Blind Equalization of Frequency-Selective Channels by Sequential Importance Sampling

Joaquín Míguez, *Member, IEEE,* and Petar M. Djurić, *Senior Member, IEEE*

***Abstract*—This paper introduces a novel blind equalization algorithm for frequency-selective channels based on a Bayesian formulation of the problem and the sequential importance sampling (SIS) technique. SIS methods rely on building a Monte Carlo (MC) representation of the probability distribution of interest that consists of a set of samples (usually called particles) and associated weights computed recursively in time. We elaborate on this principle to derive blind sequential algorithms that perform maximum *a posteriori* (MAP) symbol detection *without* explicit estimation of the channel parameters. In particular, we start with a basic algorithm that only requires the *a priori* knowledge of the model order of the channel, but we subsequently relax this assumption and investigate novel procedures to handle model order uncertainty as well. The bit error rate (BER) performance of the proposed Bayesian equalizers is evaluated and compared with that of other equalizers through computer simulations.**

***Index Terms*—Bayesian estimation, blind equalization, Monte Carlo methods, SIS algorithm.**

I. Introduction

FUTURE wideband wireless communication systems greatly depend on the development of sophisticated coding and signal processing techniques that provide high spectral efficiencies and allow a close approach to the theoretical capacity limits. One fundamental problem in this context is the detection of a symbol sequence transmitted through a frequency-selective channel. When the channel parameters are known and the transmitted symbols are independent and identically ditributed (i.i.d.) uniform random variables, optimal detection, which is based on the maximum likelihood (ML) principle, can be efficiently implemented by means of the Viterbi algorithm [1]. A straightforward way to acquire channel state information is to transmit training sequences that are known *a priori* by the transmitter and the receiver, but this approach results in an efficiency loss. Hence, a major stream of research has focused on *blind* methods, where symbols are detected without knowledge of the channel coefficients and without using any training data. This includes both linear equalizers aimed at symbol detection without explicit channel estimation (see [2], [3], and references therein), and joint channel estimation and symbol detection techniques using the expectation–maximization (EM) algorithm [4], the per survivor processing (PSP) method [5], [6], or similar procedures based on the Viterbi algorithm [7].

The last few years have witnessed a strong interest in the application of simulation-based methods to solve (hard) signal processing problems, due to the availability of powerful computing facilities. Thus, several Monte Carlo (MC) algorithms, formerly disregarded as being practically infeasible, have recently re-emerged and become popular signal processing tools. The common feature of these techniques is that they aim at building estimates from discrete random measures that approximate a desired probability distribution. Within the field of communications, both Markov chain Monte Carlo (MCMC) [8] and sequential importance sampling (SIS) techniques [9] have been applied to solve the channel equalization problem.

The most popular MCMC technique is the Gibbs sampler [8], which has been applied both in single-user and multiuser scenarios (see [10] and references therein). One important limitation of this approach is that the resulting receivers must operate in batch mode, i.e., they require the whole burst of observations containing information about the transmitted data to be available at the beginning of the processing. Moreover, it has been reported [11] that digital detectors based on the Gibbs sampler suffer from slow convergence for medium and high signal-to-noise ratios (SNRs).

The latter drawbacks are overcome by the SIS methodology. The MC estimate of the desired probability distribution built by SIS consists of particles and associated weights, both of them computed recursively as new observations are received. Under certain mild conditions and large number of particles, this MC representation allows for tight approximations of several types of estimators [12]. Specifically, SIS-based detectors attain a lower BER than receivers designed using the Gibbs sampler for the medium-and-high SNR region [11], are better suited for online processing than MCMC techniques, and naturally lend themselves to implementations with massively parallel hardware.

Propelled by the above mentioned advantages, a number of equalization methods based on SIS have been proposed in the literature, starting with the blind deconvolution technique of [13]. A suitable combination of the SIS algorithm and the well-known Kalman filter [14], termed the mixture Kalman filter (MKF), has been succesfully applied to the problem of equalizing frequency nonselective fading channels [15], [16]. Blind equalization of frequency-selective channels using SIS

Manuscript received June 20, 2002; revised October 10, 2003. This work was supported by the Ministry of Science and Technology of Spain and Xunta de Galicia under Project TIC2001-0751-C04-01 and the National Science Foundation under Award CCR-0082607. The associate editor coordinating the review of this paper and approving it for publication was Dr. Dennis R. Morgan.

J. Míguez is with the Departamento de Electrónica e Sistemas, Universidade da Coruña, Facultade de Informática, 15071 A Coruña, Spain (e-mail: jmiguez@udc.es).

P. M. Djurić is with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: djuric@ece.sunysb.edu).

Digital Object Identifier 10.1109/TSP.2004.834335

techniques has also been studied for orthogonal frequency division multiplexing (OFDM) systems [17].

In this paper, we directly address the blind equalization of a frequency-selective channel in a single carrier communication system. Within this framework, ISI is the main source of distortion to be mitigated. We elaborate on a Bayesian formulation of the problem and the SIS methodology in order to derive a blind scheme for maximum *a posteriori* (MAP) data detection that operates recursively in time and can attain nearly optimal performance in terms of BER. A key feature of the approach, shared with MKF techniques [15], [16] and the receivers in [13] and [17], is that symbol estimation is tackled without an explicit channel estimation stage. As a novelty with respect to the detectors in [13] and [15]–[17], where the number of channel coefficients is assumed known, we also study the realistic case where the model order of the channel is unknown and must be either estimated, together with the transmitted symbols, or marginalized. The problem of unknown channel order has received very little attention so far in the context of MC methods for communications, and it was very recently addressed for the first time [18]. For clarity of presentation, the derivation of the proposed equalizers is carried out within a simple framework (a single-input single-output system and binary data), but there are no theoretical obstacles preventing its extension to more complex scenarios.

The remainder of this paper is organized as follows. Section II describes the signal model for the equalization problem. The Bayesian formulation underlying the proposed blind equalizers is developed in Section III. In Section IV, the SIS algorithm is introduced and applied to the problem of equalizing an unknown channel with a known number of coefficients. Extensions of the latter method, which include the case of an unknown channel order and the technique of delayed sampling as a means of improving performance, are introduced and discussed in Section V. Illustrative computer simulations are presented in Section VI, and some concluding remarks are made in Section VII.

II. Signal Model

Consider a digital communication system where BPSK symbols $s_t \in \{\pm 1\}, t = 0, 1, 2, \ldots$ are transmitted in frames of length $T + 1$ through a frequency-selective multipath fading channel. When the coherence time of the fading process is long enough compared with the frame size, it is commonly assumed that the channel impulse response (CIR) is constant for the duration of the frame. In such cases, and assuming a receiver front-end consisting of a matched filter followed by a symbol rate sampler, the discrete-time sequence of observations can be written as [1]

$$y_t = \sum_{l=0}^{T} s_l h_{t-l} + v_t, \quad t = 0, 1, 2, \ldots, T \tag{1}$$

where $v_t \sim \mathcal{N}(0, \sigma^2)$ is an additive white Gaussian noise (AWGN) process with zero mean and variance σ^2, whereas $h_t, t \in \mathbb{Z}$, is the discrete-time equivalent CIR. In the practical case that the transmitted pulse waveforms are causal signals with limited duration, the discrete-time equivalent CIR becomes a causal finite-length sequence that can be conveniently represented by the $m \times 1$ vector

$$\mathbf{h} = [h_{m-1}, h_{m-2}, \ldots, h_0]^\top \tag{2}$$

where m is the channel order (it physically represents the number of resolvable propagation paths), and the superindex $^\top$ denotes transposition.

Assuming the transmitted bits are i.i.d. uniform random variables $s_t \sim \mathcal{U}(\pm 1)$, (1) and (2) allow for the modeling of the communication process over a frequency-selective channel by means of a dynamic system in state-space form:

$$\text{(state equation)} \quad \mathbf{s}_t = \mathbf{T}\mathbf{s}_{t-1} + \mathbf{u}_t \tag{3}$$

$$\text{(observation equation)} \quad y_t = \mathbf{s}_t^\top \mathbf{h} + v_t \tag{4}$$

where the $m \times 1$ vector $\mathbf{s}_t = [s_{t-m+1}, s_{t-m+2}, \ldots, s_t]^\top$ is the system state at time t, $\mathbf{T}$ is the $m \times m$ state-transition matrix such that $\mathbf{T}[s_{t-m-1}, \ldots, s_{t-2}, s_{t-1}]^\top = [s_{t-m}, \ldots, s_{t-1}, 0]^\top$, and the $m \times 1$ vector $\mathbf{u}_t = [0, 0, \ldots, s_t]^\top$ is the state perturbation. Notice that the channel order m determines the dimension of the state vector and, therefore, the span of the ISI.

Our aim is to find a sequential algorithm to compute the joint MAP estimate of the symbols transmitted in a single frame $s_{0:T} = \{s_0, s_1, \ldots, s_T\}$, when the channel vector $\mathbf{h}$ is unknown. The notation $\mathbf{a}_{i:j}$ is frequently used in the remainder of the paper to represent the set $\{\mathbf{a}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_j\}$.

III. Recursive Computation of the Posterior Probability

According to model (3) and (4), optimal blind equalization is achieved by MAP detection of the information bit sequence $s_{0:T}$, given the observations $y_{0:T}$. Let $p[s_{0:T} \,|\, y_{0:T}]$ represent the probability mass function (pmf) of the data sequence conditional on the corresponding series of observations. The MAP estimate of the transmitted symbols is

$$s_{0:T}^{(\mathrm{MAP})} = \arg\max_{s_{0:T}} \{p[s_{0:T} \,|\, y_{0:T}]\} \tag{5}$$

which can be solved in a "brute force" approach by computing the posterior probability of each one of the 2^{T+1} possible bit sequences and then selecting the one with the largest probability mass.

Due to computational complexity, this approach is out of question, and it is desirable to solve problem (5) sequentially and recursively, i.e., to obtain $s_{0:t}^{(\mathrm{MAP})}$ from $s_{0:t-1}^{(\mathrm{MAP})}$ when y_t is observed. To achieve this goal, we consider the following decomposition of the posterior pmf:

$$p[s_{0:t}|y_{0:t}] \propto p[y_t \,|\, s_{0:t}, y_{0:t-1}] p[s_{0:t-1} \,|\, y_{0:t-1}] \tag{6}$$

which has been obtained by applying Bayes theorem and the fact that $p[s_t \,|\, s_{0:t-1}, y_{0:t-1}] = p[s_t \,|\, s_{0:t-1}] = p[s_t]$ is a uniform pmf. Equation (6) provides the basis for sequential computation of $p[s_{0:t} \,|\, y_{0:t}]$, where it is assumed that the likelihood function $p[y_t \mid s_{0:t}, y_{0:t-1}]$ can be computed up to a proportionality constant. This is easily seen as we iterate (6) to obtain

$$p[s_{0:t} \mid y_{0:t}] \propto p[s_0]p[y_0 \mid s_0] \prod_{k=1}^{t} p[y_k \mid s_{0:k}, y_{0:k-1}]$$

where we have implicitly assumed that $s_{t<0}$ are *a priori* known. Notice that $s_{t<0}$ are neither information bits nor pilot symbols. If the data frame starts with s_0, we can usually assume the absence of signal $s_{t<0} = 0$. Otherwise, if the data frame is divided into several blocks for processing, $s_{t<0}$ can be assigned the values of previously detected bits.

The derivation of a closed-form expression for the likelihood $p[y_t \mid s_{0:t}, y_{0:t-1}]$ is addressed in Appendix A. In particular, it is shown that if the channel order m is known and the CIR vector has a prior Gaussian distribution

$$p[\mathbf{h}] = \mathcal{N}(\bar{\mathbf{h}}_{-1}, \mathbf{C}_{-1})$$

where $\bar{\mathbf{h}}_{-1}$ and $\mathbf{C}_{-1}$ are the *a priori* channel mean and covariance matrix, respectively, then the posterior channel pdf given a sequence of observations $y_{0:t}$ and transmitted symbols $s_{0:t}$ is also Gaussian, i.e.,

$$p[\mathbf{h} \mid s_{0:t}, y_{0:t}] = \mathcal{N}(\bar{\mathbf{h}}_t, \mathbf{C}_t)$$

and the distribution parameters can be updated recursively according to

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

Using the above relationships, the likelihood in (6) can be explicitly written as

$$\begin{aligned} &p[y_t \mid s_{0:t}, y_{0:t-1}] \\ &\quad = \frac{|\mathbf{C}_t|^{1/2}}{(2\pi\sigma^2 |\mathbf{C}_{t-1}|)^{1/2}} \\ &\qquad \times \exp\left\{ -\frac{1}{2} \left[\frac{y_t^2}{\sigma^2} + \bar{\mathbf{h}}_{t-1}^\top \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right.\right. \\ &\qquad \left.\left. - \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right)^\top \mathbf{C}_t \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right) \right] \right\}. \end{aligned} \tag{7}$$

Equations (6) and (7) are the basis for the recursive equalization algorithms proposed in this paper.

IV. Blind Equalization Using the SIS Algorithm

A. *SIS Algorithm with Known Channel Order*

The standard statement of the SIS algorithm (see [9, Sec. 2]) is concerned with dynamic systems in state-space form where all fixed parameters are known and only the state sequence has to be estimated. This is not the case of the model (3) and (4), where the CIR vector $\mathbf{h}$ is unknown (except for its order m); therefore, we present a slightly different derivation of the method here. We begin with an importance sampling (IS) scheme [19]. If we draw N particles from a trial pmf and weight them according to the true posterior probability of the symbols, we can write

$$\begin{aligned} s_{0:T}^{(i)} &\sim q[s_{0:T} \mid y_{0:T}] \\ \tilde{w}^{(i)} &= \frac{p\left[s_{0:T}^{(i)} \,\middle|\, y_{0:T} \right]}{q\left[s_{0:T}^{(i)} \,\middle|\, y_{0:T} \right]} \\ w^{(i)} &= \frac{\tilde{w}^{(i)}}{\sum_{k=1}^{N} \tilde{w}^{(k)}} \end{aligned}$$

where $q[s_{0:T} \mid y_{0:T}]$ is the trial or importance pmf with the same support as $p[s_{0:T} \mid y_{0:T}]$ but from which it is easier to sample, and $\{w^{(i)}\}_{i=1}^N$ is a set of normalized importance weights. These particles are said to be properly weighted, meaning that

$$\mathrm{E}_{q[s_{0:T} \mid y_{0:T}]}[g(s_{0:T})\tilde{w}] = \mathrm{E}_{p[s_{0:T} \mid y_{0:T}]}[g(s_{0:T})]$$

where $\mathrm{E}_p[\cdot]$ denotes statistical expectation with respect to the pmf in the subscript, and $g(\cdot)$ is an arbitrary integrable function of the state sequence.

The IS method can be modified so that one can build the state trajectories $s_{0:T}^{(i)}$ and the importance weights $w^{(i)}$ sequentially as new observations arrive. Consider the following factorization of the importance pmf:

$$q[s_{0:t} \mid y_{0:t}] \propto q[s_t \mid s_{0:t-1}, y_{0:t}] q[s_{0:t-1} \mid y_{0:t-1}]. \tag{8}$$

Working with (6) and (8) and the IS principle, the importance weights can be evaluated recursively in time, leading to the SIS algorithm

$$s_t^{(i)} \sim q\left[s_t \,\middle|\, s_{0:t-1}^{(i)}, y_{0:t} \right] \tag{9}$$

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p\left[y_t \,\middle|\, s_{0:t}^{(i)}, y_{0:t-1} \right]}{q\left[s_t^{(i)} \,\middle|\, s_{0:t-1}^{(i)}, y_{0:t} \right]} \quad \text{(normalized weights)} \tag{10}$$

for $i = 1, \ldots, N$. The set of particles and normalized weights at time t, $\{s_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ is a discrete random measure referred to as a *particle smoother* or *MC smoother* [9] that yields an MC estimate of the posterior pmf

$$p[s_{0:t} \mid y_{0:t}] \approx \hat{p}[s_{0:t} \mid y_{0:t}] = \sum_{i=1}^{N} \delta_i w_t^{(i)} \tag{11}$$

where $\delta_i = 1$ if $\mathbf{s}_{0:t} = \mathbf{s}_{0:t}^{(i)}$, and $\delta_i = 0$ otherwise. At time t, the particle smoother can be used to extract desired estimates using the posterior distribution. In this paper, we are interested in the MAP estimate of the data. Both joint sequence estimation at the end of the frame

$$\hat{s}_{0:T}^{(\mathrm{MAP})} := \arg\max_{s_{0:T}} \left\{ \sum_{i=1}^{N} \delta\left(s_{0:T} - s_{0:T}^{(i)} \right) w_T^{(i)} \right\} \tag{12}$$

and marginal data detection at time t,

$$\hat{s}_t^{(\mathrm{MAP})} := \arg\max_{s \in \{\pm 1\}} \left\{ \sum_{i=1}^{N} \delta\left(s - s_t^{(i)} \right) w_t^{(i)} \right\} \tag{13}$$

TABLE I
SIR AND D-SIR ALGORITHMS FOR BLIND MAP EQUALIZATION

$T+1 \equiv$frame size, $N \equiv$ number of trajectories		
Initialization:	$\mathbf{C}_{-1}^{(i)} := \mathbf{C}_{-1}$ $\overline{\mathbf{h}}_{-1}^{(i)} := \overline{\mathbf{h}}_{-1}, \forall i.$	
For $t = 0, 1, ..., T$: **For** $i = 1, ..., N$:		
Importance sampling:	SIR $s_t^{(i)} \sim q[s_t \| s_{0:t-1}^{(i)}, y_{0:t}]$	D-SIR $s_t^{(i)} \sim q[s_t \| s_{0:t-1}^{(i)}, y_{0:t+d}]$
Channel update:	$\mathbf{s}_t^{(i)} = [s_{t-m+1}^{(i)}, \ldots, s_t^{(i)}]^\top$ $\mathbf{C}_t^{(i)} = \left(\frac{\mathbf{s}_t^{(i)} \mathbf{s}_t^{(i)\top}}{\sigma^2} + \mathbf{C}_{t-1}^{(i)^{-1}} \right)^{-1}$ $\overline{\mathbf{h}}_t^{(i)} = \mathbf{C}_t^{(i)} \left(\frac{\mathbf{s}_t^{(i)} y_t}{\sigma^2} + \mathbf{C}_{t-1}^{(i)^{-1}} \overline{\mathbf{h}}_{t-1}^{(i)} \right)$	
Weight update:	SIR $\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p[y_t \| s_{0:t}^{(i)}, y_{0:t-1}]}{q_t[s_t^{(i)} \| s_{0:t-1}^{(i)}, y_{0:t}]}$	D-SIR $\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p[y_{t:t+d} \| s_{0:t}^{(i)}, y_{0:t-1}]}{q_t[s_t^{(i)} \| s_{0:t-1}^{(i)}, y_{0:t+d}]}$
Weight normalization:	$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}, \forall i$	
Resampling:	Let $N_{eff} = \frac{1}{\sum_{j=1}^N w_t^{(j)^2}}$ and choose $0 < \varepsilon < 1$. If $N_{eff} < \varepsilon$ and $t < T$ then: (1) for $j = 1, ..., N$, let $\tilde{s}_{0:t}^{(j)} = s_{0:t}^{(i)}$ with probability $w_t^{(i)}$; (2) let $s_{0:t}^{(i)} = \tilde{s}_{0:t}^{(i)}$, $i = 1, ..., N$; and (3) let $w_t^{(i)} = \frac{1}{N}$, $i = 1, ..., N$.	
MAP estimation: using equations (12) or (13)		

can be easily performed using the particle smoother.

B. Resampling

It can be shown [9], [20] that the variance of the importance weights $\{w_t^{(i)}\}_{i=1}^N$, when considered as random variables, can only increase over time. As a consequence, the resulting MC estimates deteriorate and become useless. One approach to alleviate the increase in the variance of weights is to include a *resampling* step in the SIS algorithm [9], [12]. We introduce a resampling step in the algorithm each time the effective sample size of the particle smoother,[1] which is estimated as [9]

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^N \left(w_t^{(i)^2}\right)^2} \leq N,$$

goes below a threshold ε. The resulting technique is termed the sequential importance sampling with resampling (SIR) algorithm. The SIR algorithm for blind MAP equalization is described in Table I.

C. Importance PMFs

The choice of importance function $q[\cdot]$ is up to the designer and it is usually made based on computational complexity and performance considerations. A simple choice is the prior importance function, but it is known that it is inefficient. Instead, we resort to the optimal importance function, which employs all the information available up to time t in order to propose new samples, i.e.,

$$q\left[s_t \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t}\right] = p\left[s_t \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t}\right] \propto p\left[y_t \;\middle|\; s_t, s_{0:t-1}^{(i)}, y_{0:t-1}\right].$$

Using the same methodology as in Appendix A, the likelihood in the above equation can be reduced to

$$p\left[y_t \;\middle|\; s_t, s_{0:t-1}^{(i)}, y_{0:t-1}\right] = \frac{\left|\tilde{\mathbf{C}}_t^{(i)}\right|^{1/2}}{\left(2\pi\sigma^2 \left|\mathbf{C}_{t-1}^{(i)}\right|\right)^{1/2}} \exp\Bigg\{-\frac{1}{2}\Bigg[\frac{y_t^2}{\sigma^2} + \bar{\mathbf{h}}_{t-1}^{(i)^\top} \mathbf{C}_{t-1}^{(i)^{-1}} \bar{\mathbf{h}}_{t-1}^{(i)} - \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{(i)^{-1}} \bar{\mathbf{h}}_{t-1}^{(i)}\right)^\top \times \tilde{\mathbf{C}}_t^{(i)} \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{(i)^{-1}} \bar{\mathbf{h}}_{t-1}^{(i)}\right)\Bigg]\Bigg\}$$

where $\mathbf{s}_t = [s_{t-m+1}^{(i)}, \ldots, s_{t-1}^{(i)}, s_t]^\top$ and $\tilde{\mathbf{C}}_t^{(i)^{-1}} = \mathbf{s}_t \mathbf{s}_t^\top / \sigma^2 + \mathbf{C}_{t-1}^{(i)^{-1}}$. To be specific, we draw the new sample $s_t^{(i)}$ using the trial pmf

$$q\left[s_t \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t}\right] = \frac{p\left[y_t \;\middle|\; s_t, s_{0:t-1}^{(i)}, y_{0:t-1}\right]}{\sum_{a \in \{\pm 1\}} p\left[y_t \;\middle|\; s_t = a, s_{0:t-1}^{(i)}, y_{0:t-1}\right]}. \quad (14)$$

[1]The effective sample size indicates the number of i.i.d. particles drawn from the true posterior pmf that would be necessary to obtain MC estimates with the same quality as those given by the weighted particles.

Correspondingly, the weight update equation becomes

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \sum_{a\in\{\pm 1\}} p\left[y_t \,\middle|\, s_t = a, s_{0:t-1}^{(i)}, y_{0:t-1}\right].$$

Note that the weight does not depend on the new sample that is added to the trajectory; therefore, its computation can be carried out in parallel with the sampling step.

V. Extensions

Statistical results regarding the convergence of the SIS algorithm [12] guarantee that the proposed blind MAP equalizer yields asymptotically optimal performance (as $N \to \infty$). However, there are practical situations where attaining close-to-optimal performance may require an extremely high number of particles. Environments with a very attenuated line of sight (LOS) between transmitter and receiver are typical examples. The reason is that the contribution of symbol s_t to the likelihood $p[y_t \mid s_{0:t}, y_{0:t-1}]$ depends mainly on the value of h_0 (the LOS channel component); as a consequence, both *good* and *bad* particles have similarly small likelihoods, and it is hard for the SIR algorithm to discriminate them. One promising strategy to circumvent this limitation is the use of the *delayed sampling* technique [15], [16], which will be investigated in Section V-A.

Another important practical issue is the *a priori* knowledge of the channel order m assumed so far. Normally, the maximum value of m is available to the receiver designer because it can be obtained from field measurements or statistical channel models that describe the environment where the transmission system is expected to operate. However, the *actual* channel order for a given data frame is *unknown*, and both overestimating and, especially, underestimating the value of m lead to noticeable performance losses. Solutions to this problem are explored in Section V-B.

A. Delayed Sampling

The basic idea of delayed sampling is to incorporate *future* observations when sampling the particles. Specifically, if we consider a fixed lag d, the sampling of $s_t^{(i)}$ is delayed until $y_{0:t+d}$ are available (hence the name of the technique), and the weights are also computed according to the whole set of observations.

The delayed SIS (D-SIS) algorithm with a fixed lag d can be briefly outlined as

$$s_t^{(i)} \sim q\left[s_t \,\middle|\, s_{0:t-1}^{(i)}, y_{0:t+d}\right]$$
$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p\left[y_{t:t+d} \,\middle|\, s_{0:t}^{(i)}, y_{0:t-1}\right]}{q\left[s_t \,\middle|\, s_{0:t-1}^{(i)}, y_{0:t+d}\right]}$$

where, compared with the standard SIS procedure of (9) and (10), particles are drawn from a delayed importance pmf that incorporates observations up to time $t+d$, whereas importance weights are computed accordingly, using the likelihood $p[y_{t:t+d} \mid s_{0:t}^{(i)}, y_{0:t-1}]$. The latter function can be written as

$$\begin{aligned} p\left[y_{t:t+d} \,\middle|\, s_{0:t}^{(i)}, y_{0:t-1}\right] &= \sum_{\tilde{s}_{t+1:t+d}\in\{\pm 1\}^d} p\left[y_{t:t+d}, \tilde{s}_{t+1:t+d} \,\middle|\, s_{0:t}^{(i)}, y_{0:t-1}\right] \\ &\propto \sum_{\tilde{s}_{t+1:t+d}\in\{\pm 1\}^d} p\left[y_{t:t+d} \,\middle|\, \tilde{s}_{t+1:t+d}, s_{0:t}^{(i)}, y_{0:t-1}\right] \end{aligned}$$

where the proportionality comes from the assumption that the transmitted symbols are i.i.d. uniform random variables. Using the same methodology as in Appendix A, it is straightforward to obtain the closed-form expression

$$\begin{aligned} p\left[y_{t:t+d} \,\middle|\, s_{0:t}^{(i)}, y_{0:t-1}\right] \propto & \sum_{\tilde{s}_{t+1:t+d}\in\{\pm 1\}^d} \frac{\left|\mathbf{C}_{t+d}^{(i)}\right|^{1/2}}{\left|\mathbf{C}_{t-1}^{(i)}\right|^{1/2}} \\ & \times \exp\left\{-\frac{1}{2}\left[\frac{\mathbf{y}_t^\top \mathbf{y}_t}{\sigma^2} + \bar{\mathbf{h}}_{t-1}^{(i)\top} \mathbf{C}_{t-1}^{-1\,(i)} \bar{\mathbf{h}}_{t-1}^{(i)} \right.\right. \\ & - \left(\frac{\mathbf{S}_t^{(i)}\mathbf{y}_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1\,(i)}\bar{\mathbf{h}}_{t-1}^{(i)}\right)^\top \\ & \left.\left. \times \mathbf{C}_{t+d}^{(i)}\left(\frac{\mathbf{S}_t^{(i)}\mathbf{y}_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1\,(i)}\bar{\mathbf{h}}_{t-1}^{(i)}\right)\right]\right\} \end{aligned} \quad (15)$$

where $\mathbf{y}_t = [y_t, y_{t+1}, \ldots, y_{t+d}]^\top$ is the $(d+1)\times 1$ observation vector, the $m \times (d+1)$ symbol matrix $\mathbf{S}_t^{(i)}$ is built as

$$\mathbf{S}_t = \left[\tilde{\mathbf{s}}_t^{(i)} \quad \tilde{\mathbf{s}}_{t+1}^{(i)} \quad \cdots \quad \tilde{\mathbf{s}}_{t+d}^{(i)}\right]$$

with $\tilde{\mathbf{s}}_t^{(i)} = \mathbf{s}_t^{(i)} = [s_{t-m+1}^{(i)}, \ldots, s_{t-1}^{(i)}, s_t^{(i)}]^\top$, the remaining column vectors are obtained by shifting

$$\tilde{\mathbf{s}}_k^{(i)} = \mathbf{T}\tilde{\mathbf{s}}_{k-1}^{(i)} + [0, \ldots, 0, \tilde{s}_k]^\top$$

and the matrix $\mathbf{C}_{t+d}^{-1\,(i)}$ is a covariance matrix that can be computed recursively as

$$\mathbf{C}_{t+d}^{-1\,(i)} = \mathbf{C}_{t-1}^{-1\,(i)} + \frac{1}{\sigma^2}\mathbf{S}_t^{(i)}\mathbf{S}_t^{(i)\top}.$$

Similarly to the standard SIS algorithm, the optimal delayed importance pmf is proportional to the likelihood; hence, it can be written as

$$q\left[s_t \,\middle|\, s_{0:t-1}^{(i)}, y_{0:t+d}\right] = \frac{p\left[y_{t:t+d} \,\middle|\, s_t, s_{0:t-1}^{(i)}, y_{0:t-1}\right]}{\sum_{a\in\{\pm 1\}} p\left[y_{t:t+d} \,\middle|\, s_t = a, s_{0:t-1}^{(i)}, y_{0:t-1}\right]} \quad (16)$$

and easily computed by substituting (15) into (16). Using this choice of trial pmf, the weight update equation of the D-SIS algorithm reduces to

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \sum_{a\in\{\pm 1\}} p\left[y_{t:t+d} \,\middle|\, s_t = a, s_{0:t-1}^{(i)}, y_{0:t-1}\right].$$

The delayed sequential importance sampling with resampling (D-SIR) algorithm for blind MAP equalization is similar to the

TABLE II
I-SIR ALGORITHM FOR BLIND MAP EQUALIZATION

$T+1 \equiv$ frame size, $N \equiv$ number of trajectories, $m \in \mathcal{M} = \{1, \ldots, m_{max}\}$.

Initialization:	$\forall i \in \{1,...,N\}, \forall m \in \mathcal{M}$ $\mathbf{C}_{-1}^{(i,m)} := \mathbf{C}_{-1}$ $\overline{\mathbf{h}}_{-1}^{(i,m)} := \overline{\mathbf{h}}_{-1}$.
For $t = 0, 1, ..., T$:	
For $i = 1, ..., N$:	
Posterior channel-order pmf:	$\forall m \in \mathcal{M}$ $p[m \vert s_{0:t-1}^{(i)}, y_{0:t-1}] \propto p[y_{t-1} \vert s_{0:t-1}^{(i)}, m, y_{0:t-2}] p[m \vert s_{0:t-2}^{(i)}, y_{0:t-2}]$
Importance sampling:	$s_t^{(i)} \sim q[s_t \vert s_{0:t-1}^{(i)}, y_{0:t}]$ $\mathbf{s}_t^{(i)} = [s_{t-m+1}^{(i)}, \ldots, s_t^{(i)}]^\top$
Channel update:	$\forall m \in \mathcal{M}$ $\mathbf{C}_t^{(i,m)} = \left(\frac{\mathbf{s}_t^{(i)} \mathbf{s}_t^{(i)\top}}{\sigma^2} + \mathbf{C}_{t-1}^{(i,m)^{-1}} \right)^{-1}$ $\overline{\mathbf{h}}_t^{(i,m)} = \mathbf{C}_t^{(i,m)} \left(\frac{\mathbf{s}_t^{(i)} y_t}{\sigma^2} + \mathbf{C}_{t-1}^{(i,m)^{-1}} \overline{\mathbf{h}}_{t-1}^{(i,m)} \right)$
Weight update:	$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{\sum_{m \in \mathcal{M}} p[y_t \vert s_{0:t}^{(i)}, y_{0:t-1}] p[m \vert s_{0:t-1}^{(i)}, y_{0:t-1}]}{q_t[s_t^{(i)} \vert s_{0:t-1}^{(i)}, y_{0:t}]}$
Weight normalization:	$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}, \forall i$
Resampling:	(the same as in the SIR and D-SIR algorithms)
MAP estimation: using equations (12) or (13)	

basic SIR technique, as indicated in Table I. Its superior performance is achieved at the expense of increased computational complexity, which grows exponentially with the sampling delay d.

B. Unknown Channel Order

1) Marginalization: Common statistical channel models available to transmission system designers (e.g., power-delay profiles [21], [22]) can provide probabilistic information regarding the maximum CIR duration that can be expected with a non-negligible probability or the mean value and the power (second-order moment) of the channel coefficients. With this information at hand, it is not difficult to specify a prior pmf for the channel order of the form

$$p[m], \quad m \in \mathcal{M} = \{1, 2, \ldots, m_{\max}\} \tag{17}$$

where $m_{\max}$ is the maximum CIR duration that is likely to be observed under normal transmission conditions. Starting from this simple *a priori* statistical description, the uncertainty in the knowledge of the channel order can be analytically handled within the Bayesian framework. Specifically, the parameter m can be marginalized, and the posterior pmf of the data up to time t be recursively decomposed as

$$\begin{aligned} p[s_{0:t} \mid y_{0:t}] &\propto \sum_{m \in \mathcal{M}} p[y_t \mid s_{0:t}, m, y_{0:t-1}] p[s_{0:t-1}, m \mid y_{0:t-1}] \\ &= p[s_{0:t-1} \mid y_{0:t-1}] \sum_{m \in \mathcal{M}} p[y_t \mid s_{0:t}, m, y_{0:t-1}] \\ &\quad \times p[m \mid s_{0:t-1}, y_{0:t-1}]. \end{aligned} \tag{18}$$

The *a posteriori* pmf of the channel order, which appears on the right-hand side of (18), also admits a recursive decomposition in terms of the likelihood

$$\begin{aligned} &p[m \mid s_{0:t-1}, y_{0:t-1}] \\ &\qquad \propto p[y_{t-1} \mid s_{0:t-1}, m, y_{0:t-2}] p[m \mid s_{0:t-2}, y_{0:t-2}] \end{aligned} \tag{19}$$

that enables its sequential update. Equations (18) and (19) are the basis for the integrated-order sequential importance sampling with resampling (I-SIR) algorithm for blind MAP equalization, which is summarized in Table II.

The optimal choice of the importance pmf is proportional to the expected value of the likelihood with respect to the posterior distribution of the channel order, i.e., as in (20), shown at the bottom of the page, and as a consequence, the weight update equation in Table II reduces to

$$\begin{aligned} \tilde{w}_t^{(i)} = w_{t-1}^{(i)} \sum_{a \in \{\pm 1\}} \sum_{m \in \mathcal{M}} p \left[y_t \;\middle|\; s_t = a, s_{0:t-1}^{(i)}, m, y_{0:t-1} \right] \\ \times p \left[m \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t-1} \right]. \end{aligned}$$

The efficiency of the I-SIR algorithm can be further improved by applying the technique of delayed sampling, as described in Section V-A. The complexity of the resulting algorithm may turn out prohibitive, however, since drawing and weighting a single particle would require $\mathcal{O}(|\mathcal{M}| 2^d)$ operations, where d is the smoothing lag.

2) Adaptive Estimation: An alternative to marginalization is to estimate the CIR duration jointly with the transmitted symbols. The straightforward approach is to include m in the system

$$q \left[s_t \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t} \right] = \frac{\sum_{m \in \mathcal{M}} p \left[y_t \;\middle|\; s_t, s_{0:t-1}^{(i)}, m, y_{0:t-1} \right] p \left[m \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t-1} \right]}{\sum_{a \in \{\pm 1\}} \sum_{m \in \mathcal{M}} p \left[y_t \;\middle|\; s_t = a, s_{0:t-1}^{(i)}, m, y_{0:t-1} \right] p \left[m \;\middle|\; s_{0:t-1}^{(i)}, y_{0:t-1} \right]} \tag{20}$$

TABLE III
A-SIR ALGORITHM FOR BLIND MAP EQUALIZATION

$T+1$ ≡frame size, $N \equiv$ number of trajectories, $m \in \mathcal{M} = \{1, ..., m_{max}\}$.	
Initialization:	Draw $m^{(i)} \sim p_m[m]$, Set $\mathbf{C}_{-1}^{(i)} := \mathbf{C}_{-1}$ and $\overline{\mathbf{h}}_{-1}^{(i)} := \overline{\mathbf{h}}_{-1}$, $\forall i$.
For $t = 0, 1, ..., T$: Let $j = 1$ **For** $i = 1, ..., M$: **For** $l = -1 : 1 : +1$: If $m^{(i)} + l \in \mathcal{M}$ then	
Importance sampling:	$s_t^{(j)} \sim q[s_t \| s_{0:t-1}^{(i)}, m^{(i)} + l, y_{0:t}]$ $\mathbf{s}_t^{(j)} = [s_{t-m^{(i)}-l+1}^{(i)}, \ldots, s_t^{(j)}]^\top$
Channel update:	Compute the Gaussian posterior $p[\mathbf{h} \| \mathbf{s}_{0:t}^{(j)}, m^{(i)} + l, y_{0:t}] = \mathcal{N}\left(\overline{\mathbf{h}}_t^{(j)}, \mathbf{C}_t^{(j)}\right)$ from the Gaussian prior $p_{t-1,m^{(i)}+l}[\mathbf{h}]$
Weight update:	$\tilde{w}_t^{(j)} = w_{t-1}^{(i)} \frac{p[y_t \| s_{0:t}^{(i)}, m^{(i)}+l, y_{0:t-1}]}{q_t[s_t^{(i)} \| s_{0:t-1}^{(i)}, m^{(i)}+l, y_{0:t}]}$
Increase the particle label, $j = j + 1$.	
Weight normalization:	$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}$, $\forall i$
Resampling:	(1) for $i = 1, ..., N$, let $\tilde{s}_{0:t}^{(i)} = s_{0:t}^{(j)}$ with probability $w_t^{(j)}$; (2) let $s_{0:t}^{(i)} = \tilde{s}_{0:t}^{(i)}$, $i = 1, ..., N$; and (3) let $w_t^{(i)} = \frac{1}{N}$, $i = 1, ..., N$.
MAP estimation: using equations (12) or (13)	

state by adding the constant relationship $m_t = m_{t-1}$ to model (3) and (4) and then estimate a sequence $\hat{m} = \hat{m}_{0:t}$ (together with the data) using the standard SIR method. Unfortunately, including a random *fixed* parameter as part of the system state in this way generally prevents the convergence of standard sequential MC algorithms [12], primarily due to impoverishment of the particle population.

Therefore, we consider a slightly different approach here. In particular, while using the standard recursion

$$p[s_{0:t}, m] \propto p[y_t \mid s_{0:t}, m, y_{0:t-1}] p[s_{0:t-1}, m \mid y_{0:t-1}]$$

as a basic relationship, we also allow for particles with $m = k \in \mathcal{M}$ to give birth to new trajectories with different channel orders, namely, $m = k+1$ and $m = k-1$, as long as $k \pm 1 \in \mathcal{M}$. The general proposed procedure can be outlined as follows.

- Initialization: Let $m^{(i)} \sim p[m], i = 1, \ldots, N$.
- For $t = 0, 1, \ldots$
  — For each existing particle, $(s_{0:t-1}, m)^{(i)}$, draw up to three new samples

$$\left(s_{0:t}, m^{(i)} - 1\right)^{(j)}, \quad \left(s_{0:t}, m^{(i)}\right)^{(j+1)}, \quad \text{and} \quad \left(s_{0:t}, m^{(i)} + 1\right)^{(j+2)}.$$

  (Notice that if $m^{(i)} = 1$ or $m^{(i)} = m_{\max}$ only two offsprings are possible.)
  — Compute the importance weights for all (up to $3N$) resulting trajectories.
  — Resample in order to keep only N particles.
- Detect the data and estimate the channel order.

The principal advantage of this approach with respect to the standard SIR method with a fixed random parameter is that we allow for dynamic *rejuvenation* of the particle population of the CIR length m.

A more detailed description of the resulting adaptive-order sequential importance sampling with resampling (A-SIR) algorithm is given in Table III. The main difficulty in the implementation of the method is the need to specify new Gaussian channel prior pdfs for a transition from channel order m to order $m+1$ and for the transition from m to $m-1$, given the time $t-1$ posterior for order m. Apparently, the computation of the true Gaussian densities $p[\mathbf{h} \mid s_{0:t-1}^{(i)}, m+1, y_{0:t-1}]$ and $p[\mathbf{h} \mid s_{0:t-1}^{(i)}, m-1, y_{0:t-1}]$ is analytically feasible, but it requires the processing of the whole trajectory and the whole set of observations; therefore, it is not practical. Instead, we suggest building a new channel mean by eliminating the first element in vector $\bar{\mathbf{h}}_{t-1}^{(i)}$ if the transition is from order m to order $m-1$ while padding $\bar{\mathbf{h}}_{t-1}^{(i)}$ with an extra zero element at the beginning of the vector if the transition is from m to $m+1$. Correspondingly, the new covariance matrix is built by suppressing the first row and column in matrix $\mathbf{C}_{t-1}^{(i)}$, when moving from m to $m-1$, and padding $\mathbf{C}_{t-1}^{(i)}$ with the first row and column of the $(m+1) \times (m+1)$ identity matrix when the transition is from m to $m+1$.

VI. COMPUTER SIMULATIONS

A. Simulation Setup

In order to assess the performance of the proposed MAP equalizers, we have carried out several computer simulation experiments using the discrete-time model (3) and (4) with a frame size $T+1 = 60$ and a null signal between frames ($s_{t<0} = s_{t>T} = 0$). All numerical results are presented in terms of the average bit error rate (BER) obtained for a collection of 170 random channel realizations drawn from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{0}$ is the $m \times 1$ zero-mean vector, and $\mathbf{C}$ is an $m \times m$ diagonal covariance matrix with nonzero entries conforming to the ratios $10\log_{10}(\sigma_{00}^2/\sigma_{11}^2) = 0, 10\log_{10}(\sigma_{00}^2/\sigma_{22}^2) = -6$, and $10\log_{10}(\sigma_{00}^2/\sigma_{33}^2) = -9$, where σ_{ii}^2 is the variance of the ith channel coefficient and $\sigma_{00}^2 = 0.2$. For each channel realization, the BER is numerically estimated for several values of the SNR, which is defined as

$$\text{SNR} = 10\log_{10}\left(\frac{\mathbf{h}^\top \mathbf{h}}{\sigma^2}\right).$$

We have considered seven types of equalizers in the experiments:

1) *The maximum likelihood equalizer (MLE) implemented via the Viterbi algorithm with known CIR [1]:* This is the optimal detector that yields a lower bound for the BER of the proposed schemes.
2) *A blind equalizer based on the PSP method:* This is a generalization of the Viterbi algorithm, where N_s *survivor* paths are preserved in each state of the trellis. For each path, an estimate of the CIR is computed using the recursive least squares (RLS) algorithm [6].
3) *The SIR blind equalizer [unknown CIR, known channel order, MAP sequence detection according to (12)] described in Table I:* The optimal importance pmf of (14) and resampling steps are taken whenever $N_{\text{eff}} < 0.2N$.
4) *The D-SIR blind equalizer [unknown CIR, known channel order, MAP sequence detection according to (12)] also outlined in Table I:* Here, we have the optimal importance pmf of (16) and threshold for resampling $N_{\text{eff}} = 0.2N$.
5) *The I-SIR blind equalizer [unknown CIR, unknown channel order, MAP sequence detection according to (12)] summarized in Table II:* Here, we have the optimal importance pmf of (20), a uniform channel-order prior pmf in the set $\mathcal{M} = \{1, 2, 3, 4\}$, and resampling threshold $N_{\text{eff}} = 0.2N$.
6) *The A-SIR blind MAP equalizer [unknown CIR, unknown channel order, MAP sequence detection according to (12)] shown in Table III:* The channel order is contained in the set $\mathcal{M} = \{1, 2, 3, 4\}$, and the optimal importance pmf (14) is applied according to the specific CIR duration of each particle.
7) *The blind MAP sequence detector (unknown CIR, known channel order):* This is based on the Gibbs sampler proposed in [23].

All SIR-based algorithms are initialized with *a priori* channel statistics $\bar{\mathbf{h}}_{-1} = \mathbf{0}$ and $\mathbf{C}_{-1} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of adequate dimensions.

B. Numerical Results

We start with a simple scenario consisting of *short* channels with known order $m = 2$. Fig. 1 shows the BER attained by the MLE, SIR, D-SIR, and Gibbs-sampler equalizers in the SNR region between 0 and 12 dB. The SIR and D-SIR algorithms are evaluated with $N = 100$ and $N = 300$ particles. The BER of the Gibbs receiver is only shown for the case of $N = 300$ particles and a burn-in period of 100 iterations [23]. The lag in the D-SIR algorithm is $d = 3$. Both the SIR and D-SIR equalizers attain a low BER, which is close to the bound given by the MLE equalizer with known CIR. As expected, the higher complexity of the D-SIR technique leads to a better performance with respect to the basic SIR algorithm. It is also observed that by increasing the number of particles from 100 to 300, the performance of the SIR and D-SIR detectors comes significantly closer to the MLE (less than 1 dB for the D-SIR equalizer at BER $= 10^{-3}$). Finally, it is interesting to note how the Gibbs equalizer suffers from a *BER floor* effect that can be clearly seen for SNR ≥ 8 dB (as previously reported in [11]) and, thus, provides a much poorer performance than the SIR-based receivers.

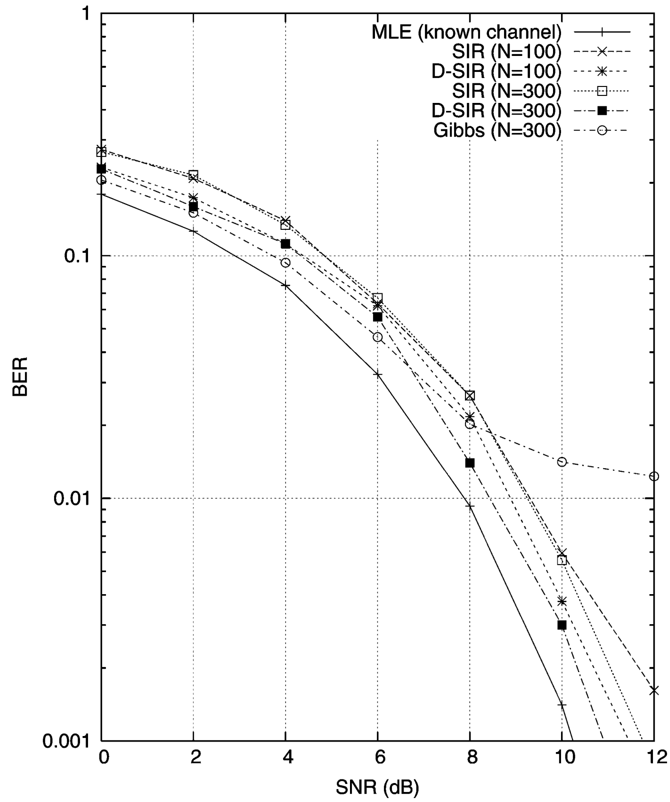

Fig. 1. BER versus SNR averaged over 170 random channel realizations of order $m = 2$. There is a burn-in period of 100 initial iterations for the Gibbs sampler.

We have also explored the effect of channel order misadjustment on the SIR equalizer. In particular, we have considered a collection of randomly selected channels with length $m = 3$ and applied the SIR equalization algorithm under the assumptions $m = 2, m = 3$, and $m = 4$. Results are plotted in Fig. 2. We clearly see how underestimating the channel order [curve labeled SIR (*estimated* $m = 2$)] leads to a large performance degradation, whereas the BER curves obtained under the assumptions $m = 3$ (the true channel order) and $m = 4$ (overestimated channel order) are very similar, with a small improvement in the case of perfect order estimation. In the figure, we can also see the BER curves obtained for the MLE with perfect channel knowledge (labeled MLE $(m = 3)$) and the MLE working with a *truncated* channel of only $m = 2$ coefficients. The performance degradation in the latter case is particularly severe. Similar results are obtained for the Gibbs-sampler equalizer, as depicted in Fig. 3, with the added handicap of the error-floor effect already illustrated in Fig. 1.

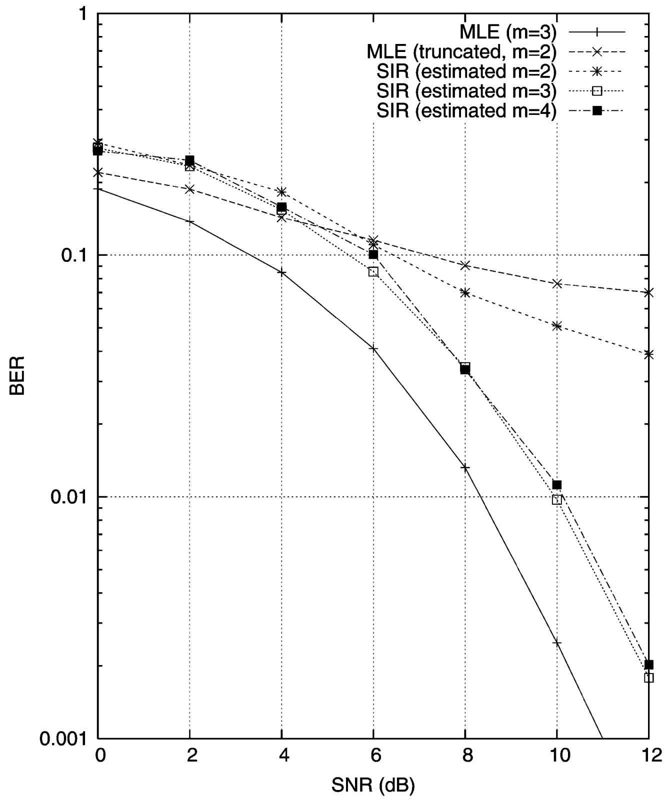

Fig. 2. BER versus SNR averaged over 170 random channel realizations of order $m = 3$. The SIR algorithm is run with different CIR duration assumptions $(m = 2, 3, 4)$ and $N = 300$ particles.

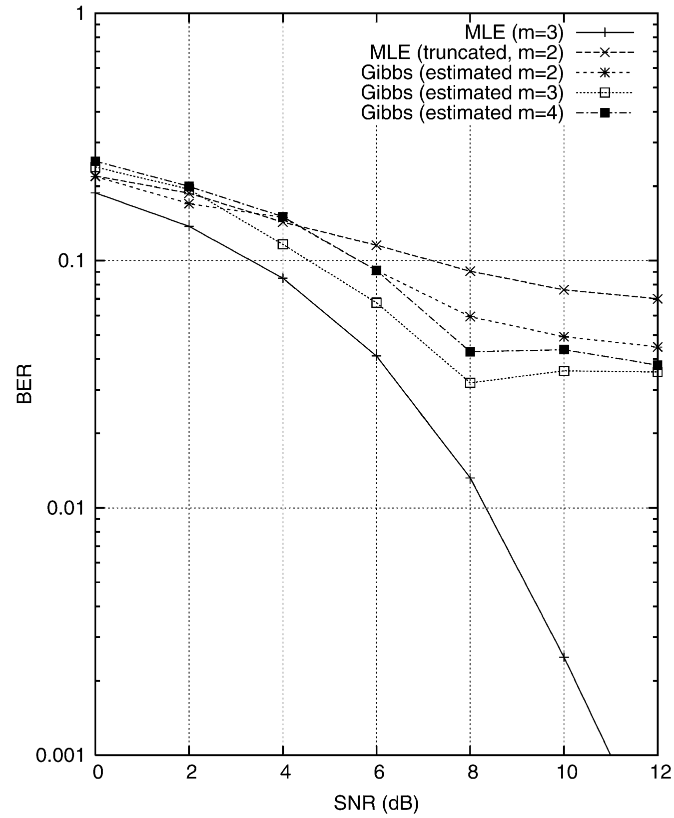

Fig. 3. BER versus SNR averaged over 170 random channel realizations of order $m = 3$. The Gibbs-sampler algorithm is run with different CIR duration assumptions $(m = 2, 3, 4)$, 100 burn-in iterations, and $N = 300$ effective particles.

Finally, we have conducted a simulation experiment to study the performance of the I-SIR and A-SIR techniques, which are specifically designed to cope with channel order uncertainty. We randomly generated a collection of 170 channels drawn from the Gaussian prior specified at the beginning of Section VI-A. For each sample channel, the order m was chosen randomly according to the uniform prior (17) with a maximum value $m_{\max} = 4$. In this scenario, we have compared the MLE with perfect CIR knowledge, the PSP equalizer with $N_s = 19$ survivors per state (which yields roughly the same computational complexity as the SIR-like algorithms with $N = 300$ particles) and overestimated CIR duration (i.e., assuming $m = m_{\max} = 4$ for all channels), the SIR equalizer with overestimated channel, the I-SIR receiver, which marginalizes m using the prior (17), and the A-SIR algorithm, which estimates m jointly with the data. The obtained results are shown in Fig. 4. The I-SIR algorithm performs very close to the optimal MLE, with a loss of only 1 dB for BER $= 10^{-3}$ and practically the same curve slope. The PSP method also yields very good results in the low and medium SNR region but presents a noticeable performance degradation (increase of the BER curve slope) in the higher SNRs. The performance advantage of the I-SIR approach with respect to the plain SIR equalizer with overestimated channel order is apparent. The A-SIR method exhibits the worst BER out of the three sequential MC approaches. Although it attains practically the same performance as the SIR equalizer in the low SNR region, its BER degrades clearly for medium to high SNR values. Nevertheless, it has a sound behavior, with a constant decreasing slope and no error floor observed for the considered range of SNR values.

VII. Conclusion

We have introduced a novel blind equalization method for frequency-selective channels based on a Bayesian formulation of the problem and the SIS methodology. An algorithm that sequentially builds an MC representation of the symbol posterior pmf given the available observations has been derived. One of the main features of the proposed blind equalizer is that an explicit estimation of the CIR is not carried out. Instead, the *a posteriori* channel distribution is recursively computed for each data trajectory in the smoother. This approach requires that the channel order be *a priori* known, which may not be realistic in many practical scenarios. To avoid this limitation, we have also proposed extensions of the basic equalizer that explicitly deal with channel order uncertainty. Computer simulation results are presented that confirm the validity of the proposed techniques.

Appendix
Derivation of the Likelihood Function

The likelihood in (6) can be written in terms of $\mathbf{h}$ as

$$p[y_t \,|\, s_{0:t}, y_{0:t-1}] = \int_{\mathbb{R}^m} p[y_t \,|\, \mathbf{h}, \mathbf{s}_t] p[\mathbf{h} \,|\, s_{0:t-1}, y_{0:t-1}] \, d\mathbf{h} \qquad (21)$$

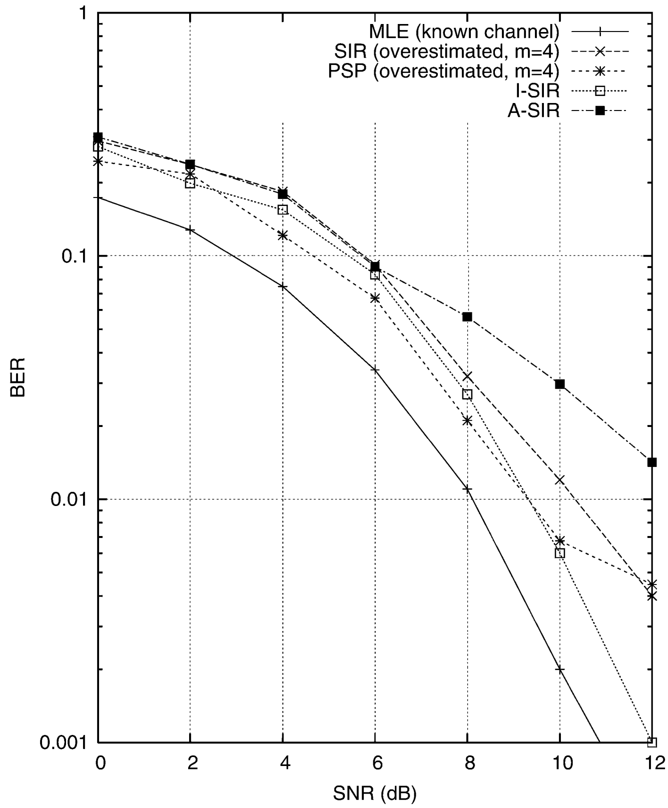

Fig. 4. BER versus SNR averaged over 170 random channel realizations of random order. For each sample channel, m is randomly drawn from the set $\{1, 2, 3, 4\}$ according to a uniform pmf. The number of particles in the MC algorithms is $N = 300$. The adaptive PSP holds $N_s = 19$ survivor paths per trellis state and assumes a length $m = 4$ CIR.

where $p[y_t \,|\, \mathbf{h}, \mathbf{s}_t] = \mathcal{N}(\mathbf{s}_t^\top \mathbf{h}, \sigma^2)$ is a Gaussian density. The posterior pdf of the channel at time t is also proportional to the integrand in (21)

$$p[\mathbf{h} \,|\, s_{0:t}, y_{0:t}] \propto p[y_t \,|\, \mathbf{h}, \mathbf{s}_t] p[\mathbf{h} \,|\, s_{0:t-1}, y_{0:t-1}]. \tag{22}$$

This property entails $p[y_t \,|\, \mathbf{h}, s_{0:t}, y_{0:t-1}] = p[y_t \,|\, \mathbf{h}, \mathbf{s}_t]$, and by iterating (22), we obtain

$$p[\mathbf{h} \,|\, s_{0:t}, y_{0:t}] \propto p[\mathbf{h}] \prod_{k=0}^{t} p[y_k \,|\, \mathbf{h}, \mathbf{s}_k]$$

where all conditional densities are Gaussian. If we additionally assume that the channel vector is *a priori* distributed according to a Gaussian model $\mathbf{h} \sim \mathcal{N}(\bar{\mathbf{h}}_{-1}, \mathbf{C}_{-1})$, where $\bar{\mathbf{h}}_{-1}$ and $\mathbf{C}_{-1}$ are the prior mean and covariance of $\mathbf{h}$, respectively, then $p[\mathbf{h} \,|\, s_{0:t}, y_{0:t}]$ is proportional to a product of Gaussian densities and, as a consequence, it is Gaussian itself.

Let $\bar{\mathbf{h}}_t$ and $\mathbf{C}_t$ denote the posterior mean and covariance of $\mathbf{h}$ given $y_{0:t}$ and $s_{0:t}$. It is possible to recursively compute $\bar{\mathbf{h}}_t$ and $\mathbf{C}_t$ from $\bar{\mathbf{h}}_{t-1}$ and $\mathbf{C}_{t-1}$. With that aim, we first expand the right-hand side of (22) to obtain

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

where $|\cdot|$ denotes the determinant of a matrix. It is not difficult to show that (23) can be rewritten as

$$\begin{aligned} &p[y_t \,|\, \mathbf{h}, \mathbf{s}_t] p[\mathbf{h} \,|\, s_{0:t-1}, y_{0:t-1}] \\ &\quad = \frac{1}{(2\pi\sigma^2)^{1/2}(2\pi)^{m/2}|\mathbf{C}_{t-1}|^{1/2}} \\ &\qquad \times \exp\left\{ -\frac{1}{2} \left[(\mathbf{h} - \bar{\mathbf{h}}_t)^\top \mathbf{C}_t^{-1} (\mathbf{h} - \bar{\mathbf{h}}_t) + y_t^2 / \sigma^2 \right.\right. \\ &\qquad + \bar{\mathbf{h}}_{t-1}^\top \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} - \left(\mathbf{s}_t y_t / \sigma^2 + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right)^\top \\ &\qquad \left.\left. \times \mathbf{C}_t \left(\mathbf{s}_t y_t / \sigma^2 + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right) \right] \right\} \end{aligned} \tag{24}$$

where

$$\mathbf{C}_t^{-1} = \frac{\mathbf{s}_t \mathbf{s}_t^\top}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \tag{25}$$

$$\bar{\mathbf{h}}_t = \mathbf{C}_t \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right). \tag{26}$$

Substituting (24) into (22), we readily obtain that

$$\begin{aligned} p[\mathbf{h} \,|\, \mathbf{s}_{0:t}, y_{0:t}] &\propto \exp\left\{ -\frac{1}{2} (\mathbf{h} - \bar{\mathbf{h}}_t)^\top \mathbf{C}_t^{-1} (\mathbf{h} - \bar{\mathbf{h}}_t) \right\} \\ &\propto \mathcal{N}(\bar{\mathbf{h}}_t, \mathbf{C}_t) \end{aligned}$$

and therefore, (25) and (26) allow for the update of the posterior channel pdf from $p[\mathbf{h} \,|\, y_{0:t-1}, s_{0:t-1}]$ to $p[\mathbf{h} \,|\, y_{0:t}, s_{0:t}]$ when y_t and s_t become available. We remark that $\bar{\mathbf{h}}_t$ is the minimum mean square error (MMSE) estimate of the CIR given the observations and the data up to time t.

Finally, we solve the integral in (21). Specifically, substituting (24) into (21), and taking into account that

$$\int_{\mathbb{R}^m} \exp\left\{ -\frac{1}{2} (\mathbf{h} - \bar{\mathbf{h}}_t)^\top \mathbf{C}_t^{-1} (\mathbf{h} - \bar{\mathbf{h}}_t) \right\} d\mathbf{h} = (2\pi)^{m/2} |\mathbf{C}_t|^{1/2}$$

we arrive at the analytical expression

$$\begin{aligned} &p[y_t \,|\, s_{0:t}, y_{0:t-1}] \\ &\quad = \frac{|\mathbf{C}_t|^{1/2}}{(2\pi\sigma^2 |\mathbf{C}_{t-1}|)^{1/2}} \exp\left\{ -\frac{1}{2} \left[\frac{y_t^2}{\sigma^2} + \bar{\mathbf{h}}_{t-1}^\top \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right.\right. \\ &\qquad \left.\left. - \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right)^\top \mathbf{C}_t \left(\frac{\mathbf{s}_t y_t}{\sigma^2} + \mathbf{C}_{t-1}^{-1} \bar{\mathbf{h}}_{t-1} \right) \right] \right\}. \end{aligned} \tag{27}$$

REFERENCES

[1] J. G. Proakis, *Digital Communications*, Singapore: McGraw-Hill, 1995.
[2] S. Haykin, Ed., *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[3] G. B. Giannakis and S. D. Halford, "Blind fractionally spaced equalization of noisy FIR channels: Direct and adaptive solutions," *IEEE Trans. Signal Processing*, vol. 45, pp. 2277–2292, Sept. 1997.
[4] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. Commun.*, vol. 47, pp. 1181–1193, Aug. 1999.
[5] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000–1011, Feb./Mar./Apr. 1994.

[6] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[7] Y. Sato, "A blind sequence detection and its application to digital mobile communications," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 49–58, Jan. 1995.

[8] W. J. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Process.*, vol. 81, no. 1, pp. 3–18, Jan. 2001.

[9] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics Comput.*, vol. 10, no. 3, pp. 197–208, 2000.

[10] R. Chen, J. S. Liu, and X. Wang, "Convergence analyzes and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. Signal Processing*, vol. 50, pp. 255–270, Feb. 2002.

[11] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering ," *IEEE Signal Processing Mag.*, vol. 20, pp. 19–38, Sept. 2003.

[12] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering," *IEEE Trans. Signal Processing*, vol. 50, pp. 736–746, Mar. 2002.

[13] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *J. Amer. Stat. Assoc.*, vol. 90, no. 430, pp. 567–576, June 1995.

[14] S. Haykin, *Adaptive Filter Theory*, Third ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[15] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2079–2094, Sept. 2000.

[16] J. Zhang and P. M. Djurić, "Joint estimation and decoding of space-time trellis codes," *J. Appl. Signal Process.*, vol. 3, pp. 305–315, 2002.

[17] Z. Yang and X. Wang, "A sequential Monte Carlo blind receiver for OFDM systems in frequency-selective fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 271–280, Feb. 2002.

[18] D. Guo, X. Wang, and R. Chen, "Wavelet-based sequential Monte Carlo blind receivers in fading channels with unknown channel statistics," in *Proc. Int. Contr. Conf.*, 2002, pp. 821–825.

[19] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 24, pp. 1317–1399, 1989.

[20] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Stat. Assoc.*, vol. 9, pp. 278–288, 1994.

[21] T. S. Rapappport, *Wireless Communications*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[22] V. K. Garg and J. E. Wilkes, *Principles & Applications of GSM*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[23] X. Wang and R. Chen, "Blind turbo equalization in Gaussian and impulsive noise," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1092–1105, July 2001.

Joaquín Míguez (M'04) received the Licenciado en Informática (M.Sc.) and Doctor en Informática (Ph.D.) degrees from Universidade da Coruña, Galicia, Spain, in 1997 and 2000, respectively.

Late in 2000, he joined the Departamento de Electrónica e Sistemas, Universidade da Coruña, where he became an associate professor in July 2003. From April 2001 through December 2001, he was a visiting scholar with the Department of Electrical and Computer Engineering, the State University of New York at Stony Brook. His research interests are in the field of statistical signal processing, with emphasis in the topics of Bayesian analysis, sequential Monte Carlo methods, adaptive filtering, stochastic optimization, and their applications to wireless communications, smart antenna systems, target tracking, and vehicle positioning and navigation.

Petar M. Djurić (SM'99) received the B.S. and M.S. degrees in electrical engineering from the University of Belgrade, Belgrade, Yugoslavia, in 1981 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Rhode Island, Kingston, in 1990.

From 1981 to 1986, he was Research Associate with the Institute of Nuclear Sciences, Vinča, Belgrade. Since 1990, he has been with the State University of New York at Stony Brook, where he is Professor with the Department of Electrical and Computer Engineering. He works in the area of statistical signal processing, and his primary interests are in the theory of modeling, detection, estimation, and time series analysis and its application to a wide variety of disciplines including wireless communications and bio-medicine. He is an Editorial Board member of *Digital Signal Processing, the EURASIP Journal on Applied Signal Processing*, and the *EURASIP Journal on Wireless Communications and Networking*.

Prof. Djurić has served on numerous conference/workshop technical committees for the IEEE and SPIE and has been invited to lecture at universities in the United States and overseas. He is the Area Editor of Special Issues of the IEEE SIGNAL PROCESSING MAGAZINE, Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and Chair of the IEEE Technical Committee on Signal Processing—Theory and Methods. He is a Member of the American Statistical Association and the International Society for Bayesian Analysis.