Applications of particle filtering to selected problems in communications

Petar M. Djurić, Jayesh H. Kotecha, Jianqui Zhang, Yufei Huang, Tadesse Ghirmai,

Mónica F. Bugallo, Joaquín Míguez

This work was supported by the National Science Foundation under Awards CCR-9903120 and CCR-0082607. Joaquín Míguez acknowledges the support of Xunta de Galicia, the Ministry of Science and Technology of Spain and the FEDER funds of the European Union (project TIC2001-0751-C04-01).

P. M. Djurić, T. Ghirmai, and M. F. Bugallo, are with the Department of Electrical and Computer Engineering at Stony Brook University, Stony Brook, NY, 11794; J. H. Kotecha is with the Department of Electrical and Computer Engineering at the University of Wisconsin at Madison, Madison; J. Zhang is with the Department of Electrical and Computer Engineering at the University of New Hamshire; Y. Huang is with Department of Electrical and Computer Engineering at the University of Texas at San Antonio, San Antonio, TX; Joaquín Míguez is with the Department of Electrónica e Sistemas, Universidade da Coruña, Facultade de Informática, 15071 A Coruña, Spain. E-mail of first author: djuric@ece.sunysb.edu

Abstract

Recent developments have demonstrated that particle filtering is an emerging and powerful methodology for sequential signal processing with a wide range of applications in science and engineering. It is based on the concept of sequential importance sampling and the use of Bayesian theory, and it is particularly useful in dealing with nonlinear and non-Gaussian problems. The underlying principle of the methodology is the approximation of relevant distributions with random measures composed of particles (samples from the space of the unknowns) and their associated weights. In this paper, first we present a brief review of the particle filtering theory and then we show how it can be used for resolving many communication problems. We demonstrate its application to blind equalization, blind detection over flat fading channels, multiuser detection, and estimation and detection of space-time codes in fading channels.

I. Introduction

Particle filtering is a technique for sequential signal processing that recently has captured the attention of many researchers in various communities including those of signal processing, statistics, and econometrics. The interest in this methodology stems from its potential for coping with difficult nonlinear and/or non-Gaussian problems.

Particle filtering is a sequential Monte Carlo methodology where the basic idea is the recursive computation of relevant probability distributions using the concepts of importance sampling and approximation of probability distributions with discrete random measures. The earliest applications of sequential Monte Carlo methods were in the area of growing polymers [19], [49], and later they expanded to other fields including physics and engineering. Sequential Monte Carlo methods found limited use in the past, except for the last decade, primarily due to their very high computational complexity and the lack of adequate computing resources of the time. The fast advances of computers in the last several years and the outstanding potential of particle filters have made them recently a very active area of research. Their potential for parallel implementation represents additional impetus for their development. The current interest in particle filtering for signal processing applications was brought on by [17]. Recent reviews and accounts of new developments on the subject can be found in [3], [10], [11].

A large portion of the theory on sequential signal processing is about signals and systems that are represented by state-space and observation equations, that is, equations of the form

$$\begin{aligned} \mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{u}_t) \\ \mathbf{y}_t &= \mathbf{g}_t(\mathbf{x}_t, \mathbf{v}_t) \end{aligned} \tag{1}$$

where $\mathbf{y}_t$ is a vector of observations, $\mathbf{x}_t$ is a state vector, $\mathbf{g}_t(\cdot)$ is a measurement function, $\mathbf{f}_t(\cdot)$ is a system transition function, $\mathbf{u}_t$ and $\mathbf{v}_t$ are noise vectors, and the subscript t denotes time index. The first equation is known as state equation, and the second, as measurement equation. The standard assumptions are that the analytical forms of the functions and the distributions of the two noises are known. Based on the observations $\mathbf{y}_t$ and the assumptions, the objective is to estimate $\mathbf{x}_t$ recursively.

The method that has been investigated the most and that has been most frequently applied in practice is the Kalman filter [1]. The Kalman filter is optimal in the important case when the equations are linear and the noises are independent, additive and Gaussian. In this situation, the distributions of interest (filtering, predictive, or smoothing) are also Gaussian and the Kalman filter can compute them exactly without approximations. For scenarios where the models are nonlinear or the noise is non-Gaussian, various approximate methods have been proposed of which the extended Kalman filter is perhaps the most prominent of all [1].

The particle filtering method has become an important alternative to the extended Kalman filter. With particle filtering, continuous distributions are approximated by discrete random measures, which are composed of weighted particles, where the particles are samples of the unknown states from the state-space, and the particle weights are "probability masses" computed by using Bayes theory. In the implementation of particle filtering, importance sampling plays a crucial role and, since the procedure is designed for sequential use, the method is also called sequential importance sampling. The advantage of particle filtering over other methods is in that the exploited approximation does not involve linearizations around current estimates but rather approximations in the representation of the desired distributions by discrete random measures.

In this paper we discuss the use of particle filtering in several important problems in communications. Figure 1 presents a diagram that sorts out the addressed problems into two groups, one related to single-user systems, and the other to multiple access systems. For single user systems, the interest revolves around detection in flat fading and equalization, where the emphasis of the latter is on time-invariant, time-variant channels, and Orthogonal Frequency Division Multiplexing (OFDM) systems. For multiple access systems, the focus is on particle filtering for multiuser detection in Code Division Multiple Access (CDMA) systems and space-time decoding in fading channels. In all these cases, the first step is defining the problem with state-space representation. For example, a general baseband communications model for a fading channel can be written as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{u}_t) \\ y_t &= \mathbf{s}_t^\top \mathbf{h}_t + v_t \end{aligned} \tag{2}$$

where y_t is the discrete time signal received at the receiver, and $\mathbf{x}_t$ is the state of the system composed of vectors of transmitted symbols $\mathbf{s}_t$ and fading channel coefficients $\mathbf{h}_t$. The state varies in time according to a known function $\mathbf{f}_t$ according to a Markov process driven by the noise $\mathbf{u}_t$. Finally, v_t is additive channel noise. The primary objective is to *sequentially* detect the transmitted symbols and/or estimate the channel as the observations arrive. From a Bayesian point of view, this implies obtaining estimates of $p(\mathbf{h}_t, s_t|y_{0:t})$, where $y_{0:t} = \{y_0, y_1, \cdots, y_t\}$, which is exactly what particle filters are designed for. Many other problems in communications can be described similarly as by (2), some of which are presented in the sequel.

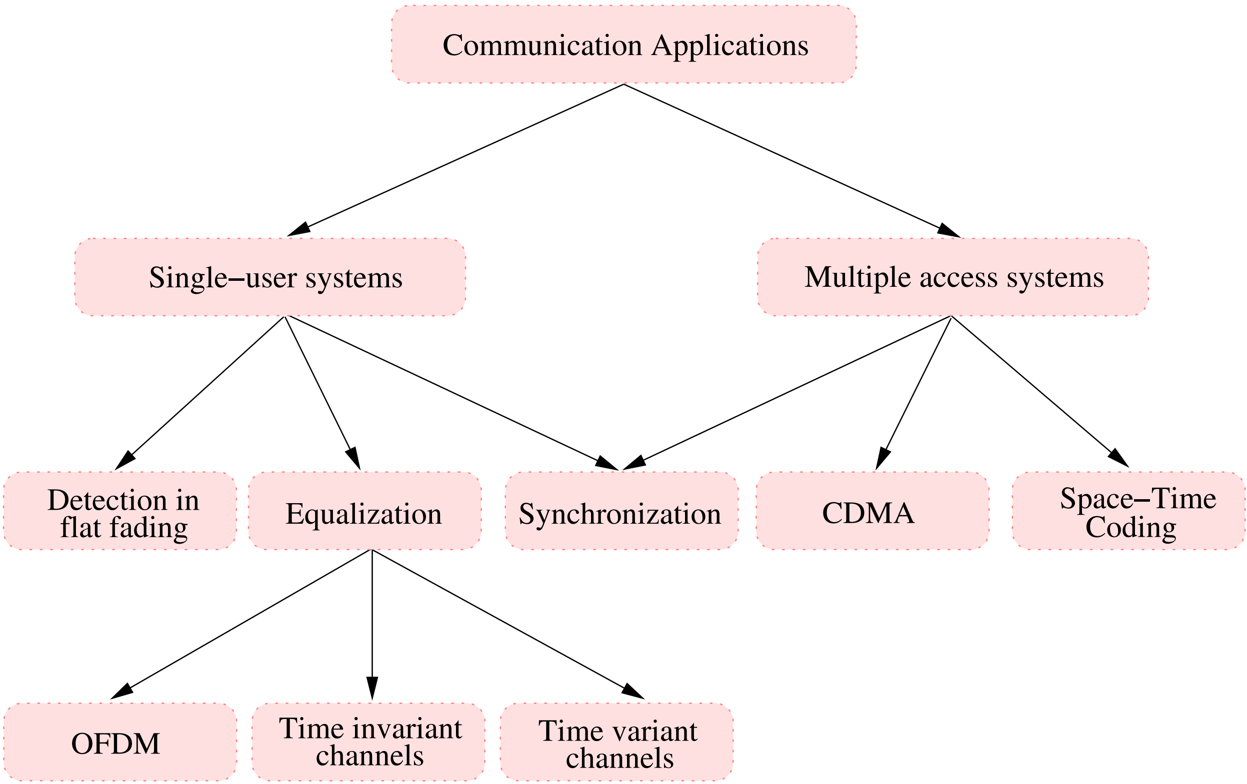

Fig. 1. Problems in communications addressed in the paper.

The paper is organized as follows. First, we present the fundamentals of particle filtering (Section II) and then we proceed with the applications to blind equalization (Section III), blind detection over flat-fading channels (Section IV), multiuser detection with particle filtering (Section V), and estimation and detection of space-time codes in fading channels (Section VI). Section VII contains a summary of additional work on applications of particle filtering to communications. Before we continue, we provide a brief summary of our notation:

a_k $\quad$ $k-$th coefficient of an autoregressive (AR) process

b_k	$k-$th coefficient of a moving-average (MA) process
C	chip-rate (processing) gain
$c(\cdot)$	code and modulation function
$\mathbf{f}_t(\cdot)$	system transition function at time t
$\mathbf{g}_t(\cdot)$	measurement function at time t
$\mathbf{h}_t$	communication channel at time t
m	running index for particles or trajectories
K	number of users in a CDMA system
L	order of communication channel
$\mathcal{L}$	symbol alphabet
M	total number of particles
$\mathcal{N}(\mu,\sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}_c(\mu,\sigma^2)$	complex Gaussian distribution with mean μ and variance σ^2
N_T	number of transmit antennas
N_R	number of receive antennas
$\pi(\cdot)$	importance sampling function
$p(\cdot)$	probability distribution function
r_a	order of an AR process
r_b	order of an MA process
$\mathbf{R}$	correlation matrix
s_t	symbol transmitted at time t
σ^2	noise variance

t	discrete time index, where $t \in \mathbb{N}$
T	size of a frame of symbols
T_s	symbol duration
τ	continuous time index
$\mathbf{u}_t$	system noise vector at time t
$\mathbf{v}_t$	observation noise vector at time t
$w_t^{(m)}$	weighting coefficient of particle m at time t
$w_t^{*(m)}$	non-normalized weighting coefficient of particle m at time t
$\mathbf{x}_t$	system state vector at time t
$\mathbf{x}_{0:t}$	$\{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_t\}$, a trajectory of states
$\mathbf{x}_t^{(m)}$	$m-$th particle at time t
$\mathbf{x}_{0:t}^{(m)}$	$m-$th trajectory of particles
ξ_k	signature of $k-$th user in CDMA transmission
$\mathbf{y}_t$	observation vector at time t
$\mathbf{y}_{0:t}$	$\{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_t\}$, a sequence of observations

In general, small and capital letters in bold face denote vectors and matrices, respectively.

II. Fundamentals of particle filtering

Consider a system/signal with a state-space representation given by (1). As already pointed out, the main task of sequential signal processing is the estimation of the state $\mathbf{x}_t$ recursively from the observations $\mathbf{y}_t$. In general, there are three probability distribution functions of interest, and they are the filtering distribution, $p(\mathbf{x}_t|\mathbf{y}_{0:t})$, the predictive distribution, $p(\mathbf{x}_{t+l}|\mathbf{y}_{0:t})$, $l \geq 1$, and the smoothing distribution, $p(\mathbf{x}_t|\mathbf{y}_{0:T})$, where $T > t$. All the information about $\mathbf{x}_t$ regarding filtering, prediction or smoothing is captured by these distributions, respectively, and so the main goal is their tracking, that is obtaining $p(\mathbf{x}_t|\mathbf{y}_{0:t})$ from $p(\mathbf{x}_{t-1}|\mathbf{y}_{0:t-1})$, or $p(\mathbf{x}_{t+l}|\mathbf{y}_{0:t})$ from $p(\mathbf{x}_{t+l-1}|\mathbf{y}_{0:t})$ or $p(\mathbf{x}_t|\mathbf{y}_{0:T})$ from $p(\mathbf{x}_{t+1}|\mathbf{y}_{0:T})$. The algorithms that exactly track these distributions are known as optimal algorithms. In many practical situations, however, the optimal algorithms are impossible to implement, primarily because the updates of the distributions require integrations that cannot be performed analytically or summations that are impossible to carry out due to the number of terms in the summations.

For the joint a posteriori distribution of $\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_t$, in case of independent noise samples which are assumed throughout the paper, we can write

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) \propto p(\mathbf{x}_0|\mathbf{y}_0) \prod_{k=1}^{t} p(\mathbf{y}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{x}_{k-1}). \tag{3}$$

It is straightforward to show that a recursive formula for obtaining $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ from $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})$ is given by

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}). \tag{4}$$

Since the transition from $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})$ to $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ is often analytically intractable, we resort to methods that are based on approximations.

In particle filtering, the distributions are approximated by discrete random measures defined by particles and weights assigned to the particles. If the distribution of interest is $p(x)$ and its approximating random measure is

$$\chi = \left\{x^{(m)}, w^{(m)}\right\}_{m=1}^{M} \tag{5}$$

where $x^{(m)}$ are the particles, $w^{(m)}$ are their weights, and M is the number of particles used in the approximation, χ approximates the distribution $p(x)$ by

$$p(x) \approx \sum_{m=1}^{M} w^{(m)} \delta(x - x^{(m)}) \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function. With this approximation, computations of expectations (which involve complicated integrations) are simplified to summations, that is, for example,

$$E\left(g(X)\right) = \int g(x)p(x)dx \tag{7}$$

is approximated by

$$E\left(g(X)\right) \approx \sum_{m=1}^{M} w^{(m)} g(x^{(m)}). \tag{8}$$

The next important concept used in particle filtering is the principle of importance sampling. Suppose we want to approximate a distribution $p(x)$ with a discrete random measure. If we can generate the particles from $p(x)$, each of them will be assigned a weight equal to $1/M$. When direct sampling from $p(x)$ is intractable, one can generate particles $x^{(m)}$ from a distribution $\pi(x)$, known also as importance function, and assign (non-normalized) weights according to

$$w^{*(m)} = \frac{p(x)}{\pi(x)} \tag{9}$$

which upon normalization become

$$w^{(m)} = \frac{w^{*(m)}}{\sum_{i=1}^{M} w^{*(i)}}. \tag{10}$$

Suppose now that the posterior distribution $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})$ is approximated by the discrete random measure $\chi_{t-1} = \{\mathbf{x}_{0:t-1}^{(m)}, w_{t-1}^{(m)}\}_{m=1}^{M}$. Note that the trajectories or streams of particles $\mathbf{x}_{0:t-1}^{(m)}$ can be considered particles of $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})$. Given the discrete random measure χ_{t-1} and the observation $\mathbf{y}_t$, the objective is to exploit χ_{t-1} in obtaining χ_t. Sequential importance sampling methods achieve this by generating particles $\mathbf{x}_t^{(m)}$ and appending them to $\mathbf{x}_{0:t-1}^{(m)}$ to form $\mathbf{x}_{0:t}^{(m)}$, and updating the weights $w_t^{(m)}$ so that χ_t allows for accurate estimates of the unknowns of interest at time t.

If we use an importance function that can be factored as

$$\pi(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) = \pi(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})\pi(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}) \tag{11}$$

and if

$$\mathbf{x}_{0:t-1}^{(m)} \sim \pi(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}) \tag{12}$$

and

$$w_{t-1}^{(m)} \propto \frac{p(\mathbf{x}_{0:t-1}^{(m)}|\mathbf{y}_{0:t-1})}{\pi(\mathbf{x}_{0:t-1}^{(m)}|\mathbf{y}_{0:t-1})} \tag{13}$$

we can augment the trajectory $\mathbf{x}_{0:t-1}^{(m)}$ with $\mathbf{x}_t^{(m)}$, where

$$\mathbf{x}_t^{(m)} \sim \pi(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(m)}, \mathbf{y}_{0:t}) \tag{14}$$

and easily associate with it an updated weight $w_t^{(m)}$ obtained according to

$$w_t^{(m)} \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(m)})\, p(\mathbf{x}_t^{(m)}|\mathbf{x}_{t-1}^{(m)})}{\pi(\mathbf{x}_t^{(m)}|\mathbf{x}_{0:t-1}^{(m)}, \mathbf{y}_{0:t})} w_{t-1}^{(m)}. \tag{15}$$

The sequential importance sampling algorithm can thus be implemented by performing the following two steps for every t:

1. Draw particles $\mathbf{x}_t^{(m)} \sim \pi(\mathbf{x}_t|\mathbf{x}_{t-1}^{(m)}, \mathbf{y}_{0:t})$, where $m = 1, 2, \cdots, M$.
2. Compute the weights of $w_t^{(m)}$ according to (15).

The importance function plays a very important role in the performance of the particle filter. This function must have the same support as the probability distribution that is being approximated. In general, the closer the importance function to that distribution, the better the approximation is. In the literature, the two most frequently used importance functions are the prior and the optimal importance function. The prior importance function is given by $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(m)})$, and it implies particle weight updates by

$$w_t^{(m)} \propto w_{t-1}^{(m)}\, p(\mathbf{y}_t \mid \mathbf{x}_t^{(m)}). \tag{16}$$

The optimal importance function minimizes the variance of the importance weights conditional on the trajectory $\mathbf{x}_{0:t-1}^{(m)}$ and the observations $\mathbf{y}_{0:t}$, and is given by $p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}^{(m)}, \mathbf{y}_{0:t})$ [11]. When the optimal function is used, the update of the weights is carried out according to

$$w_t^{(m)} \propto w_{t-1}^{(m)} p(\mathbf{y}_t \mid \mathbf{x}_{t-1}^{(m)}). \tag{17}$$

Note that implementations of particle filters with prior importance functions are much easier than those with optimal importance functions. The reason is that the computation of $p(\mathbf{y}_t \mid \mathbf{x}_{t-1}^{(m)})$ requires integration.

A major problem with particle filtering is that the discrete random measure degenerates quickly. In other words, all the particles except for a very few are assigned negligible weights. The degeneracy implies that the performance of the particle filter will deteriorate. Degeneracy, however, can be reduced by using good importance sampling functions and *resampling*.

Resampling is a scheme that eliminates particles with small weights and replicates particles with large weights. In principle, it is implemented as follows:

1. Draw M particles, $\mathbf{x}_t^{*(m)}$ from the discrete distribution χ_t.
2. Let $\mathbf{x}_t^{(m)} = \mathbf{x}_t^{*(m)}$, and assign equal weights $(1/M)$ to the particles.

The idea of resampling is depicted in Figure 2 with $M = 10$ particles. There, the left column of circles represents particles before resampling, where the diameters of the circles are proportional to the weights of the particles. The right column of circles are the particles after resampling. In general the large particles are replicated and the small particles are removed. For example, the

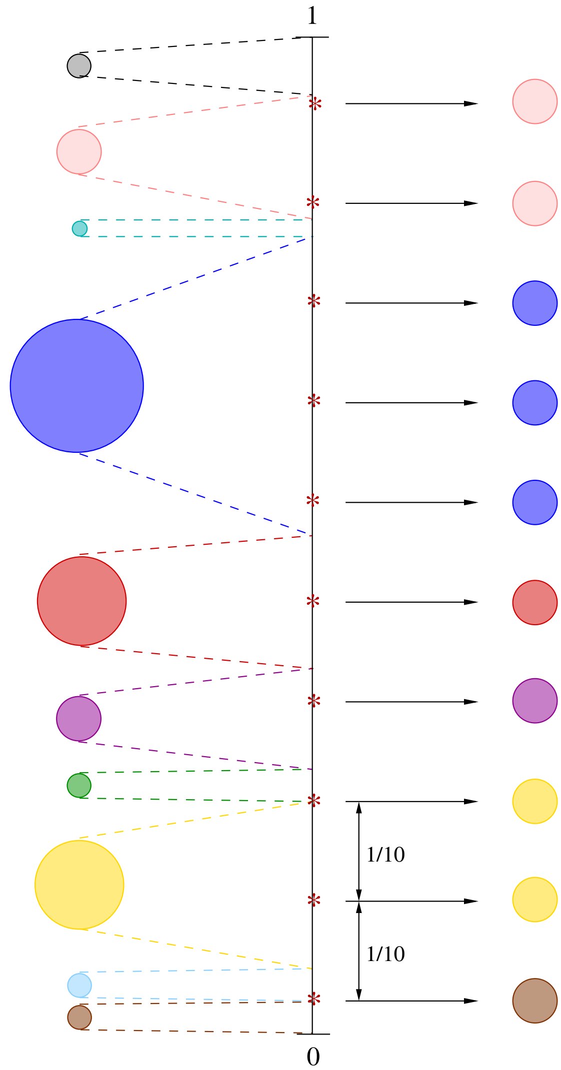

Fig. 2. A schematic description of resampling.

"blue" particle with the largest weight is replicated three times and the "yellow" particle, two times, whereas the green particles, which have small weights, are removed. Also, after resampling all the circles have equal diameters, that is, all the weights are set to $1/M$. In Figure 3, we represent pictorially the random measures and the actual probability distributions of interest as well as the three steps of particle filtering: particle generation, weight update, and resampling. In the figure, the solid curves represent the distributions of interest, which are approximated by the discrete measures. The sizes of the particles reflect the weights that are assigned to them. Finally in Figure 4, we display a flow-chart that summarizes the particle filtering algorithm. At time t, a new set of particles is generated, and their weights are computed. Thereby we obtain the random measure χ_t, which can be used for estimation of the desired unknowns. Before we proceed with the generation of the set of particles for time instant $t+1$, we estimate the *effective* particle size (a metric that measures the degeneracy of the particles [29], [36]). If the effective particle size is below a predefined threshold, resampling takes place; otherwise we proceed with the regular steps of new particle generation and weight computation.

Recently, a special class of particle filters that approximate the posterior distributions by single Gaussians has been introduced [30]. Although in their derivation it is assumed that all the relevant distributions are Gaussian, as is done with some other filters including the extended Kalman filter and its variants, they are distinguishable in that the updating of the filtering and predictive distributions is accomplished by propagating particles. This entails advantages of easier implementation than is the case with the standard particle filters and improved performance over other Gaussian based approximation filters. The Gaussian particle filter has also been used as a building block for more complex filters called Gaussian sum particle filters [31]. These filters approximate the filtering and predictive distributions by weighted Gaussian mixtures and basically represent banks of Gaussian particle filters.

Before we continue with the presentation of applications of particle filtering to communication problems, we summarize the procedure for developing particle filtering algorithms. The procedure involves the following steps:

1. Description of the problem by a discrete state-space model as in (1).

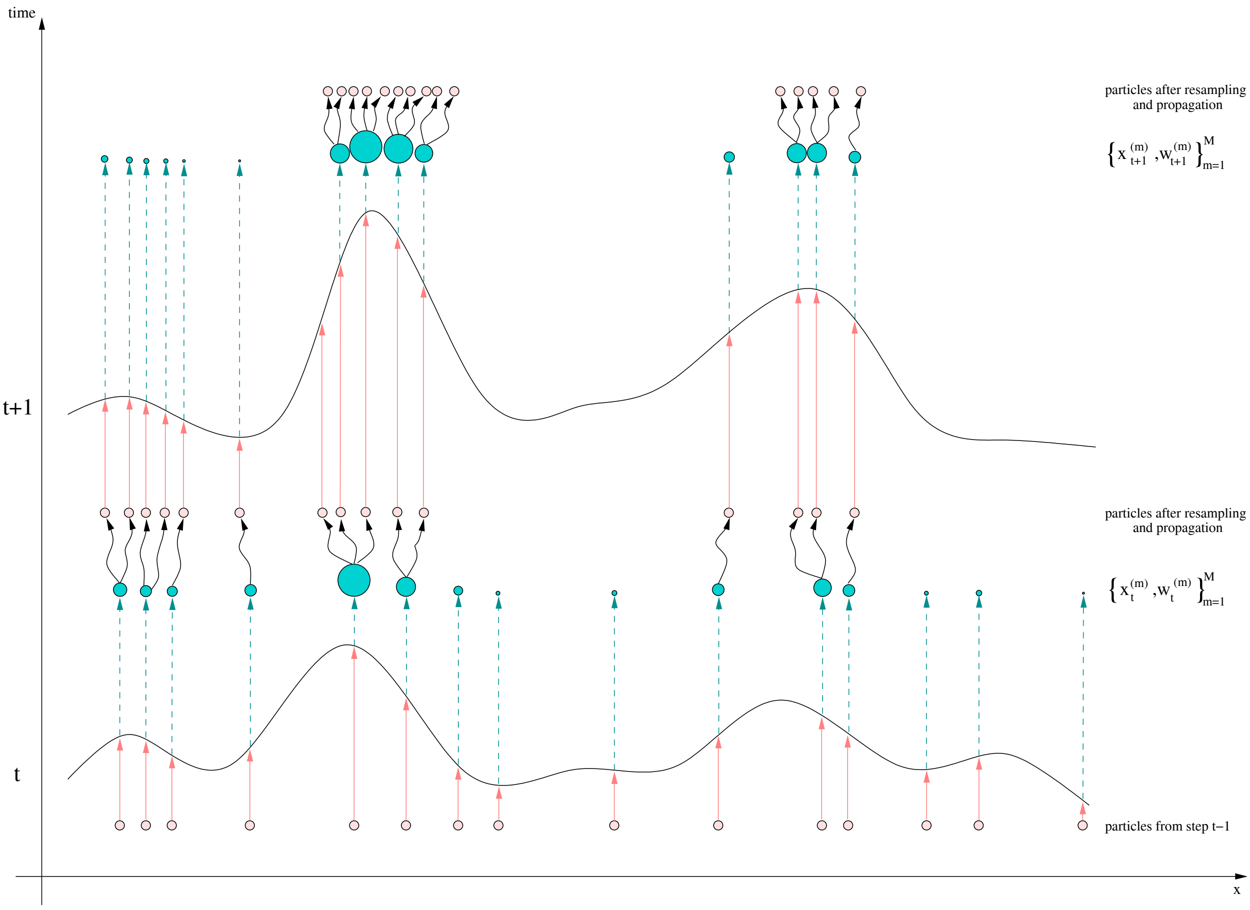

Fig. 3. A pictorial description of particle filtering.

2. Selection of a proposal function for particle generation.

3. Derivation of the equations for the weight update.

Additional issues are the choice of resampling algorithm and the schedule for resampling. We proceed with showing how these steps are applied to resolving the problem of blind equalization.

III. Blind equalization

When digital symbols are transmitted over frequency-selective channels, inter-symbol interference (ISI) occurs, which has detrimental effect on the detection at the receiver. To allow for symbol detection with reasonable error, channel equalization is needed to reverse the effect of ISI. A

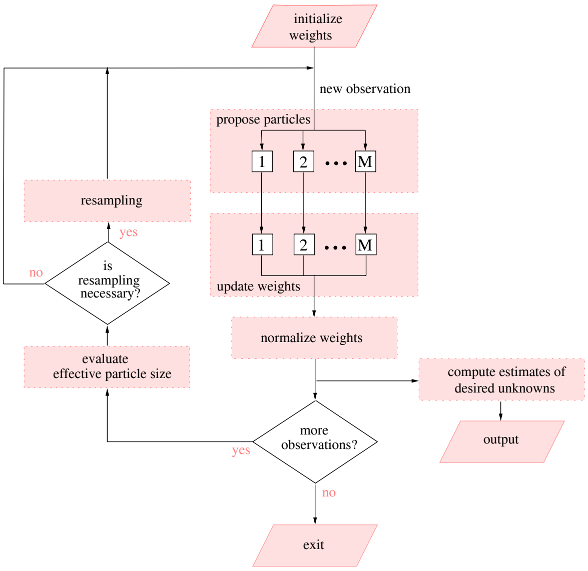

Fig. 4. A block diagram of particle filtering.

popular equalization technique applies the principle of maximum-likelihood estimation that results in the Viterbi algorithm for symbol detection. When the channel parameters are unknown, they are first estimated, usually from training data whose transmission contributes to significant overheads and bandwidth-inefficient communication.

Blind equalization involves detection of transmitted symbols without using training data. This can be accomplished either without explicit estimation of the channel parameters or by joint symbol detection and channel parameter estimation.

Recently, several researchers have employed particle filtering for problems of blind equalization. The flexibility of particle filtering has allowed for application of several variants of blind equalization including ones that involve time-invariant channels, [38], [41], time-varying channels [4], [5], [15], [16], additive Gaussian and non-Gaussian channels [44], as well as OFDM systems [55].

For convenience of discussing in greater detail some of the work done in equalization, we adopt the following signal model. When digital symbols s_t are transmitted over a frequency selective channel, the received signal can be represented as

$$y_t = \sum_{l=0}^{L-1} s_{t-l} h_{t,l} + v_t = \mathbf{h}_t^\top \mathbf{s}_t + v_t \tag{18}$$

where y_t is the received signal at time instant t, $\mathbf{s}_t^\top = [s_t \; s_{t-1} \cdots s_{t-L+1}]$, $\mathbf{h}_t^\top = [h_{t,0} \; h_{t,1} \; \cdots \; h_{t,L-1}]$ are the coefficients of the unknown FIR channel impulse response, L is the length of the channel, and v_t is an additive noise which is usually considered as a zero mean Gaussian process with a known variance σ_v^2. The objective is to detect the transmitted symbols by first obtaining the posterior distribution $p(s_{1:t}|y_{1:t})$ and then using it to perform detection.

A. Time-invariant channels

When the channels are time-invariant, we have $\mathbf{h}_t = \mathbf{h}$. If we assume Gaussian priors for the channel coefficients, we can analytically marginalize them and directly draw samples from the posterior distribution of the symbols only. This allows for performing equalization without explicitly estimating the channel coefficients. In other words, as per (15), for the update of the weighting coefficients we would use

$$w_t^{(m)} \propto \frac{p(y_t|s_{0:t}^{(m)}, y_{0:t-1})p(s_t^{(m)}|s_{0:t-1}^{(m)})}{\pi(s_t^{(m)}|s_{0:t-1}^{(m)}, y_{0:t})} w_{t-1}^{(m)}. \tag{19}$$

Following [41] and with the assumption that the data symbols $s_t \in \{-1, +1\}$ are i.i.d. uniform random variables, we can write the state space model of the observed data as

$$\begin{aligned} \mathbf{s}_t &= \mathbf{F}\mathbf{s}_{t-1} + \mathbf{u}_t \\ y_t &= \mathbf{h}^\top \mathbf{s}_t + v_t \end{aligned} \tag{20}$$

where $\mathbf{F}$ is an $L \times L$ state transition matrix given by

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{21}$$

$v_t \sim \mathcal{N}(0, \sigma_v^2)$, and $\mathbf{u}_t = [s_t \; 0 \dots 0]$. If we assume that the channel coefficients have a Gaussian prior distribution, $\mathbf{h} \sim \mathcal{N}(\bar{\mathbf{h}}_{-1}, \mathbf{C}_{-1})$, it can be shown that the posterior distribution of the channel, $p\,(\mathbf{h}^{(m)}|s_{0:t}^{(m)}, y_{0:t})$, which corresponds to the $m-$th trajectory of symbols, $s_{0:t}^{(m)}$, is Gaussian, that is, $p\,(\mathbf{h}^{(m)}|s_{0:t}^{(m)}, y_{0:t}) = \mathcal{N}(\bar{\mathbf{h}}_t^{(m)}, \mathbf{C}_t^{(m)})$. The mean and the covariance matrix are recursively updated by

$$\begin{aligned} \bar{\mathbf{h}}_t^{(m)} &= \mathbf{C}_t^{(m)} \left(\frac{\mathbf{s}_t^{(m)} y_t}{\sigma_v^2} + \mathbf{C}_{t-1}^{-1(m)} \bar{\mathbf{h}}_{t-1}^{(m)} \right) \\ \mathbf{C}_t^{-1(m)} &= \frac{\mathbf{s}_t^{(m)} \mathbf{s}_t^{\top(m)}}{\sigma_v^2} + \mathbf{C}_{t-1}^{-1(m)} \end{aligned} . \tag{22}$$

In [38], [41], it is shown that we can obtain an analytical expression for the likelihood function, $p(y_t|s_{0:t}^{(m)}, y_{0:t-1})$, by marginalizing out the channel coefficient, i.e.,

$$\begin{aligned} p(y_t|s_{0:t}^{(m)}, y_{0:t-1}) &= \frac{\left|\mathbf{C}_t^{(m)}\right|^{1/2}}{\left(2\pi\sigma_v^2 \left|\mathbf{C}_{t-1}^{(m)}\right|\right)^{1/2}} \exp\left\{ -\frac{1}{2} \left[\frac{{y_t}^2}{\sigma_v^2} + \bar{\mathbf{h}}_{t-1}^{(m)\top} \mathbf{C}_{t-1}^{-1(m)} \bar{\mathbf{h}}_{t-1}^{(m)} \right.\right. \\ &\left.\left. - \left(\frac{\mathbf{s}_t y_t}{\sigma_v^2} + \mathbf{C}_{t-1}^{-1(m)} \bar{\mathbf{h}}_{t-1}^{(m)} \right)^{\top} \mathbf{C}_t^{(m)} \left(\frac{\mathbf{s}_t y_t}{\sigma_v^2} + \mathbf{C}_{t-1}^{-1(m)} \bar{\mathbf{h}}_{t-1}^{(m)} \right) \right] \right\}. \end{aligned} \tag{23}$$

The optimal importance function is proportional to the likelihood function, or,

$$\pi(s_t|s_{0:t-1}^{(m)}, y_{0:t}) \propto p(y_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1}).$$

It is readily shown that the function has the form

$$\pi(s_t|s_{0:t-1}^{(m)}, y_{0:t}) = \frac{p(y_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1})}{p(y_t|s_t = 1, s_{0:t-1}^{(m)}, y_{0:t-1}) + p(y_t|s_t = -1, s_{0:t-1}^{(m)}, y_{0:t-1})} \tag{24}$$

and that the updating of the particle weights is carried out by

$$w_t^{(m)} \propto w_{t-1}^{(m)} \left(p(y_t|s_t = 1, s_{0:t-1}^{(m)}, y_{0:t-1}) + p(y_t|s_t = -1, s_{0:t-1}^{(m)}, y_{0:t-1}) \right). \tag{25}$$

When the particles and weights of all the trajectories are obtained, symbol detection can be carried out using smoothing. Smoothing in this context has been addressed in [8], [9], where the emphasis is on the methods for fixed-lag blind equalization and the interest lies in obtaining the smoothing distribution $p(s_{t-l}|y_{1:t})$, where l is the fixed lag, followed up by detection of s_{t-l}.

An interesting extension of the equalization algorithm is its modification to cope with unknown channel orders. The problem can be resolved by employing a marginalization strategy where the unknown channel order is marginalized. Another possibility is to estimate the channel order with the transmitted symbols.

B. Time-variant channels

For time-varying channels, such as mobile communication channels, symbol detection and channel estimation can be performed jointly following the concept of mixture Kalman filtering (MKF) [6], [11]. The problem is again formulated by writing the channel model as a state equation. As before, the channels are represented as FIR filters, but now they have time varying complex coefficients whose magnitudes are randomly varying Rayleigh processes. The coefficient variations can be approximately modeled as autoregressive-moving average (ARMA) processes given by

$$h_{t,l} = -\sum_{k=1}^{r_a} a_k h_{t-k,l} + \sum_{k=0}^{r_b} b_k u_{t-k,l} \tag{26}$$

where $h_{t,l}$ is the $l-$th coefficient at time t, $b_0 = 1$, and $u_{t,l}$ is the driving noise process of the $l-$th coefficient. The parameters $a_1, a_2, \cdots, a_{r_a}$, and $b_1, b_2, \cdots, b_{r_b}$ are functions of the fading rate of the channels and can be evaluated if the Doppler spread of the channels and the symbol rate are known. Suppose for clarity of presentation that the variation of the channel is modeled as an autoregressive process (AR) (all the b_k's except b_0 in (26) are zero). Then, if we define the channel state as $\mathbf{x}_t^\top = [\mathbf{h}_t^\top \, \mathbf{h}_{t-1}^\top \, \cdots \, \mathbf{h}_{t-r_a+1}^\top]$, where $\mathbf{h}_t^\top = [h_{t,0}, h_{t,1}, \ldots, h_{t,L-1}]$ (that is, $\mathbf{x}_t$ is an $r_a L \times 1$ vector), we obtain the following state space model:

$$
\begin{aligned}
\mathbf{s}_t &= \tilde{\mathbf{F}}\mathbf{s}_{t-1} + \mathbf{z}_t \\
\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{D}\mathbf{u}_t \\
y_t &= \mathbf{s}_t^\top \mathbf{x}_t + v_t.
\end{aligned} \tag{27}
$$

Here $\mathbf{s}_t$ is an $r_a L \times 1$ vector defined by $\mathbf{s}_t^\top = [s_t \; s_{t-1} \; \cdots \; s_{t-L+1} \; 0 \; \cdots \; 0]$, $\tilde{\mathbf{F}}$ is an $r_a L \times r_a L$ matrix given by

$$
\tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}
$$

where $\mathbf{F}$ is defined by (21), $\mathbf{z}_t$ is an $r_a L \times 1$ vector whose elements are all equal to zero except for the first one which is equal to s_t, and $\mathbf{u}_t$ is an $L \times 1$ zero mean Gaussian noise vector whose covariance matrix has diagonal elements proportional to the power of each lag. The matrix $\mathbf{A}$ has dimensions $r_a L \times r_a L$ and is constructed from the coefficients of the AR process according to

$$
\mathbf{A} = \begin{bmatrix} -a_1\mathbf{I} & -a_2\mathbf{I} & \cdots & \cdots & -a_{r_a}\mathbf{I} \\ \mathbf{I} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{28}
$$

and $\mathbf{D}$ is an $r_a L \times L$ matrix given by

$$\mathbf{D} = \begin{pmatrix} \mathbf{I} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

where $\mathbf{I}$ and $\mathbf{0}$ in the above expressions are the identity and zero matrices of sizes $L \times L$.

Now the objective is to jointly estimate the state of the channel $\mathbf{x}_t$ and to detect the transmitted symbols s_t. It should be noted that given the transmitted symbols, the state space model becomes a linear Gaussian model and the posterior distribution of the channel $p(\mathbf{x}_t|s_{0:t}, y_{0:t})$ is Gaussian for all t. This permits recursive estimation of the channel parameters using Kalman filtering [16].

We consider the importance function

$$\begin{aligned} \pi(s_t|s_{0:t-1}^{(m)}, y_{0:t}) &\propto p(y_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1})p(s_t|s_{0:t-1}^{(m)}, y_{0:t-1}) \\ &\propto p(y_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1}) \end{aligned} \tag{29}$$

where as before we assume that the symbols are i.i.d. uniform random variables. The last term can be expressed as

$$p\left(y_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1}\right) = \int p\left(y_t|\mathbf{x}_t, s_t, s_{0:t-1}^{(m)}, y_{0:t-1}\right)p\left(\mathbf{x}_t|s_t, s_{0:t-1}^{(m)}, y_{0:t-1}\right)\, d\mathbf{x}_t \tag{30}$$

where the second factor of the integrand is the predictive distribution of the Kalman filter. Then, the expression for the importance function can be rewritten as

$$\begin{aligned} \pi(s_t|s_{0:t-1}^{(m)}, y_{0:t}) &\propto \int \mathcal{N}_c(y_t; \mathbf{s}_t^{\top(m)}\mathbf{x}_t, \sigma^2)\mathcal{N}_c(\mathbf{x}_t; \boldsymbol{\mu}_t^{(m)}, \boldsymbol{\Sigma}_t^{(m)})\, d\mathbf{x}_t \\ &= \mathcal{N}_c\left(\tilde{\mathbf{s}}_t^{\top(m)}\boldsymbol{\mu}_t^{(m)}, \sigma_v^2 + \tilde{\mathbf{s}}_t^{(m)\top}\boldsymbol{\Sigma}_t^{(m)}\tilde{\mathbf{s}}_t^{(m)}\right) \end{aligned} \tag{31}$$

where for clarity, $\mathcal{N}_c(y_t; \mathbf{s}_t^{\top(m)}\mathbf{x}_t, \sigma^2)$ and $\mathcal{N}_c(\mathbf{x}_t; \boldsymbol{\mu}_t^{(m)}, \boldsymbol{\Sigma}_t^{(m)})$ denote that y_t and $\mathbf{x}_t$ have complex Gaussian distributions, respectively, $\tilde{\mathbf{s}}_t^{\top(m)} = [s_t \; s_{t-1}^{(m)} \; s_{t-2}^{(m)} \cdots s_{t-L+1}^{(m)} 0 \cdots 0]$, and $\boldsymbol{\mu}_t^{(m)}$ and $\boldsymbol{\Sigma}_t^{(m)}$ are the predictive mean and covariance of $\mathbf{x}_t$ of the $m-$th trajectory, respectively. Similarly, we can show that the corresponding weights can be computed from

$$\begin{aligned} w_t^{(m)} &\propto w_{t-1}^{(m)} \sum_{s_t \in \{\pm 1\}} p(y_t | s_{0:t-1}^{(m)}, s_t, y_{0:i-1}) \\ &= w_{t-1}^{(m)} \sum_{s_t \in \{\pm 1\}} \mathcal{N}_c\left(y_t; \mathbf{s}_t^{\top(m)} \boldsymbol{\mu}_t^{(m)}, \sigma_v^2 + \mathbf{s}_t^{\top(m)} \boldsymbol{\Sigma}_t^{(m)} \mathbf{s}_t^{(m)}\right). \end{aligned} \tag{32}$$

When implementing the algorithm, for each particle we evaluate the predictive mean and covariance of two Kalman filters, each corresponding to the symbols $s_t = 1$ and $s_t = -1$, respectively. Then we draw samples using (31), and calculate the weights by (32). Finally, the Kalman filters corresponding to the sampled symbols are updated. Once the particles with their corresponding weights are obtained, the symbol is estimated using the maximum a posteriori (MAP) criterion. If needed, a minimum mean square error (MMSE) (or other type) estimate of the channel state can be obtained from the mean updates of the posterior distribution of $\mathbf{x}_t$.

Unlike other standard methods, particle filtering can easily be extended to non-Gaussian noises. In [44], the additive complex noise is modeled as a mixture of J zero mean Gaussians having different variances. There, a latent variable λ_t is defined to indicate the distribution of v_t. The procedure draws particles from an importance function $p(s_t, \lambda_t | s_{1:t-1}^{(m)}, \lambda_{1:t-1}^{(m)}, y_{0:t})$, and the particles are used for approximation of the joint posterior distribution $p(s_t, \lambda_t | y_{0:t})$ from which the MAP estimates of the symbols are obtained. It is reported that the algorithm outperforms existing methods based on Gaussian noise [44].

Recently, a similar treatment has been extended to OFDM systems over frequency selective channels [55]. One important difference with the above treatment is that the received signal y_t

is considered to be an observation in the frequency domain where the index t there represents different subcarriers. In such systems the observed signals of all the subcarriers are simultaneously received. The channel parameters are assumed to have Gaussian a priori distribution which allows, as discussed earlier, to perform symbol detection without explicitly determining the channel parameters.

Blind equalization for satellite communications was considered in [33]. There, a state equation was developed for a nonlinear satellite channel consisting of a cascade of linear filters and a memoryless nonlinear traveling wave tube amplifier. The state equation coupled with the observation, which consists of the state variable embedded in additive Gaussian noise provided a dynamic state space model for the system. With the assumption that all the system parameters and noise variances are known, a generic particle filtering detector employing the prior importance function was applied to combat the nonlinear distortion of the channel.

C. Simulations

In our experiment we have simulated a scenario of a time-invariant channel with an impulse response of length $L = 3$. We assumed a Gaussian prior for the channel coefficients, $\mathbf{h} \sim \mathcal{N}(\overline{\mathbf{h}}_{-1}, \mathbf{C}_{-1})$, where

$$\overline{\mathbf{h}}_{-1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{C}_{-1} = \begin{bmatrix} 0.10000 & 0 & 0 \\ 0 & 0.24569 & 0 \\ 0 & 0 & 0.05475 \end{bmatrix}.$$

This choice of prior mean and covariance matrix corresponds to an environment with a strong line of sight component and two weaker, zero-mean paths. This is a fairly realistic scenario for an indoor communication system, for instance. The numerical values of the means and variances of the channel taps are selected to yield the delay power profile in Figure 5.

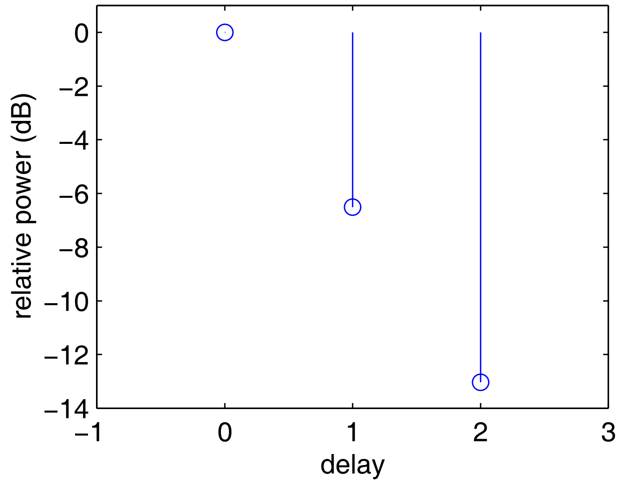

Fig. 5. Delay power profile of the 3-tap channel.

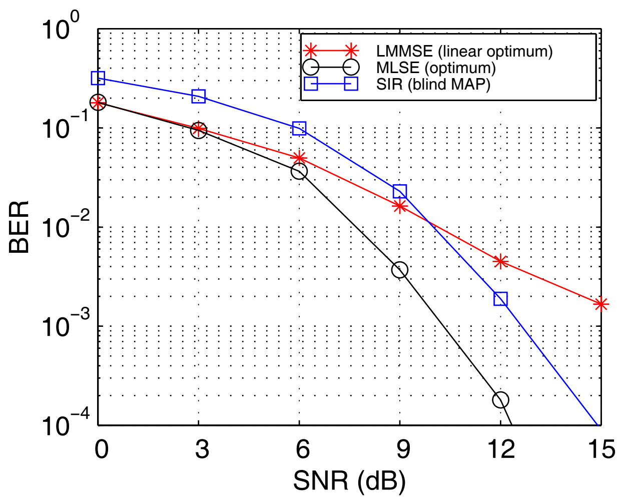

Fig. 6. BERs of different equalizers.

In order to estimate the bit error rate (BER) under this channel model, we have randomly generated 5000 signal bursts of duration $T = 40$ (i.e., 200,000 bits). A new sample channel is drawn from the above prior distribution for each burst. At the receiving end, we have simulated three equalizers:

- A one-shot linear MMSE (LMMSE) equalizer: This equalizer has perfect knowledge of the channel response, and it represents a Wiener matrix-filter that processes all the observations in a single burst at once [41]. Clearly, this is not a realistic receiver, but it yields a lower bound on the BER of simpler LMMSE equalizers.

- The Maximum Likelihood Equalizer (MLE): It, too, has perfect knowledge of the channel impulse response and is implemented via the Viterbi algorithm. This is the optimal sequence detector and, therefore, it yields a lower bound on the BER.

- The blind MAP equalizer: It is implemented via the Sequential Importance Resampling (SIR) algorithm [17]. The number of particles is $M = 300$ and resampling is carried out each time the effective particle size goes below the threshold $\varepsilon = 0.25M$.[1] The importance distribution is the optimal one.

The estimated BER curves are shown in Figure 6. Each point in the plots results from a trimmed average over the set of 2,500 signal bursts. Extreme simulation samples (the highest 0.5% and smallest 0.5%) have been discarded.

IV. Blind detection over flat fading channels

In this section, we discuss in detail the applications of particle filters to detection over flat fading channels. Figure 7 shows the baseband communications system block diagram over a frequency flat fading channel. The input signal to the system is a sequence of symbols s_t, transmitted after bandlimiting the pulses using a pulse shaping filter $g(\tau)$. The symbol period is T_s, and the channel is represented by the complex time varying process $h(\tau)$ and the additive noise $v(\tau)$.

Fading, which is the variation of the received complex amplitude, is a result of the multipath nature of the channel. Signals arriving at the receiver via multiple paths have different complex

[1] The effective sample size is defined by $M_{eff_t} = 1/\sum_{m=1}^{M}\left(w_t^{(m)}\right)^2$.

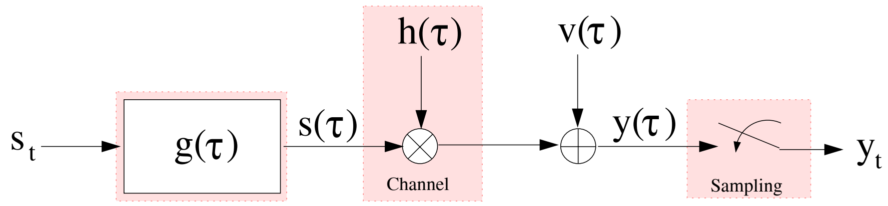

Fig. 7. Continuous-time model of the communication system.

gains and add up, resulting in a fading channel. The time variation in the number of paths, amplitudes and mainly the phases of the multiple paths produces a fading process with random nature. When the multiple paths arrive roughly within the same symbol period, the received signal does not undergo distortion in frequency and hence is called frequency flat or non-dispersive fading. The fading process is, however, highly correlated and is characterized by its bandwidth or Doppler frequency denoted as f_d. The random nature of the fading is described by the distribution of the process at each time instant. When the real and complex components are Gaussian, the resulting amplitude is Rayleigh distributed, while a line-of-sight component results in Ricean fading. For a more elaborate discussion on fading channels, see for example [43], [50].

It is important to note that the multiple paths arrive in the same symbol period, and therefore there is no inter-symbol interference. The receiver observes both a random complex gain for the transmitted symbol and additive channel noise. While the Jakes model [25] is often used to model the flat fading channel, a Markov model is preferred herein to obtain a state-space model. The flat fading process can be generated by filtering complex white noise by a low pass filter, whose spectral characteristics match that of the fading process [27]. As mentioned in the previous section, an AR (or ARMA) model adequately represents the fading process [50], [56], [60]. The AR (ARMA) parameters are chosen to match the spectral characteristics with those of the fading process. A simpler method, which uses a two-path model to build AR(2) and AR(3) processes can be found in [12] and [56] respectively; the results there closely approximate more complex path models.

The state-space representation of the baseband communications system can thus be given by

$$\begin{aligned} \mathbf{h}_t = & \quad \mathbf{A}\mathbf{h}_{t-1} + \mathbf{u}_t \\ y_t = & \quad \mathbf{s}_t^\top \mathbf{h}_t + v_t \end{aligned} \tag{33}$$

where $\mathbf{s}_t^\top = [s_t \;\; 0 \;\; 0 \;\; \ldots 0]$, and the channel $\mathbf{h}_t^\top = [h_t \; h_{t-1} \; \ldots \; h_{t-r_a}]$ is modeled as an autoregression with

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \ldots & -a_{r_a-1} & -a_{r_a} & 0 \\ 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \ldots & 0 & 1 & 0 \end{bmatrix}$$

where the AR parameters may be known or unknown, and $\mathbf{u}_t^\top = [u_t \; 0 \; \ldots \; 0]$, where u_t is complex white and Gaussian with variance σ_u^2. This is a reasonable assumption, since the AR parameters depend on the second order statistics of the channel and hence do not change as rapidly as the channel gain.

Our goal is to estimate the Bayesian posterior distribution $p\left(\mathbf{h}_t, s_t | y_{1:t}\right)$. For achieving this, we discuss two groups of algorithms of which the first assumes that the AR coefficients are known, and the second, that they are unknown.

A. Known AR coefficients

A.1 Detector I

Detector I employs the prior distribution as the importance sampling function, i.e.,

$$\begin{aligned} \pi_1(\mathbf{h_t}, s_t) & \equiv p(\mathbf{h}_t, s_t | \mathbf{h}_{t-1}^{(m)}, s_{t-1}^{(m)}) \\ & = p(\mathbf{h}_t | \mathbf{h}_{t-1}^{(m)}) p(s_t) \end{aligned} \tag{34}$$

where the last equality is due to the Markovian nature of the channel and assuming that the transmitted bits are independent of the channel and each other and are identically distributed. The weight update equation can then be written as

$$w_t^{(m)} \propto w_{t-1}^{(m)} p(y_t|\mathbf{h}_t^{(m)}, s_t^{(m)}) \tag{35}$$

$$= w_{t-1}^{(m)} p(y_t|h_t^{(m)}, s_t^{(m)}). \tag{36}$$

This detector can be used for both *linear* and *nonlinear* channel models. It can also be applied to channels with more *general fading* characteristics than the ones of Rayleigh fading, for as long as the fading process can be represented by a Markov model. This detector can also be used for *non-Gaussian* noise channels.

A.2 Detector II

Here the importance sampling function is the optimal function given by [11]

$$\begin{aligned} \pi_2(\mathbf{h}_t, s_t) &= p(\mathbf{h}_t, s_t|\mathbf{h}_{t-1}^{(m)}, s_{t-1}^{(m)}, y_t) \\ &= p(\mathbf{h}_t|\mathbf{h}_{t-1}^{(m)}, s_t, y_t)\; p(s_t|\mathbf{h}_{t-1}^{(m)}, y_t). \end{aligned} \tag{37}$$

Thus, a sample $s_t^{(m)}$ is first obtained from

$$p(s_t|\mathbf{h}_{t-1}^{(m)}, y_t) \propto p(y_t|s_t, \mathbf{h}_{t-1}^{(m)})\; p(s_t)$$

$$= \mathcal{N}(s_t h_{t-1}^{(m)},\;\; \sigma_v^2 + \sigma_u^2)\; p(s_t) \tag{38}$$

followed by a sample from $p(\mathbf{h}_t|\mathbf{h}_{t-1}^{(m)}, s_t^{(m)}, y_t) = \mathcal{N}(\boldsymbol{\mu}_t^{(m)},\;\; \boldsymbol{\Sigma}_t^{(m)})$, where

$$\begin{aligned} \boldsymbol{\Sigma}_t^{(m)} &= \left(\frac{1}{\sigma_u^2}\mathbf{I} + \frac{\mathbf{s}_t^{(m)}\mathbf{s}_t^{(m)\intercal}}{\sigma_v^2} \right)^{-1} \\ \boldsymbol{\mu}_t^{(m)} &= \boldsymbol{\Sigma}_t^{(m)} \left(\frac{\mathbf{A}\mathbf{h}_{t-1}^{(m)}}{\sigma_u^2} + \frac{\mathbf{s}_t^{(m)} y_t}{\sigma_v^2} \right). \end{aligned} \tag{39}$$

The weight update equation is given by

$$w_t^{(m)} \propto w_{t-1}^{(m)} p(y_t|\mathbf{h}_{t-1}^{(m)}, s_{t-1}^{(m)}) \tag{40}$$

where

$$p(y_t|\mathbf{h}_{t-1}^{(m)}, s_{t-1}^{(m)}) = \sum_{s_t \in \mathcal{L}} p(s_t)\mathcal{N}(y_t; s_t h_{t-1}^{(m)}, \ \ \sigma_v^2 + \sigma_u^2). \tag{41}$$

In the last expression, $\mathcal{N}(y_t; s_t^{(m)} h_{t-1}^{(m)}, \ \ \sigma_v^2 + \sigma_u^2)$ are Gaussian probability distribution functions with mean $s_t^{(m)} h_{t-1}^{(m)}$ and variance $\sigma_v^2 + \sigma_u^2$ and computed at y_t. This detector is optimal in the sense described in Subsection III-A. The generalization to non-Rayleigh fading channels and non-Gaussian channel noises is straightforward.

A.3 Detector III

This detector combines a bank of weighted Kalman filters and a particle filtering algorithm. It obtains the posterior distribution of the transmitted symbols by marginalizing the channel [7], [44]. Given the transmitted symbols, the state-space model in (33) is linear, and hence a Kalman filter can be used to track the channel. For the posterior of the symbol s_t, we can write

$$p(s_t|s_{0:t-1}^{(m)}, y_{0:t}) \propto p(y_t|s_{0:t-1}^{(m)}, s_t, y_{1:t-1})p(s_t) \tag{42}$$

where

$$p(y_t|s_{0:t-1}^{(m)}, s_t, y_{1:t-1}) = \int p(y_t|s_t, \mathbf{h}_t)p(\mathbf{h}_t|s_{0:t-1}^{(m)}, y_{0:t-1})d\mathbf{h}_t \tag{43}$$

and $p(\mathbf{h}_t|s_{0:t-1}^{(m)}, y_{0:t-1})$ is obtained from the Kalman filter. The weight update is given by

$$\begin{aligned} w_t^{(m)} &\propto w_{t-1}^{(m)} p(y_t|s_{0:t-1}^{(m)}, y_{1:t-1}) \\ &= w_{t-1}^{(m)} \sum_{s_t \in \mathcal{L}} p(y_t|s_t, s_{0:t-1}^{(m)}, y_{1:t-1})p(s_t). \end{aligned} \tag{44}$$

In [7], extensions to non-Gaussian channel noise and coded symbols are provided. Non-Rayleigh fading channels can also be tracked using Gaussian sum particle filters [32].

B. Unknown AR coefficients – Blind detection

In the previous subsection we have assumed that the AR (or ARMA) coefficients of the channel model are known. In practice, however, the channel statistics are unknown, which implies that the channel model coefficients must be estimated. One approach to their estimation is by using pilot signals. For accurate estimation, long sequences of pilot signals may be required, especially for slow fading channels. Further, for non-stationary channels, pilot signals must be constantly re-transmitted. An attractive alternative is to build a receiver for joint channel coefficients estimation and symbol detection.

The presence of unknown model coefficients does not allow for a direct extension of the above proposed particle filters. In particular, there are three related difficulties. First, the use of the prior importance function results in inefficient implementations, whereas the employment of the posterior importance function is prohibited. Second, ambiguities arise between the symbols and the model coefficients (for PSK modulated signals, there are also phase ambiguities). Third, if the model coefficients are static parameters, they may create problems because the diversity of their representations impoverishes after resampling. To provide particle filters with more diversity, rejuvenation procedures are required.

B.1 Detector I (RLS–based)

A blind algorithm for joint channel estimation and detection was first presented in [34]. The algorithm is a hybrid method which updates the unknown AR coefficients using the recursive least squares (RLS) method, while the channel and the symbols undergo particle filtering updates. The prior is used as importance sampling function, and the AR coefficients are marginalized as follows:

$$\begin{aligned} \pi(\mathbf{h}_t, s_t|\mathbf{h}_{0:t-1}^{(m)}, s_{0:t-1}^{(m)}, y_{0:t-1}) &\equiv p(\mathbf{h}_t|\mathbf{h}_{0:t-1}^{(m)}, y_{0:t-1})p(s_t) \\ &= p(s_t)\int p(\mathbf{h}_t|\mathbf{h}_{t-1}^{(m)}, \mathbf{a})p(\mathbf{a}|\mathbf{h}_{0:t-1}^{(m)})d\mathbf{a} \end{aligned} \tag{45}$$

where $\mathbf{a}^\top = [a_1, \ldots, a_{r_a}]$. This results in a hybrid algorithm, which is a weighted bank of RLS filters that update $\mathbf{a}$, and whose weights are computed recursively

$$w_t^{(m)} \propto w_{t-1}^{(m)} p(y_t|h_t^{(m)}, s_t^{(m)}). \tag{46}$$

B.2 Detector II

Another blind particle filtering detector was reported in [21]. The detector adopts an AR(2) model for the fading channel

$$h_t = -a_1 h_{t-1} - a_2 h_{t-2} + u_t. \tag{47}$$

The parameters a_1 and a_2 are obtained from

$$a_1 = -2r_d \cos(2\pi\Omega_d/\sqrt{2}) \quad \text{and} \quad a_2 = r_d^2 \tag{48}$$

where r_d is the pole radius of the AR(2) model, and $\Omega_d = f_d T$ is the normalized maximum Doppler frequency. The expressions described by (48) impose an interesting and important relationship between the underlying fading channel and the model coefficients. By considering the physical communication system, the relationship enables us to combat the ambiguity between the symbols and the coefficients and to implement a fully blind detector. Furthermore, to achieve efficient implementation, the detector first marginalizes out h_t using MKF, and then employs a hybrid importance function for s_t and the model coefficients a_1 and a_2 [22]. The hybrid importance function is expressed as

$$\begin{aligned}
&\pi(s_t, \mathbf{a}_t | s_{0:t-1}^{(m)}, \mathbf{a}_{0:t-1}^{(m)}, y_{0:t}) \\
&= p(s_t | \mathbf{a}_t, s_{0:t-1}^{(m)}, \mathbf{a}_{0:t-1}^{(m)}, y_{0:t}) p(\mathbf{a}_t | \mathbf{a}_{t-1}^{(m)}) \\
&= p(s_t | \mathbf{a}_t^{(m)}, \mathbf{a}_{0:t-1}^{(m)}, \mathbf{s}_{0:t-1}^{(m)}, y_{0:t}) \delta(a_{1,t} - a_{1,t-1}^{(m)}) \delta(a_{2,t} - a_{2,t-1}^{(m)}) \qquad (49)
\end{aligned}$$

where $\mathbf{a}_t = [a_{1t}, a_{2t}]^\top$, and the last equality is obtained based on the state equations $a_{1,t} = a_{1,t-1}$ and $a_{2,t} = a_{2,t-1}$. The corresponding weights are computed by

$$\begin{aligned}
w_t^{(m)} &\propto w_{t-1}^{(m)} p(y_t | \mathbf{a}_{0:t}^{(m)}, \mathbf{s}_{0:t-1}^{(m)}, y_{0:t-1}) \\
&= w_{t-1}^{(m)} \sum_{s_t \in \mathcal{L}} p(y_t | s_t, \mathbf{a}_{0:t}^{(m)}, \mathbf{s}_{0:t-1}^{(m)}, y_{0:t-1}). \qquad (50)
\end{aligned}$$

Note that, as suggested by its name, the hybrid importance function (49) is a combination of the posterior and the prior importance functions. As opposed to the posterior importance function, it is easily implementable because the sampling from (49) and the computation of the weight in (50) can be readily carried out. Finally, to overcome the impoverishment of the particle representation in a generic implementation, an auxiliary particle filter with a smoothing kernel can be applied.

B.3 Detector III

In addition to linear models, alternative modeling of fading channels may be preferred, especially if one wants to capture the nonlinearities of channels. A wavelet-based nonparametric modeling of fading channels was used in [18] where the fading process is decomposed using wavelet expansions, i.e.,

$$h_t = \boldsymbol{\phi}_t^\top \boldsymbol{\alpha}_t \qquad (51)$$

where $\boldsymbol{\phi}_t$ is the wavelet basis vector at time t, and α_k denotes the wavelet coefficients, where the two vectors are of size k. A blind receiver employing MKF was proposed for joint estimation of the wavelet coefficients and symbol detection. What is more, the receiver treats k as a static unknown parameter and evolves according to $k_t = k_{t-1}$. As a result, the blind receiver updates the number of wavelets dynamically and requires no channel statistics.

C. Some Examples

In this subsection, we present results of two sets of experiments are presented. In all the simulations, it is assumed that data are transmitted continuously without pilot symbols or any form of reinitialization bits in between. Hence, the data are transmitted even during harsh conditions when the instantaneous signal-to-noise ratio (SNR) is very low, which happens during zero fading.

C.1 Rayleigh fading with Gaussian channel noise – AR parameters known

The first experiment considered a Rayleigh fading channel with additive Gaussian noise. A BPSK modulation scheme was used for data transmission, with the symbols $s_t = -1$ and $s_t = 1$ being equally likely. Data were differentially encoded to mitigate the phase ambiguity problem. All three detectors from subsection IV-A were implemented for a channel with normalized Doppler spreads set to $f_d = 0.001$, which corresponds to slow fading, and $f_d = 0.01$, which is a fast fading scenario. An AR(3) process was used to model the channel, where the AR coefficients are a function of the Doppler spread and are obtained from the method suggested in [56]. The particle filter detectors were compared with the *clairvoyant* detector, which performs matched filtering and detection assuming that the channel is known exactly at the receiver. Thus, it serves as an unachievable upper performance bound. The number of particles chosen for Detectors I and II was $M = 1000$, while for Detector III was $M = 50$.

In Figure 8, BERs as functions of SNR for the slow fading case ($f_d = 0.001$) are plotted. The AR coefficients are given by $(a_1, a_2, a_3) \equiv (-2.9916, 2.9833, -0.9917)$. From the figure, it can

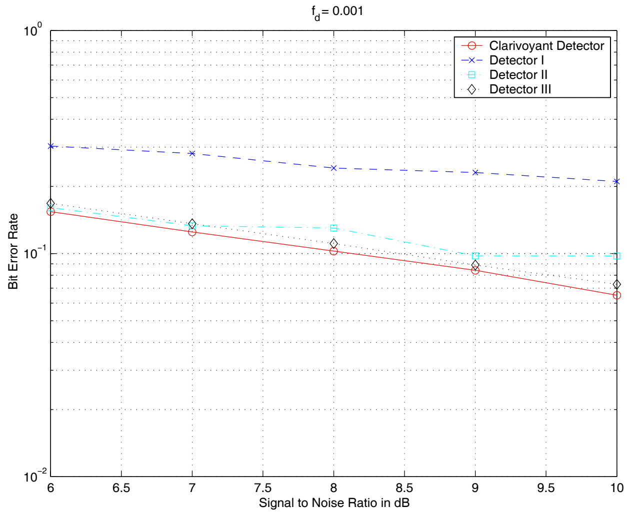

Fig. 8. Bit error rate performance of Detectors I, II and III over flat Rayleigh fading with $f_d = 0.001$ for differentially encoded BPSK signaling.

be seen that Detectors II and III perform similarly. However, it was observed that as the SNR was increased, Detector II degraded in performance. Detector I does not work well in slow fadings compared to the other two detectors.

Figure 9 shows the BERs as functions of SNR for the fast fading case, where $f_d = 0.01$. The AR coefficients are given by $(a_1, a_2, a_3) \equiv (-2.9145, 2.8344, -0.9197)$. Detectors I and II have equally good performance and are only slightly better than Detector III.

C.2 Rayleigh fading with Gaussian channel noise – AR parameters unknown

Figure 10 displays the BER performance versus SNR of the blind Detector II from subsection IV-B.2. The fading process was generated using the Jakes' method with 8 oscillators and $\Omega_d = 0.03$. We can see that Detector II clearly outperforms the differential detector and performs closely to the lower bound achieved by the pilot aided MKF. For the pilot aided MKF, 1000 pilot symbols were first used to estimate the AR coefficients (the modified covariance method [20], [26] was employed

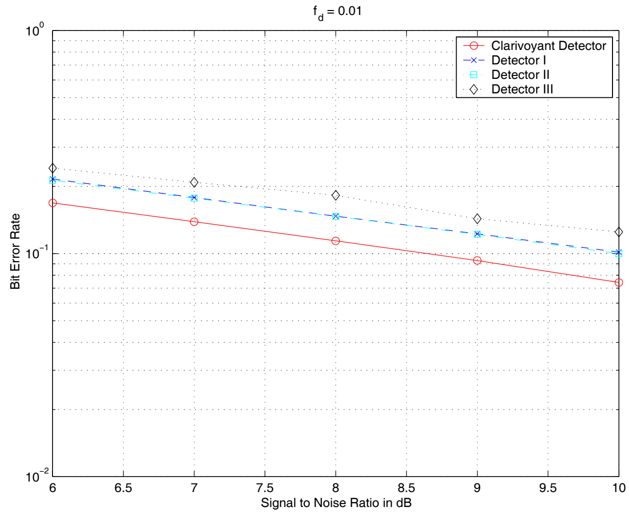

Fig. 9. Bit error rate performance of Detectors I, II and III over flat Rayleigh fading with $f_d = 0.01$ for differentially encoded BPSK signaling.

for the estimation), and then MKF was implemented with the estimated coefficients set as true coefficients.

V. Particle filtering for multiuser detection

Multiuser detection (MUD) has received a great deal of attention since the eighties due to its potential for increasing CDMA system capacity. A specific feature of MUD is that it does not treat the multiple access interference (MAI) present in CDMA systems as noise but as information. Since the optimum MUD is exponential in complexity, numerous approximate detectors have been developed to reduce that complexity. However, the performance of these detectors is suboptimal since they use interim hard decisions.

Application of particle filtering to MUD requires a representation of the system by a dynamic state space model. Since the symbols of a CDMA system are uncorrelated across different time slots, it is not obvious how to construct such representation.

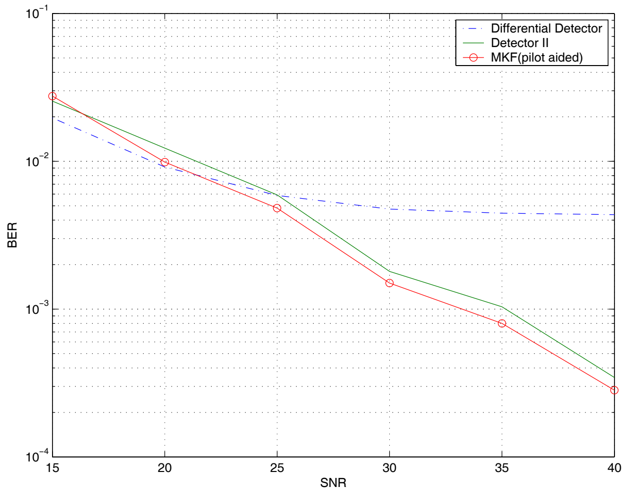

Fig. 10. BER performances of Detector II, the pilot aided MKF, and the differential detector. The fading channel was generated by Jakes' method and $\Omega_d = 0.03$.

The earliest application of particle filters to MUD appeared in [2] and subsequently in [45]. The state-space model is constructed using time dynamics of the fading channels and the symbols of all the users at a time slot are treated as one super symbol. As a result, the super symbol has a large alphabet whose size grows exponentially with the number of users. Recently in [58], [59], an alternative state-space representation of CDMA systems was proposed. It is based on whitened matched filter (WMF) outputs, where the dynamics of the system evolves with user index. This representation allows for an efficient application of particle filtering. Here we also note that MUD methods based on Markov chain Monte Carlo sampling were also proposed [53], [54].

A. State-space representations

Consider a synchronous CDMA system with chip-rate (processing gain) C and K users. Let T_s denote symbol duration and $\xi_k(\tau)$ the normalized deterministic signature waveform assigned to the $k-$th user. Here $\tau \in [0, T_s]$, and $k \in \{1, \ldots, K\}$. Let $s_k(\tau) \in \{-1, +1\}$ be the symbol transmitted by the $k-$th user, $h_k(\tau)$ the fading coefficient of the $k-$th user, and $v(\tau)$ the received zero mean complex white Gaussian noise with variance σ_v^2. We can express the received signal, $y(\tau)$, as

$$y(\tau) = \sum_{k=1}^{K} h_k(\tau) s_k(\tau) \xi_k(\tau) + v(t) \qquad \tau \in [0, T_s]. \tag{52}$$

After sampling at the system chip rate and modeling the Rayleigh flat fading channels of all users as ARMA processes as in (26), where the ARMA coefficients are chosen to fit the spectra of the fading processes. Here, without loss of generality, we assume that $r_a = r_b = r$. We can write for the state space representation

$$\boldsymbol{\rho}_t = \widetilde{\mathbf{A}} \boldsymbol{\rho}_{t-1} + \mathbf{E}\mathbf{u}_t \tag{53}$$

$$\mathbf{h}_t = \mathbf{B} \boldsymbol{\rho}_t \tag{54}$$

where $\boldsymbol{\rho}_t^\top = [\rho_{1,t}\ \rho_{1,t-1}\ \cdots\ \rho_{1,t-r}\ \rho_{2,t}\ \cdots\ \rho_{K,t-r}]$ is an auxiliary state vector of size $(r+1)K \times 1$ introduced to facilitate the representation, and $\mathbf{h}_t = [h_{1,t}\ \cdots\ h_{K,t}]^\top$ is the fading coefficient vector of all users. The matrices $\widetilde{\mathbf{A}}$ and $\mathbf{B}$ are known and of sizes $(r+1)K \times (r+1)K$ and $K \times (r+1)K$ respectively, where $\widetilde{\mathbf{A}}$ is a block diagonal matrix, i.e., $\widetilde{\mathbf{A}} = \text{diag}\{\mathbf{A}, \mathbf{A}, \cdots, \mathbf{A}\}$, with $\mathbf{A}$ being an $(r_a+1) \times (r_a+1)$ matrix defined by (28), and

$$\mathbf{B}^\top = \begin{bmatrix} \mathbf{b} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{b} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{b} \end{bmatrix}$$

where $\mathbf{b}^\top = [b_0\ b_1\ \cdots b_r]$ is a vector of the MA parameters of the ARMA process. Finally, $\mathbf{E}$ is an $(r+1)K \times K$ matrix with zero elements except for 1's at positions $((k-1)r+1, k)$, where $k = 1, 2, \cdots, K$, and $\mathbf{u}_t$ is a $K \times 1$ noise vector. The observations can now be written as

$$y_t = \mathbf{s}_t^\top \boldsymbol{\Xi} \mathbf{B} \boldsymbol{\rho}_t + v_t \tag{55}$$

where $\mathbf{s}_t = [s_{1,t} \ \cdots, s_{K,t}]^\top$, and $\boldsymbol{\Xi}$ is a $K \times K$ diagonal matrix of spreading codes.

In the algorithm from [2], [45], (53) and (55) form the representation of the system that evolves based on the chip duration, whereas the parameter of interest $\mathbf{s}_t$ is static within each symbol time. Samples of $\mathbf{s}_t$ are taken from an alphabet of size 2^K, and the channel state is integrated out. This is equivalent to the MKF algorithm proposed in [6], [11]. With large number of users, the alphabet of $\mathbf{s}_t$ grows exponentially and the calculation of the importance weight becomes computationally very expensive. In [45] and [46], it was found that deterministic methods which preserve the most likely particles were the most efficient.

In the algorithm from [58], [59], WMF outputs were utilized for the system representation because they are sufficient statistics of the transmitted symbols [52]. We can express the matched filter output as

$$\mathbf{y}_t = \mathbf{R}\mathbf{H}_t\mathbf{s}_t + \mathbf{v}_t \tag{56}$$

where $\mathbf{R}$ is the crosscorrelation matrix whose $ij-$th element is defined by $\mathbf{R}_{ij} =< \xi_i, \xi_j >= \int_0^{T_s} \xi_i(\tau)\xi_j(\tau)d\tau$, $\mathbf{H}_t = \text{diag}\{h_{1,t}, \cdots, h_{K,t}\}$ is the diagonal matrix of the channel fading coefficients which are considered static within each symbol interval, $\mathbf{s}_t^\top = [s_{1,t} \ \cdots, s_{K,t}]$ is the user symbol vector, and $\mathbf{v}_t$ is a complex-valued Gaussian vector with independent real and imaginary components and covariance matrix equal to $\sigma_v^2\mathbf{R}$.

The cross-correlation matrix is positive definite, and Cholesky factorization can be employed. There exists a unique lower triangular matrix $\mathbf{F}$ such that $\mathbf{R} = \mathbf{F}^\top\mathbf{F}$. When we apply $\mathbf{F}^{-\top} = (\mathbf{F}^\top)^{-1}$

to the matched filter output, we obtain

$$\bar{\mathbf{y}}_t = (\mathbf{F}^\top)^{-1}\mathbf{y}_t \tag{57}$$

$$= \mathbf{F}\mathbf{H}_t\mathbf{s}_t + \bar{\mathbf{v}}_t \tag{58}$$

$$= \mathbf{F}\mathbf{S}_t\mathbf{h}_t + \bar{\mathbf{v}}_t \tag{59}$$

where $\mathbf{S}_t = \mathrm{diag}(s_{1,t}, s_{2,t}, \cdots, s_{K,t})$ is the user symbol matrix. It can be verified that the covariance matrix of $\bar{\mathbf{v}}_t$ is $\sigma_v^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Since the noise becomes i.i.d. white Gaussian, $\bar{\mathbf{y}}_t$ is called the whitened matched filter output. Component-wise, it can be written as

$$\bar{y}_{k,t} = \sum_{i=1}^{k} F_{k,i}h_{i,t}s_{i,t} + \bar{v}_{k,t}. \tag{60}$$

B. Particle filtering implementation

B.1 MUD when the fading coefficients are known

In this case the unknowns are the user symbols which are uncorrelated across different time slots. In the following presentation, we drop the time subscript t. First, note that s_k is independent of $\bar{y}_{1:k-1}$. Therefore, the posterior distribution of the symbols of the first k users can be factored according to

$$\begin{aligned} p(s_{1:k}|\bar{y}_{1:k}) &= \frac{p(\bar{y}_k|s_{1:k}, \bar{y}_{1:k-1})p(s_{1:k}|\bar{y}_{1:k-1})}{p(\bar{y}_k|\bar{y}_{1:k-1})} \\ &\propto p(\bar{y}_k|s_{1:k})p(s_k)p(s_{1:k-1}|\bar{y}_{1:k-1}). \end{aligned} \tag{61}$$

Now if we choose an importance function of the form

$$\begin{aligned} \pi(s_{1:k}|\bar{y}_{1:k}) &= p(s_k|s_{1:k-1}, \bar{y}_{1:k})p(s_{k-1}|s_{1:k-2}, \bar{y}_{1:k-1}) \cdots p(s_1|y_1) \\ &= p(s_k|s_{1:k-1}, \bar{y}_{1:k})\pi(s_{1:k-1}|\bar{y}_{1:k-1}) \end{aligned} \tag{62}$$

we can create trajectories from $p(s_{1:K}|\bar{y}_{1:K})$ using importance sampling recursively and the standard particle filtering procedure.

During the $k-$th recursion, the trajectory $s_{1:k}^{(m)}$ is weighted with respect to $p(s_{1:k}^{(m)}|\bar{y}_{1:k})$ according to

$$\begin{aligned} w_k^{(m)} &= \frac{p(s_{1:k}^{(m)}|\bar{y}_{1:k})}{\pi(s_{1:k}^{(m)}|\bar{y}_{1:k})} \\ &\propto w_{k-1}^{(m)} p(\bar{y}_k|s_{1:k-1}^{(m)}, \bar{y}_{1:k-1}) \\ &= w_{k-1}^{(m)} \eta_k^{(m)}. \end{aligned} \quad (63)$$

Observe that at recursion k, the $\eta_k^{(m)}$'s are all that we need to have for calculating the weights $w_k^{(m)}$.

We note that drawing particles from the importance function $\pi(s_k|s_{1:k-1}^{(m)}, \bar{y}_{1:k}) = p(s_k|s_{1:k-1}^{(m)}, \bar{y}_{1:k})$ is easy since $s_k \in \{+1, -1\}$. In particular, we can write

$$\begin{aligned} p(s_k = 1|s_{1:k-1}^{(m)}, \bar{y}_{1:k}) &\propto p(\bar{y}_k|s_k = 1, s_{1:k-1}^{(m)}, \bar{y}_{1:k-1}) p(s_k = 1|s_{1:k-1}, \bar{y}_{1:k-1}) \\ &= p(\bar{y}_k|s_k = 1, s_{1:k-1}^{(m)}) p(s_k = 1) \end{aligned} \quad (64)$$

where $p(\bar{y}_k|s_k, s_{1:k-1}^{(m)})$ and $p(s_k)$ are the likelihood function and the prior distribution at recursion k, respectively, and they can be easily computed. An analogous expression can be written for $p(s_k = -1|s_{1:k-1}^{(m)}, \bar{y}_{1:k})$. Next, observe that $\eta_k^{(m)} = \sum_{s_k} p(\bar{y}_k|s_k, s_{1:k-1}^{(m)}) p(s_k)$, and thus the incremental weight is proportional to the sum of the importance function from (64), which is also readily obtained.

As in other applications of particle filtering, resampling is required in the algorithm. However, the weight must be clearly associated with all the particles in the trajectory at all times (that is, for all users). This is unlike in other applications where only the present particles in the trajectories are retained and therefore after resampling, the connection between the present weights and previous particles is lost.

The complexity of the algorithm is $O(KM)$, i.e., proportional to the product of the number of particles and the number of users. If the number of particles is fixed, the complexity is only linear with respect to the number of users.

B.2 MUD when the fading coefficients are unknown

When considering the joint channel estimation and MUD problem, we have to incorporate the state-space representation of the Rayleigh fading channel in (53) and (54) in the final state-space representation. If we assume that the channel is static within each symbol duration T_s, the WMF output at time t can be written as

$$\bar{\mathbf{y}}_t = \mathbf{F}\mathbf{S}_t\mathbf{B}\boldsymbol{\rho}_t + \bar{\mathbf{v}}_t. \tag{65}$$

With $\mathbf{f}_k^\top$ representing the $k-$th row of $\mathbf{F}$, the WMF output corresponding to the $k-$th user can be expressed as $\bar{y}_{k,t} = \mathbf{f}_k^\top\mathbf{S}_t\mathbf{B}\boldsymbol{\rho}_t + \bar{v}_{k,t}$. If we process a data block of T symbols, we can organize the WMF outputs into a sequence of KT observations, i.e., $\bar{\mathbf{y}}_{1:KT} = \{\bar{y}_{1,1} \ \cdots \ \bar{y}_{K,1} \ \cdots \ \bar{y}_{1,T} \ \cdots \ \bar{y}_{K,T}\}$. We introduce now a new index $i = (t-1)K + k$, and we summarize the state-space representation by

$$\begin{aligned}
\widetilde{\mathbf{S}}_i &= \begin{cases} \mathbf{X}_i, & \text{if } \mathrm{mod}(i,K) = 1 \\ \widetilde{\mathbf{S}}_{i-1} + \mathbf{X}_i, & \text{else} \end{cases} \\
\boldsymbol{\rho}_i &= \begin{cases} \mathbf{A}\boldsymbol{\rho}_{i-1} + \mathbf{E}\mathbf{u}_i, & \text{if } \mathrm{mod}(i,K) = 1 \\ \boldsymbol{\rho}_{i-1}, & \text{else} \end{cases} \\
\bar{y}_i &= \mathbf{f}_k^\top\widetilde{\mathbf{S}}_i\mathbf{B}\boldsymbol{\rho}_i + \bar{v}_i
\end{aligned} \tag{66}$$

where $\mathbf{X}_i$ is a $K \times K$ matrix whose only non-zero element is $(\mathbf{X}_i)_{kk} = s_{k,i}$. As we will see later, such representation facilitates the application of particle filtering.

The posterior distribution can be factored based on the state-space equations in (66),

$$p(\widetilde{\mathbf{S}}_{1:i}|\bar{\mathbf{y}}_{1:i}) \quad \propto \quad p(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}_{i-1})p(\widetilde{\mathbf{S}}_{1:i-1}|\bar{\mathbf{y}}_{1:i-1}) \int p(\bar{y}_i|\widetilde{\mathbf{S}}_i, \boldsymbol{\rho}_i)p(\boldsymbol{\rho}_i|\widetilde{\mathbf{S}}_{1:i-1}, \bar{\mathbf{y}}_{1:i-1})d\boldsymbol{\rho}_i \tag{67}$$

and if we choose the importance function in the form of

$$\pi(\widetilde{\mathbf{S}}_{1:i}|\bar{\mathbf{y}}_{1:i}) = \pi(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}_{1:i-1}, \bar{\mathbf{y}}_{1:i})\pi(\widetilde{\mathbf{S}}_{1:i-1}|\bar{\mathbf{y}}_{1:i-1})$$

we can obtain particles using the MKF algorithm. Specifically, the optimal importance function is $p(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i})$ which can be evaluated from

$$p(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i}) \propto p(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}) \int p(\bar{y}_i|\widetilde{\mathbf{S}}_i, \boldsymbol{\rho}_i)p(\boldsymbol{\rho}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i-1})d\boldsymbol{\rho}_i. \tag{68}$$

Note that given $\widetilde{\mathbf{S}}^{(m)}_{1:i-1}$, the state-space model in (66) is linear in $\boldsymbol{\rho}_i$ and is Gaussian, and the term $p(\boldsymbol{\rho}_i \mid \widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i-1})$ can be obtained via the prediction step of the Kalman filter. In turn, the integration in (68) can be evaluated analytically. The term $p(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1})$ is equal to zero if its first $k-1$ elements on the diagonal do not match those of $\widetilde{\mathbf{S}}_{i-1}$; otherwise, it is equal to $1/2$.

The importance weight becomes

$$\begin{aligned} w_i^{(m)} &= \frac{p\left(\widetilde{\mathbf{S}}^{(m)}_{1:i}|\bar{\mathbf{y}}_{1:i}\right)}{p\left(\widetilde{\mathbf{S}}^{(m)}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i}\right)\pi\left(\widetilde{\mathbf{S}}^{(m)}_{1:i-1}|\bar{\mathbf{y}}_{1:i-1}\right)} \\ &\propto w_{i-1}^{(m)} p\left(\bar{y}_i|\widetilde{\mathbf{S}}_{1:i-1}, \bar{\mathbf{y}}_{1:i-1}\right) \\ &= w_{i-1}^{(m)} \sum_{\widetilde{\mathbf{S}}_i} p\left(\widetilde{\mathbf{S}}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}\right) \int p\left(\bar{y}_i|\widetilde{\mathbf{S}}_i, \boldsymbol{\rho}_i\right) p\left(\boldsymbol{\rho}_i|\widetilde{\mathbf{S}}^{(m)}_{1:i-1}, \bar{\mathbf{y}}_{1:i-1}\right) d\boldsymbol{\rho}_i. \end{aligned} \tag{69}$$

Further details of the algorithm can be found in [59].

C. Simulations

For the algorithm based on WMF outputs, first we simulated a synchronous CDMA system with $K = 15$ users and a chip-rate of $C = 30$. The spreading codes were generated randomly and the same spreading code was used in all simulated detectors. Residual resampling [37] was performed after every 5 users. In Figure 11 we present a performance comparison for the equal power case with other popular CDMA MUD detectors: the 3-stage successive interference cancellation detector with decorrelating first stage (3-stage), the Gibbs sampler,[2] and the decorrelating decision feedback detector (DDF). For the Gibbs sampler, we experimented with two detectors whose difference was in the length of the burn-in periods (periods until convergence). The first detector drew 100 samples of which the first 50 were discarded (Gibbs-50). The second detector generated 150 samples, and the first 100 samples were discarded (Gibbs-100). As a reference, the Breadth-first tree-search algorithm which is optimum as described in [47] was simulated to provide a lower bound. In the simulation of the particle filtering method, we used 50 particles (PF-50) and 100 particles (PF-100). The performance curve in Figure 11 was obtained by averaging the BER of all 15 users. We can see that in the equal power case, particle filtering can provide near-optimum performance. It seems that the performance gain by increasing the number of particles from 50 to 100 is only marginal.

In the near-far case, the targeted user (the first user), had an SNR of 9 dB. The signal strength of the remaining 14 users was identical and compared with that of the targeted user, and was varied from -10 dB to 10 dB. The results of the experiment are shown in Figure 12. It is clear that particle filtering performs almost always better than the 3-stage detector and although it performs worse than the Gibbs sampler with weaker interferers, it is more consistent than the

[2]In our simulations we include the Gibbs sampler because it is another Monte Carlo based technique where estimates and detection are based on generated samples from desired a posteriori distributions.

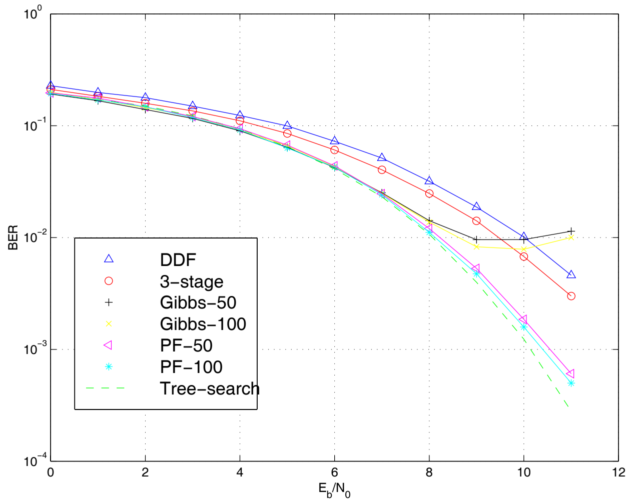

Fig. 11. Performance comparison of various detectors for $C = 30, K = 15$, equal powers, and known channels.

Gibbs sampler. Also, it is near optimum in the range from -4 dB to 10 dB.

Then we simulated the performance of the algorithm based on WMF outputs in the case of joint channel estimation and MUD. We simulated a quasi-static Rayleigh flat fading case where the channel was considered static within a block of $L + l_0$ symbol times. The first l_0 symbols were known to the receiver and they were sent as pilot signals. The fading coefficient within each block for each user was a complex Gaussian random variable with unit variance and zero mean. The fading coefficients of different blocks were considered independent from each other. We simulated blocks of $L = 10$ and $l_0 = 1$. For comparison, we also simulated the SAGE-JDE($\uparrow$) algorithm from [28]. For every simulated point, at least 300 bit errors were accumulated. In Figure 13, we see that the particle filtering algorithm consistently outperformed SAGE-JDE($\uparrow$).

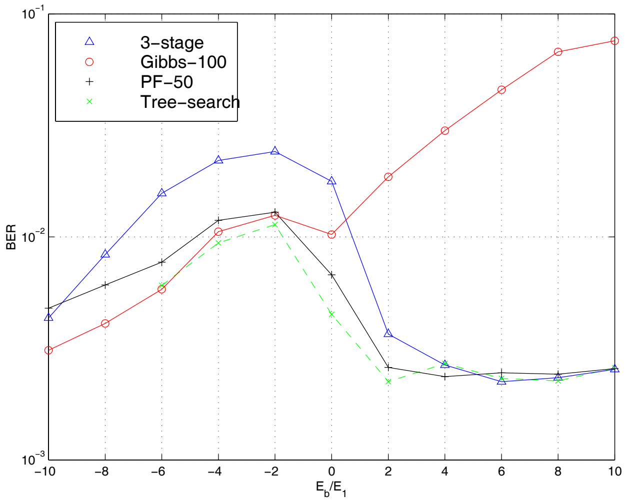

Fig. 12. Performance comparison of various detectors for $C = 30, K = 15$, known channels and near-far resistance of 9 dB.

VI. Estimation and detection of space-time codes in fading channels

Space-time coding (STC) originally introduced in [13] and further developed in [51] provides a framework for exploiting spatial and temporal diversity to increase data rate in wireless communications. Although space-time trellis codes (STTCs) are deemed to possess the best coding efficiency among space-time codes, they are hard for detection especially when unknown time varying fading coefficients are involved. As demonstrated in [7] and [44], it is quite straightforward to represent binary or M-ary convolutional (trellis) coded systems in fading channels using state-space models, and in [57] particle filtering was considered for this problem.

Suppose that a communication system employs N_T transmit and N_R receive antennas and that a sequence of user data symbols, $s_0, ..., s_t$, where $s_t \in \mathcal{A}$, is put through a trellis space-time encoder. The new state vector of the trellis space-time encoder at time t is determined according to the state transition equation $\mathbf{s}_t = f(\mathbf{s}_{t-1}, s_t)$, where $\mathbf{s}_{t-1}$ is the previous state, and $\mathbf{s}_t$ is the new user state. Based on the current state vector, the space-time encoder then generates a set of N_T symbols,

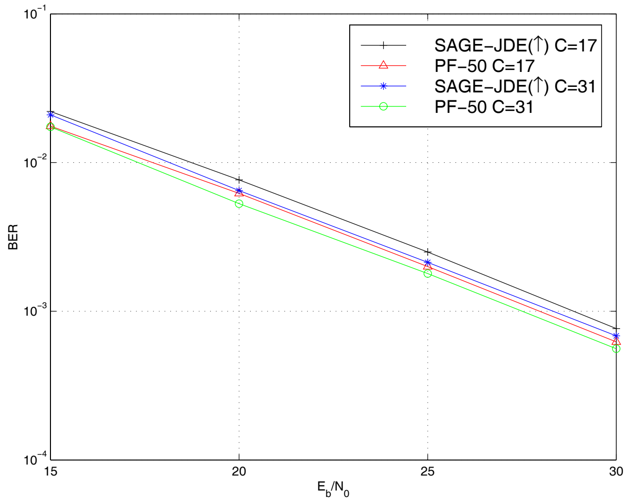

Fig. 13. $C = 17/31, K = 10$ equal power in Rayleigh quasi-static flat fading.

$\mathbf{c}^\top(\mathbf{s}_t) = [c_1(\mathbf{s}_t), ..., c_{N_T}(\mathbf{s}_t)]$, to be transmitted by the N_T antennas, where $c_i(\cdot)$ denotes the code and modulation function of the $i-$th antenna. Suppose $h_{n_T n_R,t}$ is the fading coefficient from the n_T-th transmit antenna to the n_R-th receive antenna at time t. Let $\mathbf{h}_{n_R,t}^\top = [h_{1n_R,t} \cdots h_{N_T n_R,t}]$ represent the set of channel states from all transmit antennas to the n_R-th receive antenna. If we arrange all the channel states at time t into a single $N_T N_R \times 1$ vector $\mathbf{h}_t^\top = [\mathbf{h}_{1,t}^\top \cdots \mathbf{h}_{N_R,t}^\top]$, all the received signals at time t can be written in vector form as

$$\mathbf{y}_t = \mathbf{C}(\mathbf{s}_t)\mathbf{h}_t + \mathbf{v}_t \tag{70}$$

where $\mathbf{y}_t = [y_{1,t} \cdots y_{N_R,t}]^\top$ is the received signal vector, and $\mathbf{v}_t = [v_{1,t} \cdots v_{N_R,t}]^\top$ is the noise vector. The code and modulation matrix $\mathbf{C}(\mathbf{s}_t)$ is an $N_T \times N_T N_R$ matrix of the form

$$\mathbf{C}^\top(\mathbf{s}_t) = \begin{bmatrix} \mathbf{c}(\mathbf{s}_t) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}(\mathbf{s}_t) & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{c}(\mathbf{s}_t) \end{bmatrix}.$$

Note that here $\mathbf{0}$ is an $N_T \times 1$ zero vector. This somewhat odd matrix representation is selected to simplify the description of the joint estimation and decoding algorithm.

The fading coefficients from the N_T-th to the N_R-th antenna are modeled in the same way as in (53) and (54) except that the subscript $n_T n_R$ is added, i.e.,

$$\boldsymbol{\rho}_{n_T n_R,t} = \mathbf{A}_{n_T n_R} \boldsymbol{\rho}_{n_T n_R,t-1} + \mathbf{E}_{n_T n_R} \mathbf{u}_{n_T n_R,t} \quad (71)$$

$$\mathbf{h}_{n_T n_R,t} = \mathbf{B}_{n_T n_R} \boldsymbol{\rho}_{n_T n_R,t} \quad (72)$$

where $\mathbf{A}_{n_R n_T}$, $\mathbf{E}_{n_T n_R}$, and $\mathbf{B}_{n_T n_R}$ have the same definition as before.

We can represent the whole system in a compact state-space form as follows

$$\boldsymbol{\rho}_t = \widetilde{\mathbf{A}} \boldsymbol{\rho}_{t-1} + \mathbf{E}\mathbf{u}_t \quad (73)$$

$$\mathbf{h}_t = \mathbf{B} \boldsymbol{\rho}_t \quad (74)$$

where $\boldsymbol{\rho}_t = [\boldsymbol{\rho}_{11,t}, \cdots, \boldsymbol{\rho}_{n_T 1,t}, \cdots, \boldsymbol{\rho}_{1 n_R,t}, \cdots \boldsymbol{\rho}_{n_T n_R,t}]^\top$ is the extended state vector, and the matrices $\widetilde{\mathbf{A}}$, $\mathbf{B}$, and $\mathbf{E}$ are constructed accordingly from the matrices $\mathbf{A}_{n_R n_T}$, $\mathbf{E}_{n_T n_R}$, and $\mathbf{B}_{n_T n_R}$, $n_T = 1, 2, \cdots, N_T$, $n_R = 1, 2, \cdots, N_R$, respectively. Note that in the algorithm described above, it is not required that the fading coefficients or the noise vector at the receive antennas be independent as required in most other algorithms.

Since the space-time code is trellis coded, besides the current and previous received signals $\mathbf{y}_{0:t}$, future observations also hold information about the current user state. Hence it is appropriate to use the delayed importance function as well as the delay weight method in evaluating the posterior distribution function. The number of delays can be chosen according to the constraint length of the trellis code.

A. Simulation

We simulated an STTC system with two transmit and one receive antenna. The ARMA model for the fading coefficient is described in [7], and it was of order $(3, 3)$. This model corresponds to a fast fading scenario with a normalized Doppler frequency (with respect to the symbol rate $1/T_s$) $f_d T_s = 0.05$.

For our simulations, we designed a special STTC based on the Tarokh 8-state STTC [51] that can combat phase ambiguity utilizing the time correlation of the channels. The details of the design can be found in [57]. We compared the performance of the system using our STTC with that of a system that uses Tarokh's STTC and pilot signals. The SNR was varied between 20 and 25 dB. During our simulations, we found that we had to apply resampling frequently, which is essential for good performance of the algorithm. We used the same residual resampling process as described in [7] and resampling was performed every 5 steps. Because the constraint length is one in the STTC, the number of delayed weights and delayed samples was one. For every simulated point, at least 100 symbol errors were accumulated.

The comparison results are shown in Figure 14. There we also represent the performance of a detector that has exact information about the channel states. The simulations show that it is viable to use particle filtering for decoding STTC with unknown fading channels. In addition, a significant performance improvement can be obtained by using STTCs that can combat phase ambiguity [57].

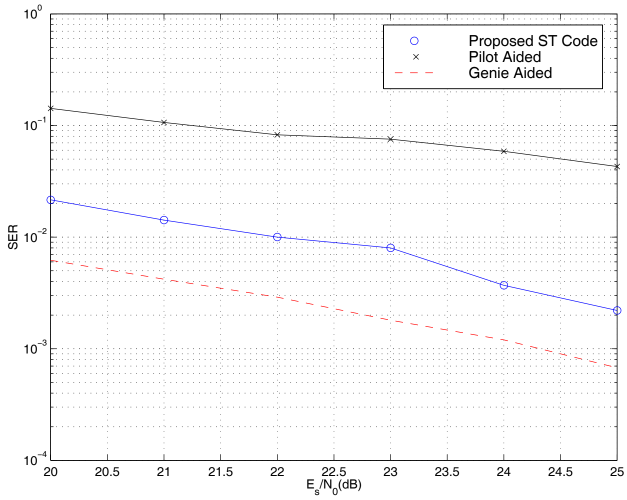

Fig. 14. Comparison results of a genie aided detector and particle filter detectors that use Tarokh's and a newly proposed STTCs.

VII. Summary of additional work on applications of particle filtering to communications

Besides the presented applications, there has been research in applying particle filtering to other problems in communications such as synchronization of communication systems [14], [40] and detection of signals in BLAST systems[3] [23]. In this section, we briefly summarize some of this work.

It is broadly recognized that many practical communication channels present a high degree of structure and that they can be accurately characterized through a set of reference parameters with a clear physical meaning. Since the observed signals collected by the receiver are affected by these parameters, they should be estimated and compensated prior to data detection in order to achieve *optimal* or *close-to-optimal* performance. The generalized synchronization problem consists of the recovery of a set of such physical parameters, namely the symbol timing, phase offset and carrier

[3]BLAST stands for Bell Laboratories Layered Space-Time.

frequency error. Up to date, many different techniques [39], [48] have been proposed in order to solve the synchronization problem, but they are based on approximate and heuristic methods, as optimal estimation of the parameters of interest seems to be analytically intractable [39].

Before particle filters are applied for synchronization, again, the observed signals are expressed in a state-space form. Synchronization, however, poses the additional difficulty in that the parameters of interest are usually modeled as fixed [39], while standard particle filtering algorithms are aimed at tracking time-varying unknowns.

In order to address the optimal recovery of the reference parameters and data detection, two different approaches are considered. In the first approach [40], the idea of rejuvenation [35] is applied in deriving a recursive algorithm aimed at approximating the joint smoothing probability distribution of the transmitted symbols and their (fixed) synchronization parameters. The other approach, [14], allows the synchronization parameters to evolve in a random way using *artificial evolution* mechanisms (the parameters are modeled as autoregressive stochastic processes [24], [42]) and therefore, they can be estimated using the traditional particle filtering approach. According to computer simulations, the two methods appear promising for joint data detection and synchronization.

In BLAST systems, spatial diversity is exploited to improve efficient transmission in broadband wireless communications. It is achieved by using multiple transmitting and receiving antennas, and it provides significant capacity gains. An optimum solution to the detection problem in BLAST systems is based on the maximum likelihood (ML) principle. Its complexity, however, is exponential with the number of transmitting antennas, and therefore its use in practice is prohibitive. A reasonable trade-off between complexity and performance was proposed where the detection proceeds along the signal layers in a decreasing order of their signal-to-noise ratio and where the detection in each layer is a two-step scheme of cancellation and nulling [13]. A weakness of this approach, known also as a V-BLAST receiver, is the propagation of error.

As an alternative, one can derive a dynamic state-space model of the BLAST system. The states evolve in space rather than in time, and the construction of the model is based on QR decomposition and the output of the feedforward filters. The proposed particle filters for the considered state-space model do not suffer from error propagation, and the simulations show that they greatly outperform the V-BLAST method and have near optimum performance [23].

References

[1] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.

[2] C. Andrieu, A. Doucet, and A. Touzni., "Adaptive MAP multi-user detection for fading CDMA channels," in *Proceedings of the IEEE Workshop on SSAP*, Pocono Manor, PA, 2000.

[3] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[4] T. Bertozzi, D. Le Ruyet, G. Rigal, and H. Vu-Thien, "Joint data-channel estimation using the particle filtering on multipath fading," in *the Proceedings of the ICT*, Tahiti Papeete, French Polynesia, 2003.

[5] T. Bertozzi, D. Le Ruyet, G. Rigal, and H. Vu-Thien, "On particle filtering for digital communications," in *the Proceedings of the IEEE Workshop on SPAWC*, Rome, Italy, 2003.

[6] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society*, vol. 62, no. Part 3, pp. 493–508, 2000.

[7] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat fading channels via mixture Kalman filtering," *IEEE Transactions on Information Theory*, vol. 46, pp. 2079–2094, 2000.

[8] T. C. Clapp and S .J. Godsill, "Fixed-lag blind equalization and sequence estimation in digital communication systems using sequential importance sampling," in *the Proceedings of IEEE ICASSP*, Phoenix, AZ, 1999.

[9] T. C. Clapp and S. J. Godsill, "Fixed-lag smoothing using sequential estimation in digital communications using sequential importance sampling," in *Bayesian Statistics 6*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds. Oxford University Press, 1999.

[10] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[11] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, pp. 197–208, 2000.

[12] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous Code-Division Multiple-Access channel," *IEEE Transactions on Communications*, vol. 41, no. 2, pp. 285–290, 1993.

[13] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[14] T. Ghirmai, M. F. Bugallo, and P. M. Djurić, "Joint data detection and symbol timing and symbol timing estimation using particle filtering," in *the Proceedings of IEEE ICASSP*, Hong Kong, 2003.

[15] T. Ghirmai, P. M. Djurić, M. F. Bugallo, and J. Kotecha, "Multisample receivers for time-varying channels using particle filtering," in *the Proceedings of the IEEE on Workshop SPAWC*, Rome, Italy, 2003.

[16] T. Ghirmai, J. Kotecha, and P. M. Djurić, "Joint channel estimation and data detection over frequency selective fading channels using sequential Monte Carlo filtering," in *the Proceeding of the CISS*, Princeton, NJ, 2002.

[17] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[18] D. Guo, X. Wang, and R. Chen, "Wavelet-based sequential Monte Carlo blind receivers in fading channels with unknown channel statistics," in *the Proceedings of IEEE ICC*, New York, NY, 2002.

[19] J. M. Hammersley and K. W. Morton, "Poor man's Monte Carlo," *Journal of the Royal Statistical Society, Series B*, vol. 16, no. 1, pp. 23–38, 1954.

[20] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, New York, 1996.

[21] Y. Huang and P. M. Djurić, "A blind particle filtering detector for joint channel estimation, tracking and data detection over flat fading channels," in *the Proceedings of EUSIPCO*, Toulouse, France, 2002.

[22] Y. Huang and P. M. Djurić, "A new importance function for particle filtering and its application to blind detection in flat fading channels," in *the Proceedings of IEEE ICASSP*, Orlando, FL, 2002.

[23] Y. Huang, J. Zhang, and P. M. Djurić, "Detection with particle filtering in blast systems," in *the Proceedings of the IEEE ICC*, Anchorage, AK, 2003.

[24] R. Iltis, "Joint estimation of PN code delay and multipath using the extended Kalman filter," *IEEE Transactions*

on Communications, vol. 38, pp. 1677–1685, 1990.

[25] W. C. Jakes, *Microwave Mobile Communnication*, IEEE Press, New York, 1993.

[26] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[27] S. M. Kay, *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[28] A. Kocian and B. H. Fleury, "Iterative joint symbol detection and channel estimation for DS/CDMA via the SAGE algorithm," in *the Proceedings of IEEE International Symposium on PIMRC*, London, UK, 2000.

[29] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, pp. 278–288, 1994.

[30] J. Kotecha and P. M. Djurić, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, Accepted for publication.

[31] J. Kotecha and P. M. Djurić, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, Accepted for publication.

[32] J. Kotecha and P. M. Djurić, "Gaussian sum particle filtering detector for impulsive fading channels," in *the Proceedings of EURASIP Workshop on NSIP*, Baltimore, MD, 2001.

[33] J. Kotecha and P. M. Djurić, "Sequential detector for nonlinear channels with application to satellite communications," in *the Proceedings of IEEE ICASSP*, Orlando, FL, 2002.

[34] J. H. Kotecha and P. M. Djurić, "Sequential Monte Carlo sampling detector for Rayleigh fast-fading channels," in *the Proceedings of IEEE ICASSP*, Istanbul, Turkey, 2000.

[35] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., pp. 197–223. Springer, 2001.

[36] J. S. Liu, "Metropolized independent sampling with comparison to rejection sampling and importance sampling," *Statsitics and Computing*, vol. 6, pp. 113–119, 1996.

[37] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2001.

[38] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 567–576, 1995.

[39] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, 1997.

[40] J. Miguez, M. F. Bugallo, and P. M. Djurić, "A sequential Monte Carlo algorithm for blind timing recovery and data detection," in *the Proceedings of IEEE Workshop on SPAWC*, Rome, Italy, 2003.

[41] J. Míguez and P. M. Djurić, "Blind equilization by sequential importance sampling," in *the Proceedings of IEEE ISCAS*, Phoenix, AZ, 2002.

[42] M. J. Omidi, S. Pasupathi, and P. G. Gulak, "Joint data and Kalman estimation for Rayleigh fading channels," *Wireless Personal Communication*, , no. 10, pp. 319–339, 1999.

[43] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 3 edition, 1995.

[44] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Particle filtering for demodulation in fading channels with non-Gaussian additive noise," *IEEE Transacations on Communications*, vol. 49, no. 4, pp. 579–582, 2001.

[45] E. Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Particle filtering for multiuser detection in fading CDMA channels," in *Proceedings of 11th IEEE Workshop on SSP*, Singapore, 2001.

[46] E. Punskaya, A. Doucet, and W.J. Fitzgerald, "On the use and misuse of particle filtering in digital communications," in *the Proceedings of EUSIPCO*, Toulouse, France, 2002.

[47] L. K. Rasmussen, T. J. Lim, and T. Aulin, "Breadth-first maximum-likelihood detection for multiuser CDMA," *IEEE Transactions on Communications*, vol. 45, pp. 1176–1178, 1997.

[48] J. Riba and G. Vázquez, "Bayesian recursive estimation of frequency and timing exploiting the cyclostationary property," *Signal Processing*, vol. 40, pp. 21–37, 1994.

[49] M. N. Rosenbluth and A. W. Rosenbluth, "Monte Carlo calculation of the average extension of molecular chains," *Journal of Chemical Physics*, vol. 23, no. 2, pp. 356–359, 1956.

[50] G. L. Stuber, *Mobile Communications*, Kluwer Academic Publishers, Norwell, MA, 1996.

[51] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Transactions on Information Theory*, pp. 744–765, 1998.

[52] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

[53] X. Wang and R. Chan, "Adaptive Bayesian multiuser detection for synchronous CDMA with Gaussian and non-Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 48, no. 7, pp. 2013–2029, 2000.

[54] Z. Yang and X. Wang, "Blind turbo multiuser detetction for long-code multipath CDMA," *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 112–125, 2002.

[55] Z. Yang and X. Wang, "A sequential Monte Carlo blind receiver for OFDM systems in frequency-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 271–280, 2002.

[56] H. Zamiri-Jafrian and S. Pasupathi, "Adaptive MLSD receiver with identification of flat fading channels," in

the Proceedings of IEEE ICASSP, Munich, Germany, 1997.

[57] J. Zhang and P. M. Djurić, "Joint estimation and decoding of space-time trellis codes," *EURASIP Journal on Applied Signal Processing*, vol. 2002, pp. 305–315, 2002.

[58] J. Zhang and P. M. Djurić, "Reduced complexity Monte Carlo smoothing with application to communications," in *the Proceedings of CISS*, Princeton, NJ, 2002.

[59] J. Zhang, Y. Huang, and P. M. Djurić, "Multiuser detection with particle filtering," in *the Proceedings of EUSIPCO*, Toulouse, France, 2002.

[60] R.A. Ziegler and J. M. Cioffi, "Estimation of time-varying digital radio channels," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 2, pp. 134–151, 1992.