0:n-1} \middle| \mathbf{y}_{1:n-1}\right)$, the general particle filtering receiver proceeds as follows:

Particle Filtering Algorithm

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $(\widetilde{d}_n^{(i)}, \widetilde{\boldsymbol{\theta}}_n^{(i)}) \sim \pi(d_n, \theta_n | d_{1:n-1}^{(i)}, \theta_{0:n-1}^{(i)}, \mathbf{y}_{1:n})$.
- For $i = 1, \ldots, N$, evaluate the importance weights $w_n^{(i)}$ up to a normalizing constant.
- For $i = 1, \ldots, N$, normalize $w_n^{(i)}$ to obtain $\tilde{w}_n^{(i)}$.

Selection Step

- Multiply/discard particles with respect to high/low $\tilde{w}_n^{(i)}$ to obtain N unweighted particles $\left\{d_{1:n}^{(i)}, \boldsymbol{\theta}_{1:n}^{(i)}\right\}_{i=1}^{N}$.

7.2.3 Implementation issues

The choice of importance distribution and selection scheme is discussed in [Doucet *et al.*, 2001b]; depending on those chosen, the computational complexity of the algorithm varies.

7.2.3.1 Importance density

Prior density. The simplest solution is to take the prior as an importance distribution, i.e.

$$\begin{aligned}\pi(d_n, \boldsymbol{\theta}_n | d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n}) &= p(d_n) p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) \\ &= p(d_n) \prod_{t=2H(n-1)+1}^{2Hn} p(\theta_t | \theta_{t-1}),\end{aligned}$$

then w_n becomes

$$w_n \propto p\left(\mathbf{y}_n | \mathbf{y}_{1:n-1}, d_{1:n}, \boldsymbol{\theta}_{0:n}\right) = \prod_{t=2H(n-1)+1}^{2Hn} p\left(y_t | d_{1:n}, \theta_{0:t}, y_{1:t-1}\right), \quad (7.15)$$

and requires evaluation of $2H$ one-step Kalman filter updates for each symbol.

Sequential Importance Sampling

(prior as an importance distribution)

- For $i = 1, \ldots, N$,

  sample $\widetilde{d}_n^{(i)} \sim p(d_n)$, set $w_n^{(i)} = 1$,

  For $t = 2H(n-1)+1, \ldots, 2Hn$,

  sample $\widetilde{\theta}_t^{(i)} \sim p(\theta_t | \theta_{t-1}^{(i)})$,

  perform one-step Kalman filter update

  $(w_n^{(i)} = w_n^{(i)} p\left(y_t | d_{1:n}, \theta_{0:t-1}^{(i)}, \widetilde{\theta}_t^{(i)}, y_{1:t-1}\right))$.

- For $i = 1, \ldots, N$, normalize $w_n^{(i)}$ to obtain $\tilde{w}_n^{(i)}$.

If H is long, it is useful to resample the particles at intermediate steps between $t = 2H(n-1)+1$ and $t = 2Hn$. One can also use Markov chain Monte Carlo (MCMC) steps to rejuvenate the particle and in particular d_n.

Suboptimal importance density. Of course, using the prior distribution in our case can be inefficient, as no information carried by the observations is used to explore the state space. The optimal choice, in a sense of minimizing the conditional variance of the importance weights [Doucet *et al.*, 2000], would consist of taking

$$\pi(d_n, \boldsymbol{\theta}_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n}) = p(d_n, \boldsymbol{\theta}_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n}),$$

as an importance density. From Bayes' rule $p(d_n, \boldsymbol{\theta}_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n})$ may be expressed as

$$p\left(d_n, \boldsymbol{\theta}_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n}\right) = \frac{p\left(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, d_n, \boldsymbol{\theta}_{0:n-1}, \boldsymbol{\theta}_n\right) p(d_n) p(\boldsymbol{\theta}_n |\, \boldsymbol{\theta}_{n-1})}{p(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1})}, \tag{7.16}$$

in which case,

$$\begin{aligned} w_n &= p(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}) \\ &= \int_{\breve{\boldsymbol{\theta}}_n} \sum_{m=1}^{M} \Big[p(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, d_m = m, \boldsymbol{\theta}_{0:n-1}, \breve{\boldsymbol{\theta}}_n) \\ &\qquad \times p(d_m = m) p\left(\breve{\boldsymbol{\theta}}_n \middle|\, \boldsymbol{\theta}_{n-1}\right) d\breve{\boldsymbol{\theta}}_n \Big], \end{aligned} \tag{7.17}$$

cannot be computed analytically. However, a suboptimal importance density of the following form

$$\pi(d_n, \boldsymbol{\theta}_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n-1}, \mathbf{y}_{1:n}) = p(d_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n}, \mathbf{y}_{1:n}) p(\boldsymbol{\theta}_n |\, \boldsymbol{\theta}_{n-1})$$

with

$$p\left(d_n |\, d_{1:n-1}, \boldsymbol{\theta}_{0:n}, \mathbf{y}_{1:n}\right) = \frac{p\left(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, d_n, \boldsymbol{\theta}_{0:n}\right) p(d_n)}{p(\mathbf{y}_n |\, \mathbf{y}_{1:n-1}, d_{1:n-1}, \boldsymbol{\theta}_{0:n})}, \tag{7.18}$$

can be employed instead, resulting in

$$\begin{aligned} w_n &\propto p(\mathbf{y}_n|\,\mathbf{y}_{1:n-1}, d_{1:n-1}, \boldsymbol{\theta}_{0:n}) \\ &= \sum_{m=1}^{M} p(\mathbf{y}_n|\,\mathbf{y}_{1:n-1}, d_{1:n-1}, d_m = m, \boldsymbol{\theta}_{0:n-1}, \boldsymbol{\theta}_n^{(i)}) p(d_m = m), \end{aligned} \tag{7.19}$$

where $\boldsymbol{\theta}_n^{(i)}$ is drawn from the prior Gaussian distribution with mean $\gamma\boldsymbol{\theta}_{n-1}^{(i)}$ and variance σ_θ^2 :

$$\boldsymbol{\theta}_n^{(i)} \sim \mathcal{N}\left(\gamma\boldsymbol{\theta}_{n-1}^{(i)}, \sigma_\theta^2\right) \qquad \text{for } i = 1, \ldots, N. \tag{7.20}$$

The importance weight $w_n^{(i)}$ in (7.19) does not actually depend on $d_n^{(i)}$, and the weights evaluation and selection steps can be done prior to the sampling of $d_n^{(i)}$ as follows:

Evaluation of Importance Weights

(suboptimal importance distribution)

- For $i = 1, \ldots, N$,

  For $m = 1, \ldots, M$, $w_n^{(i,m)} = 1$,

  For $t = 2Hn + 1, \ldots, 2H(n+1)$,

  sample $(\widetilde{\theta}_t^{(i)}) \sim p(\theta_t|\,\theta_{t-1}^{(i)})$,

  for $m = 1, \ldots, M$, perform one-step Kalman filter update $\left(w_n^{(i,m)} = w_n^{(i,m)} p\left(y_t|\,d_{1:n-1}^{(i)}, d_m = m, \theta_{0:t-1}^{(i)}, \widetilde{\theta}_t^{(i)}, y_{1:t-1}\right)\right)$.

- Evaluate the importance weight $w_n^{(i)}$ up to a normalizing constant:

$$w_n^{(i)} \propto \sum_{m=1}^{M} w_n^{(i,m)} p(d_m = m),$$

- For $i = 1, \ldots, N$, normalize $w_n^{(i)}$ to obtain $\tilde{w}_n^{(i)}$.

For each symbol detection, this procedure requires the evaluation of the M $2H$-step ahead Kalman filters, which is quite computationally expensive. Further research should, therefore, concentrate on development of other more efficient suboptimal importance distributions on a case by case basis.

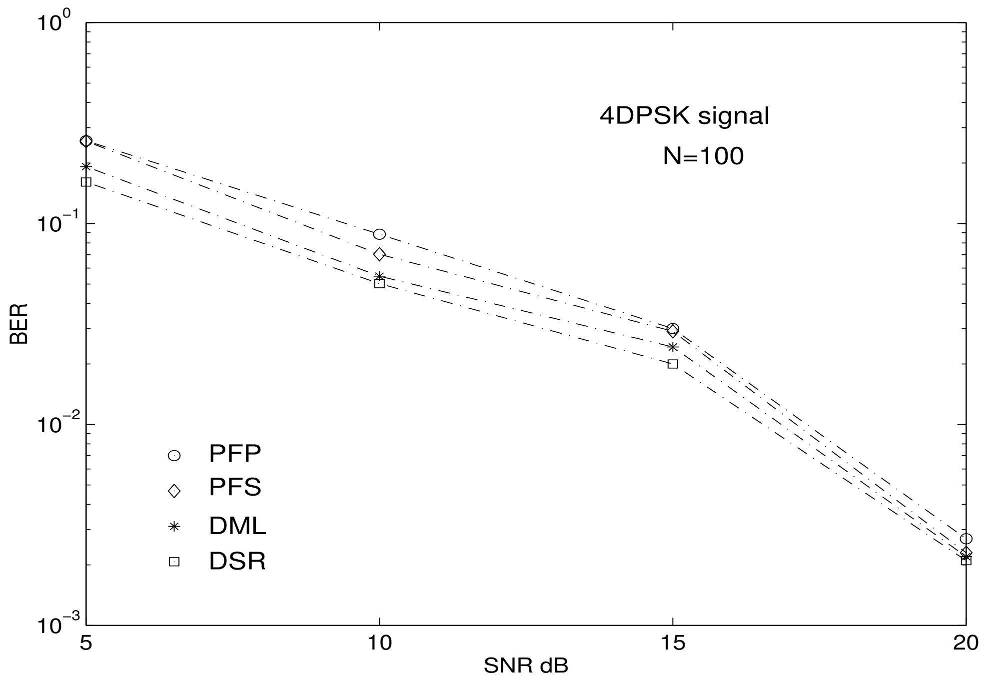

Figure 7.1: Bit-error-rate performance.

7.2.3.2 Selection

As far as the selection step is concerned, a stratified sampling scheme [Kitagawa, 1996] is employed in this work. The algorithm is discribed in details in Chapter 3, and as was already pointed out, has the minimum variance one can achieve in the class of unbiased schemes [Crisan, 2001], and can be implemented in $O(N)$ operations.

7.2.4 Deterministic Particle Filter

The use of the suboptimal importance distribution described in the previous section increases the efficiency of the algorithm in comparison with the

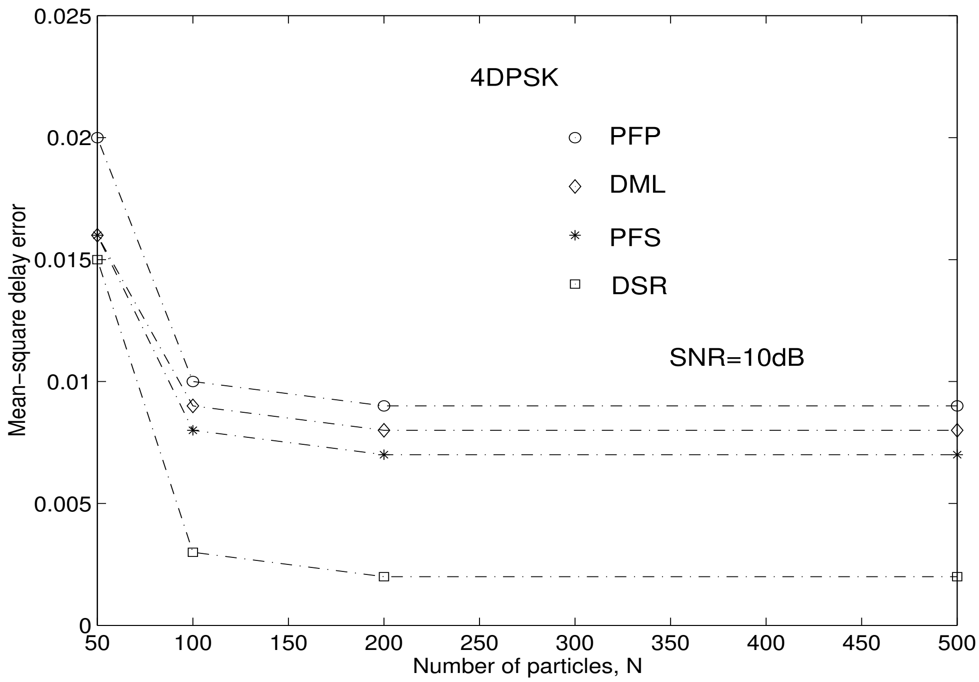

Figure 7.2: Mean-square delay error for the different number of particles.

standard approach using the prior. However, as shown in Chapter 6, if one already opts for (7.19), and all the calculations have to be performed anyway, it might be better to base our approximation of $p\left(d_{1:n}, d\boldsymbol{\theta}_{0:n}\middle|\, \mathbf{y}_{1:n}\right)$ directly on

$$\hat{p}_{N\times M}\left(d_{1:n}, d\boldsymbol{\theta}_{0:n}\middle|\,\mathbf{y}_{1:n}\right) = \sum_{i=1}^{N}\sum_{m=1}^{M}\tilde{w}_n^{(i,m)}\delta\left(\{d_{1:n}, d\boldsymbol{\theta}_{0:n}\} - \left\{d_{1:n-1}^{(i)}, d_n = \mu_m, \boldsymbol{\theta}_{0:n-1}^{(i)}, \boldsymbol{\theta}_n^{(i)}\right\}\right),$$

with corresponding weights $\tilde{w}_n^{(i,m)}$ being equal to

$$w_n^{(i,m)} \propto \tilde{w}_{n-1}^{(i)} p(\mathbf{y}_n|\,\mathbf{y}_{1:n-1}, d_{1:n-1}^{(i)}, d_m = m, \boldsymbol{\theta}_{0:n-1}^{(i)}, \boldsymbol{\theta}_n^{(i)}) p(d_m = m),$$

and $\boldsymbol{\theta}_n^{(i)}$ still drawn from its prior (7.20). Indeed, all possible "extensions" of the existing state sequences at each step n are considered in this case, and one does not discard unnecessarily any information by selecting randomly one path out of the M available. In the above expression, $\tilde{w}_{n-1}^{(i)}$ is the weight

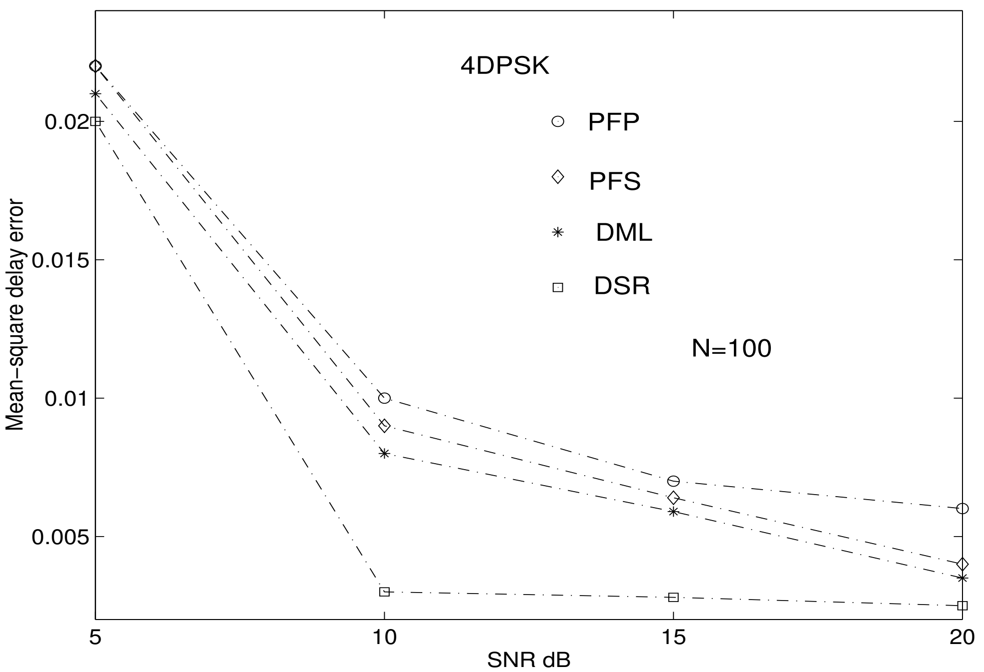

Figure 7.3: Mean-square delay error via SNR.

of the "parent" particle, which has M "offspring" instead of the usual one, resulting in a total number of $N \times M$ particles at each stage. This number increases exponentially with time, and, therefore, a selection procedure has to be employed at each step n.

The simplest way to perform such selection is to choose the N most likely offspring and discard the others (as, for example, in [Tugnait & Haddad, 1979]); Chapter 6 shows the superiority of this approach over other methods in the fully discrete framework. A more complicated procedure involves preserving the particles with high weights and resampling the ones with low weights, thus reducing their total number to N. An algorithm of this type is presented in [Fearnhead, 1998] but other selection schemes can be designed. Contrary to the case involving the discrete parameters only, in this scenario a resampling scheme with replacement could be employed, since $\boldsymbol{\theta}_n^{(i)}$ is chosen randomly. Therefore, stratified resampling could be used in order to select N particles from $N \times M$ available.

Whether we choose to preserve the most likely particles, employ the se-

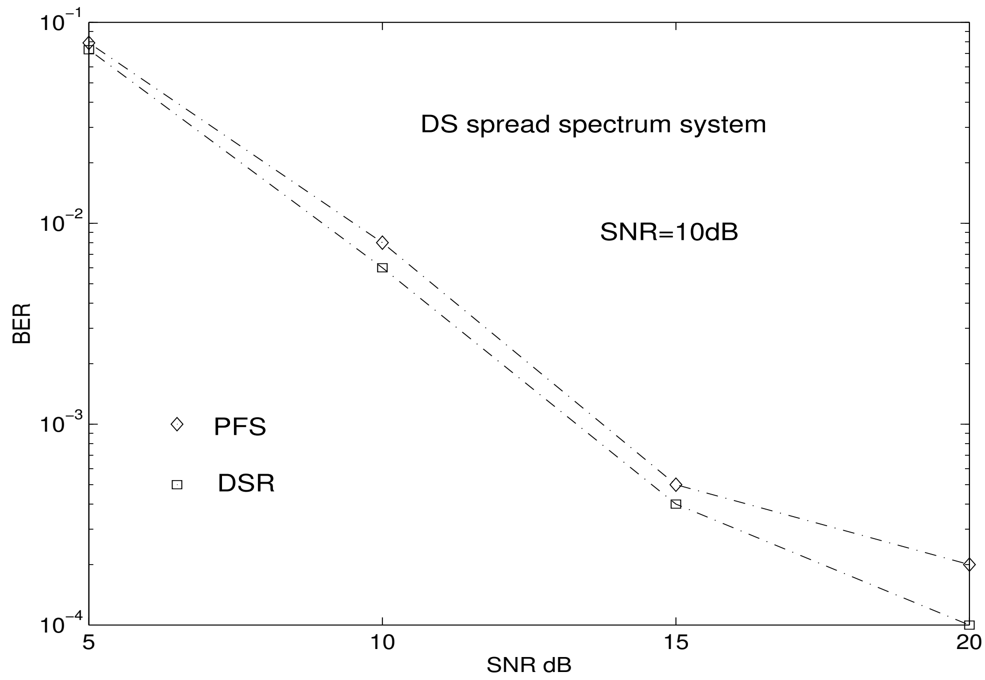

Figure 7.4: Bit error rate.

lection scheme proposed in [Fearnhead, 1998], or stratified resampling, the computational load of the resulting algorithms at each time step n is that of $N \times M \times 2H$ Kalman filters, and the selection step in first two cases is implemented in $O(N \times M \log N \times M)$ operations. Of course, if M is large, which is the case in many applications, all these methods are too computationally extensive to be used, and one should resort to a standard particle filter.

7.3 Simulation Results

In the following experiments the bit-error-rate (BER) and the tracking delay error $\left(\theta_t - \hat{\theta}_t\right)$ were evaluated by means of computer simulations. Gray-encoded M-ary differential phase shift keyed (MDPSK) signals were employed with mapping function

$$s_n = \exp(j\phi_n), \text{ and } \phi_n = \sum_{j=1}^{n} \sum_{m=1}^{M} \frac{2\pi m}{M} \delta\left(d_j - r_m\right).$$

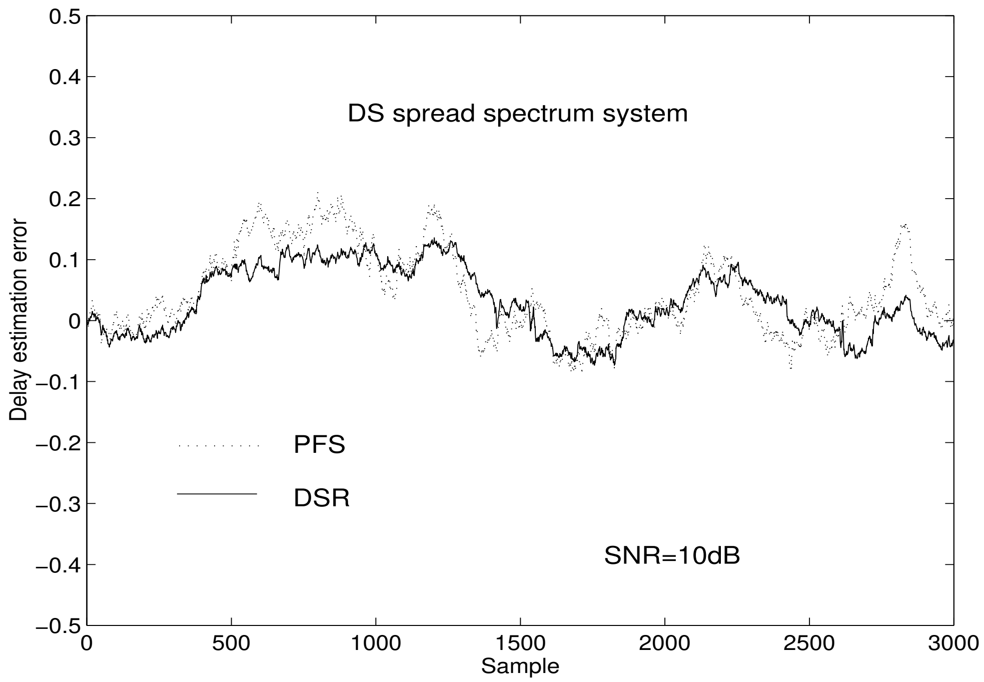

Figure 7.5: The error in delay estimation. SNR=10 dB.

In order to assess the performance of the proposed approaches we first applied them to a simpler case of synchronization under flat fading condition, $G = 1$, for a system with no spectrum spreading employed, $c_1 = 1$, $H = 1$. In the first experiment, 4DPSK signals were considered with the average signal to noise ratio (SNR) varying from 5 to 20dB. The AR coefficients for the channel, Equation (7.4), were set to $\alpha^{(0)} = 0.999$, $\sigma_f^{(0)} = 0.01$, and the delay model parameters in Equation (7.5) were chosen to be the same, $\gamma = 0.999$ and $\sigma_\theta = 0.01$. The bit-error-rate (BER) obtained by the particle filtering receiver employing prior (PFP) and suboptimal (PFS) importance distributions, and the deterministic receiver preserving N most likely particles (DML) and using stratified resampling (DSR) is presented in Figure 7.1. The number of particles used in these algorithms was equal to $N = 100$, and little or no improvement in BER was gained by increasing this number for deterministic schemes. For the randomized approaches, the number of particles required to achieve the BER of DSR algorithm was equal to $N = 1200$. In Figure 7.2, the mean-square delay error (MSE) is presented as a function

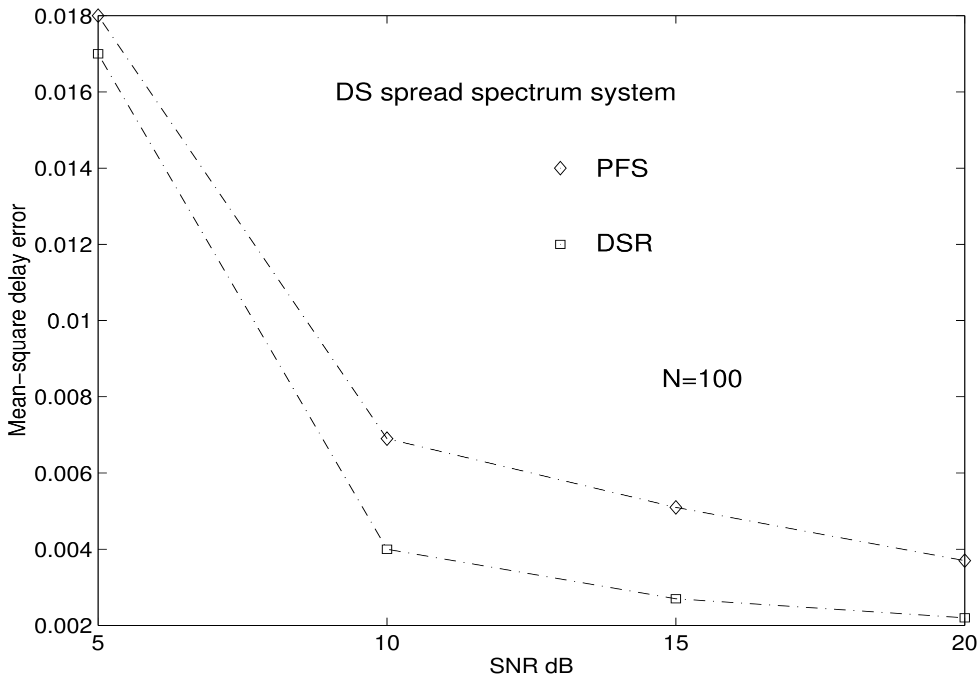

Figure 7.6: Mean-square delay error via SNR.

of the number of particles N for SNR=10dB:

$$\hat{\theta}_{MSE} = \frac{1}{2HL_d} \sum_{n=1}^{2HL_d} \left(\theta_n - \hat{\theta}_n\right)^2 ,$$

where L_d is a length of the symbol sequence, $L_d = 1000$. The results for the different signal to noise ratios (SNR) are given in Figure 7.3. As one can see, the deterministic particle filter with stratified resampling slightly outperforms the receiver selecting most likely particles, and is more efficient than both standard particle filtering schemes.

In the second experiment we applied this algorithm to perform joint symbols/channel coefficients/code delay estimation for DS spread spectrum systems with $H = 15$, $G = 4$. A binary DPSK modulation scheme was employed with the multipath channel response and AR coefficients chosen as in [Iltis, 1990, channel B]. As it is shown in Figure 7.4, the algorithm employing 100 particles exhibits good bit-error-rate (BER) performance. A tracking error trajectory for 100 information symbols (corresponding to 1500 chips and 3000 channel samples) and an average signal to noise ratio (SNR) equal to 10 dB is presented in Figure 7.5. Figure 7.6 also illustrates the mean-square delay error as a function of SNR for $L_d = 1000$.

In this chapter, we proposed the application of particle filtering techniques to a challenging problem of joint symbols, channel coefficients and code delay estimation for DS spread spectrum systems in multipath fading. The algorithm was designed to make use of the structure of the model, and incorporated a variance reduction technique. The work was based on the results on the superiority of the DML approach in a fully discrete environment obtained in Chapter 6. The method could not be applied straightforwardly, however, and several procedures combining both deterministic and randomized schemes were considered. The algorithms were tested and compared. Although computer simulations showed that all methods were capable of providing good performance, in this particular case involving additional continuous-valued parameters, the deterministic scheme employing stratified resampling turned out to be the most efficient one. The choice of the algorithm might, however, be application dependent so further investigation is necessary. The receiver can be easily extended to address multiuser DS CDMA transmission, or simplified to consider the channel tracking only since it is naturally incorporated in the proposed algorithm. Future research should concentrate on the development of suboptimal importance distributions and selection schemes capable of increasing the algorithm efficiency.

8 Conclusion

The revolution in global communications is leading to unprecedented access to information and knowledge. Affordability creates accessibility, and as the number of users and the demand for services grows, so does the need for efficient signal processing techniques capable of coping with the increasingly difficult communication environment. Research into new methodologies pushing the bounds of performance provides a means of meeting the challenging requirements of the future, and ensures that our capabilities remain at the cutting edge of progress. In this thesis, the use in digital communications of one of such promising and fairly recent advanced technology - Sequential Monte Carlo methods - was investigated.

We aimed at developing a general framework for addressing a variety of problems arising in the field, and considered several specific tasks as an illustration of the general approach. Derivation of efficient particle filtering algorithms based on the structure and particular features of the model was the other our objective.

We presented a general model of a modern digital communication system at the beginning of the thesis, reviewed currently existing algorithms, and started our investigation with the simplest scenario of demodulation in flat Rayleigh fading conditions.

An efficient particle filtering demodulator was, first, developed, which took into account the structure of the model, and combined variance reduction strategies based on Kalman filtering and the use of optimal importance distribution. The ability to incorporate the tracking of the channel into the algorithm, and straightforward extension to fixed-lag smoothing in order to obtain more accurate estimates, are two of the advantages of the proposed approach. Although, in general, SMC techniques tend to be quite computationally demanding, a relatively small number of particles was required for this scenario; moreover, the use of parallel processors for the implementation was also a possibility. In addition, the particle filter allowed to significantly reduce the number of the inserted pilot symbols required when the differential encoding was not employed. The bit-error-rate results for both Gaussian and non-Gaussian additive noise and a variety of linear modulation schemes showed that the use of the developed receiver can significantly improve the communication quality of the transmission in fading conditions.

We then demonstrated how easily the algorithm could be extended to treat more complicated scenarios in the example of joint symbol decoding and demodulation, space diversity combining and multiuser detection. Although in all these cases the particle filtering receiver outperformed the methods routinely used in communication applications, the issues connected to computational complexity of the approach arose, and some other applications of the SMC were suggested as one of the future research directions. Another disadvantage of the algorithm, was, of course, its inability to take interleaving into consideration.

The problems considered up to this point involved the discrete unknown parameters only and belonged to the class of jump Markov linear systems. Several other deterministic and randomized schemes were applicable in this framework, and their review and comparison was, probably, the key contribution of this thesis. To the best of our knowledge, such comparison had never been made before, and the results were surprising. The basic deterministic approach preserving the limited number of most likely particles (DML) turned out to be the most efficient one. This does not mean that the SMC methods are of no use in communication systems. Indeed, all schemes were capable of providing optimal performance given a large number of particles. In addition, the DML would be too computationally expensive to apply in a number of different scenarios, whereas the particle filter with certain improvements is still relevant. These results, however, are important for many other applications involving the same class of models not only in digital communications but also in the other areas, and, therefore, should be investigated further.

Moreover, other algorithms based on the SMC techniques could be designed taking these results into account. We proposed such an algorithm at the end of the thesis where the joint code delay estimation and symbol detection for direct sequence spread spectrum systems were considered. This problem required the estimation of both discrete and continuous-valued unknown parameters, and, therefore, it was impossible to apply the deterministic schemes straightforwardly. We suggested, however, several procedures combining both deterministic and randomized approaches, which were tested and compared. The results were in favour of the deterministic particle filter employing stratified resampling, which demonstrated excellent performance in comparison with other methods, and enabled performing of efficient symbol detection in difficult multipath conditions.

Future work

Some directions for future research and application of this work have already been highlighted throughout this thesis. To conclude, however, we would like to briefly identify three main areas where further investigation would be of significant interest:

- **Applications.** First of all, of course, are the multiple extensions and simplifications of the proposed algorithm to consider a large number of digital communication applications for which the use of particle filtering might be beneficial. There is an unlimited number of examples, including, for instance, demodulation in asynchronous CDMA, or simple synchronization (with information symbols assumed to be known). The model presented in Chapter 2 itself could be significantly extended to take into account, for example, Doppler effects or narrowband interferences (typical for underwater acoustic channels). In addition, some of the parameters previously assumed known (such as ARMA model coefficients for the channel characteristics) could be estimated. For multiuser detection, the application of the SMC methods in a different framework, involving the estimation of symbol from one source at a time, should be studied. An efficient deterministic particle filter developed in Chapter 7 could also be applied in other fields of engineering and science, such as genetics, economics and finances, to name a few.

- **Algorithm efficiency.** The performance of the algorithm could be significantly improved if a suboptimal importance distribution allowing more efficient sampling could be developed. At the moment, however, there is no general tool allowing us to design such importance distributions, and the problem has to be addressed on a case by case basis, depending on the task one is addressing. Some suggestions are given at the end of Chapter 6. Another area for improvement is development of other variance reduction techniques (in addition to Rao-Blackwellisation) in the framework of SMC. Quasi-Monte Carlo methods and stratification are some of them.

- **Deterministic schemes.** Further research should be directed at investigation of the deterministic schemes. One of the improvements could be the use of the discount parameter (Chapter 6), and consequent optimization of the algorithm with respect to it. Other, more efficient, selection schemes could also be proposed. The ideas initiated in Chapter 6 and 7 could be incorporated into the design of efficient particle filters for continuous-valued parameter estimation.

While this thesis was in preparation, a great deal of work has appeared in the literature on the application of particle filtering, [Dong *et al.*, 2003; Iltis, 2003; Huang & Djuric, 2002; Ghirmai *et al.*, 2003] among others, and the number is rapidly growing. Although there is still a lot to be done, one may only hope that all these efforts will eventually help to increase the ease and breadth of communicating.

References

Ackerson, G. & Fu, K. (1970). On state estimation in switching environments. *IEEE Transactions on Automatic Control*, **15**, 10–17.

Akashi, H. & Kumamoto, H. (1977). Random sampling approach to state estimation in switching environments. *Automatica*, **13**, 429–434.

Alspach, D. & Sorenson, H. (1972). Non-linear Bayesian estimation using Gaussian sum approximation. *IEEE Transactions on Automatic Control*, **17**, 439–447.

Anderson, B. & Moore, J. (1979). *Optimal FIltering*. Prentice-Hall, Englewood Cliffs.

Bello, P. (1963). Characterization of randomly time-variant linear channels. *IEEE Transactions on Communications Systems*, **11**, 360–393.

Berzuini, C., Best, N., Gilks, W. & Larizza, C. (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of American Statistical Association*, **92**, 1403–1412.

Biglieri, E., Proakis, J. & Shamai(Shitz), S. (1998). Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, **44**, 2619–2692.

Blom, H. & Bar-Shalom, Y. (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions on Automatic Control*, **33**, 780–783.

Bucy, R. & Youssef, H. (1974). Nonlinear filter representation via spline functions. In *Proceedings of the 5th Symposium on Nonlinear Estimation*, 51–60.

Carpenter, J., Clifford, P. & Fearnhead, P. (1999). An improved particle filter for non-linear problems. *IEE proceedings - Radar, Sonar and Navigation*, **146**, 2–7.

Casella, G. & Robert, C. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika*, **83**, 81–94.

Cavers, J. (1991). An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Transactions on Vehicular Technology*, **40**, 686–693.

Chan, K. & Bateman, A. (1992). The performance of reference-based M-ary PSK with trellis coded modulation in Rayleigh fading. *IEEE Transactions on Vehicular Technology*, **41**, 190–198.

Chen, R., Wang, X. & Liu, J. (2000). Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering. *IEEE Transactions on Information Theory*, **46**, 2079–2094.

Collings, I. & Moore, J. (1994). An HMM approach to adaptive demodulation of QAM signals in fading channels. *International Journal of Adaptive Control and Signal Processing*, **8**, 457–474.

Collings, I. & Moore, J. (1995). An adaptive hidden Markov model approach to FM and M-ary DPSK demodulation in noisy fading channels. *Signal Processing*, **47**, 71–84.

Crisan, D. (2001). Particle filters - A theoretical perspective. In *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, eds. A. Doucet, J.F.G. de Freitas and N.J. Gordon.

Crisan, D. & Doucet, A. (2000). Convergence of sequential Monte Carlo methods. Tech. Rep. CUED/F-INFENG/TR. 381, Cambridge University Engineering Department.

Crisan, D., Moral, P.D. & Lyons, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields*, **5**, 293–318.

Dai, Q. & Schwedyk, E. (1994). Detection of bandlimited signals over frequency selective Rayleigh fading channels. *IEEE Transactions on Communications*, **42**, 941–950.

Daum, F. (1988). New exact nonlinear filters. In *Bayesian Analysis of Time Series and Dynamic Models*, Dekker, ed. J.C. Spall.

Davis, L. & Collings, I. (1999). Multi-user MAP decoding for flat-fading CDMA channels. In *Proceedings of the 5th International Conference on Digital Signal Processing for Communication Systems*, 79–86.

Dong, B., Wang, X. & Doucet, A. (2003). A new clas of soft MIMO demodulation algorithms. *IEEE Transactions on Signal Processing*, to appear.

Doucet, A., Godsill, S. & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197–208.

Doucet, A., de Freitas, J. & Gordon, N., eds. (2001a). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag: New-York.

Doucet, A., Gordon, N. & Krishnamurthy, V. (2001b). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, **49**, 613–624.

Du, J. & Vucetic, B. (1991). New 16QAM trellis codes for fading channels. *Electronics Letters*, **27**, 1009–1010.

Fagin, S. (1964). Recursive linear regression theory, optimal filter theory, and error analysis. *IEEE International Convention Record*, **12**, 216–240.

Fearnhead, P. (1998). *Sequential Monte Carlo methods in filter theory*. Ph.D. thesis, University of Oxford.

Fitzgerald, R. (1968). Error divergence in optimal filtering problems. In *Proceedings of the 2nd IFAC symposium AC space*, Vienna, Austria.

Forney, G. (1972). Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interferences. *IEEE Transactions on Information Theory*, **18**, 363–378.

Georghiades, C. & Han, J. (1997). Sequence estimation in the presence of random parameters via the EM algorithm. *IEEE Transactions on Communications*.

Gertsman, M. & Lodge, J. (1997). Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels. *IEEE Transactions on Communications*, **45**, 788–799.

Geweke, J. (1989). Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica*, **57**, 1317–1339.

Ghirmai, T., Bugallo, M., Miguez, J. & Djuric, P. (2003). Joint symbol detection and timing estimation usign particle filtering. In *Proceedings ICASSSP 2003*.

Gilks, W. & Berzuini, C. (1998). Monte Carlo inference for dynamic Bayesian models, Medical Research Council, Cambridge, UK.

Godsill, S. & Clapp, T. (2001). Improvement strategies for Monte Carlo particle filters. In *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, eds. A. Doucet, J.F.G. de Freitas and N.J. Gordon.

Gordon, N. (1994). *Bayesian Methods for Tracking*. Ph.D. thesis, Imperial College, University of London.

Gordon, N., Salmond, D. & Smith, A. (1993). Novel approach to nonlinear/non-gaussian Bayesian state estimation. *IEE Proceedings-F*, **140**, 107–113.

Gordon, N.J. & Whitby, A. (1995). Bayesian approach to target tracking in the presence of glint. In *Proceedings of the SPIE Signal and Data Processing of Small Targets*, vol. 2561, 472–483.

Haeb, R. & Meyr, H. (1989). A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels. *IEEE Transactions on Communications*, **37**, 748–759.

Handschin, J. & Mayne, D. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control*, **9**, 547–559.

Higuchi, T. (1995a). Kitagawa Monte Carlo filter from the perspective of genetic algorithm. Research Memorandum, The Institute of Statistical Mathematics, Tokyo, Japan.

Higuchi, T. (1995b). Kitagawa Monte Carlo filter using the genetic algorithm operators. Research Memorandum, The Institute of Statistical Mathematics, Tokyo, Japan.

Hoeher, P. (1992). A statistical discrete-time model for the WSSUS multipath channel. *IEEE Transactions on Vehicular Technology*, **41**, 461–468.

Hoeher, P. & Lodge, J. (1999). Turbo DPSK: Iterative differential PSK demodulation and channel decoding. *IEEE Transactions on Communications*, **47**, 837–843.

Hou, W. & Chen, B. (2000). Adaptive detection in asynchronous code-division multiple access systems in multipath fading channels. *IEEE Transactions on Communications*, **48**, 863–873.

Huang, Y. & Djuric, P. (2002). A blind particle filtering detector for for joint channel estimation, tracking and data detection over flat fading channels. In *Proceedings EUSIPCO 2002*.

Hurzeler, M. & Kunsch, H.R. (1998). Monte Carlo approximations for general state space models. *Journal of Computational and Graphical Statistics*, **7**, 175–193.

Iltis, R. (1990). Joint estimation of PN code delay and multipath using the Extended Kalman Filter. *IEEE Transactions on Communications*, **38**, 1677–1685.

Iltis, R. (2001). A DS-CDMA tracking mode receiver with joint channel/delay estimation and MMSE detection. *IEEE Transactions on Communications*, **49**, 1770–1779.

Iltis, R. (2003). A Sequential Monte Carlo Filter for joint linear/nonlinear state estimation with application to DS-CDMA. *IEEE Transactions on Signal Processing*, to appear.

Jakes, W. (1974). *Microwave Mobile Communications*. Wiley, New York.

Jazwinski, A. (1973). *Stochastic Processes and Filtering Theory*. Academic Press.

Kalman, R. & Bucy, R. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions ASME Series D*, **83**, 95–108.

Kam, P. & Ching, H. (1992). Sequence estimation over the slow nonselective Rayleigh fading channel with diversity reception and its application to Viterbi decoding. *IEEE Journal on Selected Areas in Communications*, **COM-10**, 562–570.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of non-stationary time series (with discussions). *Journal of the American Statistical Association*, **82**, 1032–1063.

Kitagawa, G. (1993). A Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space models. In *Proceedings of the 2nd US-Japan Joint Seminar on Statistical Time Series Analysis*, 110–131, Honolulu, Hawaii.

Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the gaussian-sum smoother. *Annals Institute of Statistical Mathematics*, **46**, 605–623.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, **5**, 1–25.

Komninakis, C., Fragouli, C., Sayed, A. & Wesel, R.D. (1999). Channel estimation and equalization in fading. In *Proceedings of the 33rd Asilomar*

Conference on Signals, Systems, and Computers, vol. 2, 1159–1163, Pacific Grove, CA.

Kong, A., Liu, J. & Wong, W. (1994). Sequential imputations and Bayesian missing data problems. *Journal American Statistical Association*, **89**, 278–288.

Künsch, H. (2003). Recursive Monte Carlo filters: Algorithms and theoretical analysis. Tech. Rep. 112, Eidgenossische Technische Hochschule, Switzerland.

Lin, S. & Costello, D. (1983). *Error Control Coding: Fundamentals and Applications*. Prentice Hall.

Liu, J. (1996). Metropolised independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, **6**, 113–119.

Liu, J. & Chen, R. (1995). Blind deconvolution via sequential imputation. *Journal of the American Statistical Association*, **90**, 567–576.

Liu, J. & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93**, 1032–1044.

Liu, J. & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, eds. A. Doucet, J.F.G. de Freitas and N.J. Gordon.

Liu, Y. & Blostein, S. (1995). Identification of frequency non-selective fading channels using decision feedback and adaptive linear prediction. *IEEE Transactions on Communications*, **43**.

Lodge, J. & Moher, M. (1990). Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels. *IEEE Transactions on Communications*, **38**.

Moral, P.D. & Salut, G. (1995). Non-linear filtering using Monte Carlo particle methods. *C.R. Academy of Science Paris*, **320**, 1147–1152, (in French).

Paskov, S. & Traub, J. (1995). Faster valuation of financial derivatives. *Journal of Portfolio Management*, **22**, 113–120.

Pitt, M. & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, **94**, 590–599.

Pitt, M. & Shephard, N. (2001). Auxiliary variable based particle filters. In *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, eds. a. doucet, j.f.g. de freitas and n.j. gordon edn.

Proakis, J. (1995). *Digital Communications*. McGraw Hill, 3rd edn.

Raphaeli, D. (2000). Suboptimal Maximum-Likelihood multiuser detection of synchronous CDMA on frequency-selective multipath channels. *IEEE-TransComs*, **48**, 875–885.

Rappaport, T. (1996). *Wireless Communications: Principles and Practice*. Prentice Hall.

Ripley, B. (1987). *Stochastic Simulation*. Wiley, New York.

Robert, C.P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

Rubin, D. (1988). Using the SIR algorithm to simulate the posterior distributions. In *Bayesian Statistics 3*, 395–402, Oxford: Oxford University Press, eds. J.M. Bernado, M.H. DeGroot, D.V. Lindley and A.F.M. Smith.

Sampei, S. & Sunaga, T. (1993). Rayleigh fading compensation for QAM in land mobile communications. *IEEE Transactions on Vehicular Technology*, **42**, 137–147.

Seymour, J. (1994). *Improved synchronization in the mobile communication environment*. Ph.D. thesis, Purdue University.

Seymour, J. & Fitz, M. (1995). Near-optimal symbol-by-symbol detection schemes for flat rayleigh fading. *IEEE Transactions on Communications*, **43**, 1525–1533.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, UK.

Sorenson, H. (1988). Recursive estimation for nonlinear dynamic systems. In *Bayesian Analysis of Time Series and Dynamic Models*, Dekker, ed. J.C. Spall.

Sorenson, H. & Alspach, D. (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica*, **7**, 465–479.

Sorenson, H. & Stubberud, A. (1968). Recursive filtering for systems with small but nonnegligible nonlinearities. *International Journal of Control*, **7**, 271–280.

Stuber, G. (1996). *Principles of mobile communication*. Norwell, MA: Kluwer.

Torrance, J. & Hanzo, L. (1995). Comparative study of pilot symbol assisted modem schemes. In *Proceedings of the Radio Receivers and Associated Systems Conference (RRAS)*, 36–41, Bath, UK.

Tugnait, J. (1982). Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control*, **25**, 1054–1065.

Tugnait, J. & Haddad, A. (1979). A detection-estimation scheme for state estimation in switching environment. *Automatica*, **15**, 477–481.

Turin, W. & van Nobelen, R. (1998). Hidden Markov modeling of flat fading channels. *IEEE Journal on Selected Areas in Communications*, **16**, 1809–1817.

Verdu, S. (1998). *Multiuser Detection*. Cambridge University Press.

Viterbi, A. (1995). *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley.

Vitetta, G. & Taylor, D. (1995). Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels. *IEEE Transactions on Communications*, **43**, 2750–2758.

West, M. & Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 2nd edn.

Wishner, R., Tabaczynski, J. & Athans, M. (1968). On the estimation of the state of noisy nonlinear multivariate systems. In *Proceedings of the IFAC symposium Multivariate Control Systems*, Dusseldorf, Germany.

Wishner, R., Athans, M. & Tabaczynski, J. (1969). A comparison of three filters. *Automatica*, **5**, 487–496.

Wolf, H. (1968). An assessment of the work in minimum variance orbit determination. Tech. Rep. 68-3, Analytical Mechanics Associates Inc., Westbury, New York.

Yang, Z. & Wang, X. (2002). Blind detection of OFDM signals in multipath fading channels via Sequential Monte Carlo. *IEEE Transactions on Signal Processing*, **50**, 271–280.

Zaritskii, V., Svetnik, V. & Shimelevich, L. (1975). Monte Carlo technique in problems of optimal data processing. *Automation and Remote Control*, **12**, 95–103.