Wavelet-Based Sequential Monte Carlo Blind Receivers in Fading Channels With Unknown Channel Statistics

Dong Guo, Xiaodong Wang, *Member, IEEE*, and Rong Chen

***Abstract*—Recently, an adaptive Bayesian receiver for blind detection in flat-fading channels was developed by the present authors, based on the sequential Monte Carlo methodology. That work is built on a parametric modeling of the fading process in the form of a state-space model and assumes the knowledge of the second-order statistics of the fading channel. In this paper, we develop a nonparametric approach to the problem of blind detection in fading channels, without assuming any knowledge of the channel statistics. The basic idea is to decompose the fading process using a wavelet basis and to use the sequential Monte Carlo technique to track both the wavelet coefficients and the transmitted symbols. A novel resampling-based wavelet shrinkage technique is proposed to dynamically choose the number of wavelet coefficients to best fit the fading process. Under such a framework, blind detectors for both flat-fading channels and frequency-selective fading channels are developed. Simulation results are provided to demonstrate the excellent performance of the proposed blind adaptive receivers.**

***Index Terms*—Adaptive shrinkage, fading channel, resampling, sequential Monte Carlo, wavelet.**

I. Introduction

SIGNAL detection in fading channels has been a key problem in communications, and an array of methodologies have been developed to tackle this problem. Specifically, the optimal detector for flat-fading channels with known channel statistics are studied in [19] and [26], which has a prohibitively high complexity. Suboptimal receivers in flat-fading channels employ a two-stage structure, with a channel estimation stage followed by a sequence detection stage. Channel estimation is typically implemented by a Kalman filter or a linear predictor and is facilitated by per-survivor processing [35], decision-feedback [25], pilot symbols [23], or a combination of the above [21]. Other suboptimal receivers for flat-fading channels include the method based on a combination of hidden Markov model and Kalman filtering [8] and the approach based on the expectation–maximization (EM) algorithm [17].

Manuscript received July 17, 2002; revised April 17, 2003. This work was supported in part by the U.S. National Science Foundation under Grants CCR-0225721, CCR-0225826, and DMS-0225692. The associate editor coordinating the review of this paper and approving it for publication was Prof. Nicholas D Sidiropoulos.

D. Guo and X. Wang are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: wangx@ee.columbia.edu).

R. Chen is with the Department of Information and Decision Science, University of Illinois at Chicago, Chicago, IL 60607 USA.

Digital Object Identifier 10.1109/TSP.2003.819990

Sequence detection in frequency-selective fading channels has also received considerable attention recently. In [9] and [33], several MLSE receiver structures are developed that are based on the known second-order statistics of the fading process. When the fading statistics are unknown, they are usually estimated from the data in a training-assisted mode or decision-directed mode [11], [34]. Furthermore, symbol-by-symbol maximum *a posteriori* (MAP) schemes for equalizing time-varying fading channels have also been studied [2], where channel estimation is facilitated by some *ad hoc* Kalman-type nonlinear estimators. Another approach to equalization of frequency-selective fading channels is based on modeling the time-varying channel impulse response function by a superposition of deterministic time-varying basis functions (e.g., complex exponentials) with time-invariant coefficients [18].

Recently, in [5], a blind adaptive Bayesian receiver is developed for flat-fading channels. It is based on the powerful sequential Monte Carlo (SMC) technique for numerical Bayesian computation; for SMC methods and its applications in diverse fields, see [1], [3], [4], [6], [12], [14]–[16], [20], [22], [24], [31], and references therein. The blind adaptive Bayesian receiver achieves near-optimum performance without the use of any training/pilot symbols or decision feedback. The technique in [5] assumes that the fading channel process follows a linear dynamic model (i.e., ARMA model) and that the model parameters are known to the receiver. However, some practical fading processes exhibit spectral characteristics that require very-high order ARMA model to fit. Moreover, in some applications, the channel fading statistics may not be known to the receiver at all. Hence, in this paper, we address the problem of blind adaptive detection in fading channels with unknown channel statistics.

Our approach is to decompose the fading process using a wavelet basis and then use the SMC technique to estimate both the wavelet coefficients and the data symbols. Wavelet-based signal processing enjoys a very strong optimality property for general inverse problems in that their use can achieve accurate and parsimonious representation of the signal of interest. Some recent works have addressed the use of wavelet to model fading channels [27], [28]. In these methods, the number of wavelet basis is fixed *a priori*, and the wavelet coefficients are obtained by using training symbols and standard adaptive algorithms (e.g., LMS, RLS). Our approach is blind in nature, and moreover, the number of wavelet coefficients are dynamically chosen during the SMC procedure via resampling. Wavelet-based SMC receivers for both flat-fading and

frequency-selective fading channels are developed. Note that the proposed SMC algorithms will work under any orthogonal decomposition. In this paper, we mainly focus on the wavelet decomposition because it has a better time-frequency property, as suggested in [10], [27], and [28], than the Fourier transform, which means that it usually needs less number of coefficients than the Fourier transform. Moreover, the low frequency of the fading process varies slowly with time and, hence, enables the SMC algorithms to track their dynamics.

The rest of this paper is organized as follows. The wavelet representations of fading channels are discussed in Section II. The SMC blind receivers for flat-fading and frequency-selective fading channels are developed in Sections III and IV, respectively. Simulation results are provided in Section V. Section VI contains the conclusions.

II. Wavelet Modelling of Fading Channels

A. Flat-Fading Channel Model

Consider a discrete-time baseband communication system signaling through a flat-fading channel with additive white Gaussian noise. The transmitted data symbols $\{s_n\}$ take values from a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\mathcal{A}|}\}$. The input–output relationship is given by

$$y_n = s_n \alpha_n + \nu_n, \quad n = 1, 2, \ldots \tag{1}$$

where y_n, α_n, and ν_N are, respectively, the received signal, the fading coefficient, and the noise sample at time n. It is assumed that the processes $\{\alpha_n\}$, $\{s_n\}$, and $\{\nu_n\}$ are mutually independent and that ν_n is a complex Gaussian distribution

$$\nu_n \sim \mathcal{N}_c(0, \sigma^2). \tag{2}$$

The fading process is assumed to be Rayleigh, that is, $\{\alpha_n\}$ is a zero-mean complex Gaussian process with a Jakes' autocorrelation function given by [32]

$$E\left\{\alpha_m \alpha^*_{m+n}\right\} = J_0(2\pi f_d T n) \tag{3}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and zeroth order, f_d is the maximum Doppler shift, and T is the symbol interval. Note that in [5], it is assumed that the fading process $\{\alpha_n\}$ follows an ARMA model. However, for practical fading processes, e.g., Jakes' fading processes, very-high order models are needed to fit the fading spectrum given by (3). Hence, in this paper, we drop such a model assumption and treat the general fading processes via a nonparametric approach using the wavelet decomposition.

B. Wavelet Representation of Fading Processes

Suppose a discrete-time signal $\mathbf{c}^0$ has a finite length K_0, i.e., $\mathbf{c}^0 = [c_1^0, \ldots, c_{K_0}^0]^T$. In Mallat's algorithm of multiresolution analysis, a one-stage FIR filter, say lowpass filter, and downsampling by a factor of two can be given in a matrix form [10] as (4), shown at the bottom of the page, where $2N$ is the length of lowpass filter $\mathcal{H}$, and the length of $\mathbf{c}^1$ is $K_1 \triangleq K_0 + 2N - 1$. The impulse response of the lowpass filter is given by

$$h_n = \frac{1}{\sqrt{2}} \int \phi\left(\frac{x}{2}\right) \phi(x - n) \mathrm{d}x \tag{5}$$

where ϕ is the generator of the multiresolution wavelet basis. Specifically, in $\boldsymbol{H}^0$, we adopt an commonly used periodic method to mitigate edge effects [10]. Similarly, the detailed signal of $\mathbf{c}^0$ after downsampling is given by

$$\boldsymbol{d}^1 \downarrow 2 = \boldsymbol{G}^0 \boldsymbol{c}^0 \tag{6}$$

where $\boldsymbol{G}^0$ is similarly defined as $\boldsymbol{H}^0$ in (4) with $\{h_n\}$ replaced by $\{g_n\}$, which is, in general, given by

$$g_n = (-1)^n h_{1-n}. \tag{7}$$

It is easy to show that the following orthogonality condition holds:

$$\begin{bmatrix} \boldsymbol{H}^0 \\ \boldsymbol{G}^0 \end{bmatrix} \begin{bmatrix} \boldsymbol{H}^{0T} & \boldsymbol{G}^{0T} \end{bmatrix} = \boldsymbol{I}. \tag{8}$$

Hence, we can then write the one-step wavelet decomposition of $\mathbf{c}^0$ as

$$\boldsymbol{x}^1 \triangleq \begin{bmatrix} \boldsymbol{c}^1 \downarrow 2 \\ \boldsymbol{d}^1 \downarrow 2 \end{bmatrix}_{K_1 \times 1} = \begin{bmatrix} \boldsymbol{H}^0 \\ \boldsymbol{G}^0 \end{bmatrix}_{K_1 \times K_0} \boldsymbol{c}^0 \tag{9}$$

where $\boldsymbol{x}^1$ contains the first-level wavelet coefficients. Now, apply the same decomposition to $\mathbf{c}^1 \downarrow 2$ to obtain

$$\boldsymbol{c}^2 \downarrow 2 = [\boldsymbol{H}^1]_{\frac{K_2}{2} \times \frac{K_1}{2}} \boldsymbol{c}^1 \downarrow 2 \tag{10}$$

$$\text{and} \quad \boldsymbol{d}^2 \downarrow 2 = [\boldsymbol{G}^1]_{\frac{K_2}{2} \times \frac{K_1}{2}} \boldsymbol{c}^1 \downarrow 2 \tag{11}$$

$$\underbrace{\begin{bmatrix} c_0^1 \\ c_2^1 \\ c_4^1 \\ \vdots \\ \vdots \\ c_{K_1-3}^1 \\ c_{K_1-1}^1 \end{bmatrix}}_{\mathbf{c}^1 \downarrow 2} = \underbrace{\begin{bmatrix} h_{2N-2} & h_{2N-1} & & & & & h_0 & h_1 & \ldots & h_{2N-3} \\ & & & & \ddots & & & & & \\ h_0 & h_1 & \ldots & & h_{2N-1} & & & & & \\ & & h_0 & h_1 & \ldots & h_{2N-1} & & & & \\ & & & \ddots & & \ddots & & & & \\ h_2 & h_3 & \ldots & h_{2N-1} & & & & h_0 & h_1 & \end{bmatrix}}_{\boldsymbol{H}^0}_{\frac{K_1}{2} \times K_0} \underbrace{\begin{bmatrix} c_0^0 \\ c_1^0 \\ c_2^0 \\ \vdots \\ \vdots \\ c_{K_0-2}^0 \\ c_{K_0-1}^0 \end{bmatrix}}_{\boldsymbol{c}^0} \tag{4}$$

where $K_1 = K_0 + 2N - 1$. Thus, $\boldsymbol{c}^2 \downarrow 2$ and $\boldsymbol{d}^2 \downarrow 2$ have the same length $(K_2/2)$. Denote

$$\boldsymbol{x}^2 \triangleq \begin{bmatrix} \boldsymbol{c}^2 \downarrow 2 \\ \boldsymbol{d}^2 \downarrow 2 \\ \boldsymbol{d}^1 \downarrow 2 \end{bmatrix}_{\left(K_2+\frac{K_1}{2}\right)\times 1} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{H}^1 \\ \boldsymbol{G}^1 \end{bmatrix}_{K_2\times\frac{K_1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\frac{K_1}{2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}^0 \\ \boldsymbol{G}^0 \end{bmatrix} \boldsymbol{c}^0 \tag{12}$$

where $\boldsymbol{x}^2$ contains the second-level wavelet coefficients, which is composed of the downsampled second-level signal approximation and two levels of the downsampled detail signals. Repeating the above decomposition L times and keeping all the detailed signals and the last level signal approximation, we then obtain the Lth-level wavelet coefficients

$$\boldsymbol{x}^L \triangleq \begin{bmatrix} \boldsymbol{c}^L \downarrow 2 \\ \boldsymbol{d}^L \downarrow 2 \\ \vdots \\ \boldsymbol{d}^1 \downarrow 2 \end{bmatrix}_{K\times 1} \tag{13}$$

where we have (14), shown at the bottom of the page. Since $K_i = K_{i-1} + 2N - 1$, $i = 1, \ldots L$, then the length K of the Lth-level wavelet coefficients is given by

$$K = \frac{L+2}{2}K_0 + \frac{L^2+3L}{4}(2N-1). \tag{15}$$

It then follows from (14) that

$$\boldsymbol{c}^0 = \boldsymbol{\Phi}\boldsymbol{x}^L \tag{16}$$

where the perfect reconstruction matrix is given by $\boldsymbol{\Phi} \triangleq (\mathcal{W}^T\mathcal{W})^{-1}\mathcal{W}^T$. Since $\left[\begin{smallmatrix}\mathbf{H}^i \\ \mathbf{G}^i\end{smallmatrix}\right]$, $i = 0, 1 \ldots, L-1$ is orthogonal, we have

$$\boldsymbol{\Phi} = [\boldsymbol{H}^{0T} \quad \boldsymbol{G}^{0T}] \begin{bmatrix} \boldsymbol{H}^{1T} & \boldsymbol{G}^{1T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{\frac{K_1}{2}} \end{bmatrix} \cdots \begin{bmatrix} \boldsymbol{H}^{(L-1)T} & \boldsymbol{G}^{(L-1)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{\sum_{i=0}^{L-1}\frac{K_i}{2}} \end{bmatrix}. \tag{17}$$

Now, consider the fading channel coefficients $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{K_0}]^T$, where K_0 is the block size of the signal. By applying the wavelet decomposition (14) on $\boldsymbol{\alpha}$, we get

$$\boldsymbol{x} = \mathcal{W}\boldsymbol{\alpha}. \tag{18}$$

Then, the discrete-time fading process $\boldsymbol{\alpha}$ can be expressed in terms of the wavelet coefficients $\boldsymbol{x}$ as

$$\boldsymbol{\alpha} = \boldsymbol{\Phi}\boldsymbol{x}. \tag{19}$$

Denote $\boldsymbol{\phi}_t^T$ as the tth row of $\boldsymbol{\Phi}$. Then, (19) can be written as

$$\alpha_t = \boldsymbol{\phi}_t{}^T\boldsymbol{x}, \quad t = 1, \ldots, K_0. \tag{20}$$

1) Wavelet Shrinkage: As noted in (15), the number of the wavelet coefficients K is a function of the length of the original signal K_0, the length of the lowpass (or highpass) filter N, and the number of the levels of the decomposition L. Note that the wavelet coefficient space is structured, roughly, according to the location and scale of the functional information contained in each coefficient. Only a few large coefficients explain most of the functional form in the signal, whereas the remaining majority are comparatively small and, therefore, can be discarded. This is demonstrated by the following numerical example. In this example, we choose the length of the fading process segment $\boldsymbol{\alpha}$ as $K_0 = 128$, the Daubechies filter with order $N = 2$, and the decomposition level $L = 7$. Then, the size of the wavelet coefficients $\boldsymbol{x}$ is $K = 143$. For a given fading process realization, we compute the wavelet coefficients $\boldsymbol{x}$ and then truncate it by keeping only the first κ elements. Hence, the fading process $\boldsymbol{\alpha}$ is approximated by

$$\hat{\boldsymbol{\alpha}}^\kappa = \boldsymbol{\Phi}[:, 1:\kappa]\boldsymbol{x}[1:\kappa]. \tag{21}$$

The approximation error is then

$$\varepsilon^\kappa = \frac{1}{K_0}E\left\{\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}^\kappa\|^2\right\}. \tag{22}$$

In Fig. 1, we plot ε^κ as a function of the number of wavelet coefficients for different values of normalized Doppler shift f_dT. It is seen that in general, for a fixed approximation error, the slower the fading process is, the fewer wavelet coefficients are needed to approximate the fading process. For fading processes with fading rate $f_dT \leq 0.01$, with 32 wavelet coefficients, the approximation error is below -20 dB. For a very fast fading process, e.g., $f_dT > 0.01$, more wavelet coefficients are needed to approximate it well. Similar observations are made for Daubechies filters with order higher than two. Various wavelet shrinkage methods exist for choosing the number of wavelet coefficients, such as the hard shrinkage method, the visual shrinkage method [13], and the adaptive Bayesian shrinkage method [7]. In this paper, we propose a sequential Monte Carlo method for joint adaptive wavelet shrinkage, wavelet coefficients estimation, and symbol detection.

$$\boldsymbol{x}^L = \underbrace{\begin{bmatrix} \begin{bmatrix} \boldsymbol{H}^{L-1} \\ \boldsymbol{G}^{L-1} \end{bmatrix}_{K_L\times\frac{K_{L-1}}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\sum_{i=0}^{L-2}\frac{K_i}{2}} \end{bmatrix} \cdots\cdots \begin{bmatrix} \begin{bmatrix} \boldsymbol{H}^1 \\ \boldsymbol{G}^1 \end{bmatrix}_{K_2\times\frac{K_1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\frac{K_1+K_0}{2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}^0 \\ \boldsymbol{G}^0 \end{bmatrix}}_{\mathcal{W}} \boldsymbol{c}^0. \tag{14}$$

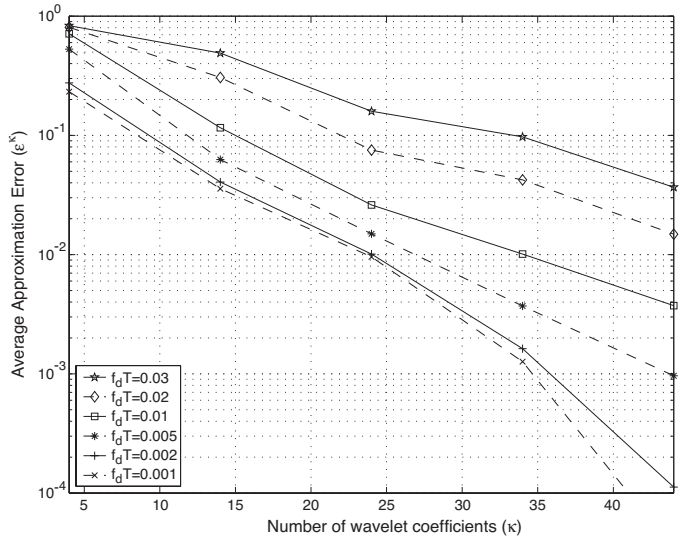

Fig. 1. Average approximation error versus the number of wavelet coefficients in the wavelet representation of the fading processes. The Daubechies filter with order 2 is used.

III. SMC Blind Receiver for Flat-Fading Channels

A. Sequential Monte Carlo Methods

Consider the following dynamic system modeled in a state-space form

$$\begin{aligned} \text{state equation } \boldsymbol{z}_t &= f_t(\boldsymbol{z}_{t-1}, \boldsymbol{u}_t) \\ \text{observation equation } \boldsymbol{y}_t &= g_t(\boldsymbol{z}_t, \boldsymbol{v}_t) \end{aligned} \tag{23}$$

where $\boldsymbol{z}_t$, $\boldsymbol{y}_t$, $\boldsymbol{u}_t$, and $\boldsymbol{v}_t$ are, respectively, the state variable, the observation, the state noise, and the observation noise at time t. Let $\boldsymbol{Z}_t = (\boldsymbol{z}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_t)$ and $\boldsymbol{Y}_t = (\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_t)$. Suppose an online inference of $\boldsymbol{Z}_t$ is of interest; that is, at current time t, we wish to make a timely estimate of a function of the state variable $\boldsymbol{Z}_t$, say $h(\boldsymbol{Z}_t)$, based on the currently available observation $\boldsymbol{Y}_t$. With the Bayes theorem, we realize that the optimal solution to this problem is $E\{h(\boldsymbol{Z}_t)|\boldsymbol{Y}_t\} = \int h(\boldsymbol{Z}_t)p(\boldsymbol{Z}_t|\boldsymbol{Y}_t)\mathrm{d}\boldsymbol{Z}_t$. In most cases, an exact evaluation of this expectation is analytically intractable because of the complexity of such a dynamic system. Sequential Monte Carlo methods, which are based on importance sampling, give us viable choices to the required estimation. The basic idea of SMC is to draw m random samples $\{\boldsymbol{Z}_t^{(j)}\}_{j=1}^m$ from some *trial* distribution $q(\boldsymbol{Z}_t|\boldsymbol{Y}_t)$ if it is difficult to draw the samples directly from the target distribution $p(\boldsymbol{Z}_t|\boldsymbol{Y}_t)$. By associating the weight

$$w_t^{(j)} = \frac{p\left(\boldsymbol{Z}_t^{(j)}\middle|\boldsymbol{Y}_t\right)}{q\left(\boldsymbol{Z}_t^{(j)}\middle|\boldsymbol{Y}_t\right)} \tag{24}$$

to the sample $\boldsymbol{Z}_t^{(j)}$, we can approximate the quantity of interest $E\{h(\boldsymbol{Z}_t)|\boldsymbol{Y}_t\}$ as

$$E_p\{h(\boldsymbol{Z}_t)|\boldsymbol{Y}_t\} \cong \frac{1}{W_t}\sum_{j=1}^{m} h\left(\boldsymbol{Z}_t^{(j)}\right) w_t^{(j)} \tag{25}$$

where $W_t \triangleq \sum_{j=1}^m w_t^{(j)}$. The pair $(\boldsymbol{Z}_t^{(j)}, w_t^{(j)})$, $j = 1, \cdots, m$ is called a *properly weighted sample* with respect to the target distribution $p(\boldsymbol{Z}_t|\boldsymbol{Y}_t)$.

To implement an online estimation of the posterior density, a set of random samples properly weighted with respect to $p(\boldsymbol{Z}_t|\boldsymbol{Y}_t)$ is needed for any time t. A Markovian structure of the state equation allows us to implement a recursive importance sampling strategy. Suppose a set of properly weighted samples $\{(\boldsymbol{Z}_{t-1}^{(j)}, w_{t-1}^{(j)})\}_{j=1}^m$ with respect to $p(\boldsymbol{Z}_{t-1}|\boldsymbol{Y}_{t-1})$ are available at time $(t-1)$. Then, a set of a properly weighted samples $\{\boldsymbol{Z}_t^{(j)}, w_t^{(j)}\}_{j=1}^m$ with respect to $p(\boldsymbol{Z}_t|\boldsymbol{Y}_t)$ at time t are given by the following procedure [4], [22], [24]. For $j = 1, \ldots, m$, do the following.

- Draw a sample $\boldsymbol{z}_t^{(j)}$ from a trial distribution $q(\boldsymbol{z}_t|\boldsymbol{Z}_{t-1}^{(j)}, \boldsymbol{Y}_t)$, and let $\boldsymbol{Z}_t^{(j)} = (\boldsymbol{Z}_{t-1}^{(j)}, \boldsymbol{z}_t^{(j)})$;
- Compute the importance weight

$$w_t^{(j)} = w_{t-1}^{(j)} \frac{p\left(\boldsymbol{Z}_t^{(j)}\middle|\boldsymbol{Y}_t\right)}{p\left(\boldsymbol{Z}_{t-1}^{(j)}\middle|\boldsymbol{Y}_{t-1}\right) q\left(\boldsymbol{z}_t^{(j)}\middle|\boldsymbol{Z}_{t-1}^{(j)}, \boldsymbol{Y}_t\right)}.$$

The algorithm is initialized by drawing a set of i.i.d. samples $\boldsymbol{z}_0^{(1)}, \ldots, \boldsymbol{z}_0^{(m)}$ from $p(\boldsymbol{z}_0|\boldsymbol{y}_0)$, where $\boldsymbol{y}_0$ represents the "null" information, and $p(\boldsymbol{z}_0|\boldsymbol{y}_0)$ corresponds to the prior distribution of $\boldsymbol{z}_0$.

B. Sequential Monte Carlo Receiver

Consider again the wavelet representation (44) of the fading process. In this subsection, we assume that the first κ wavelet coefficients are used to model the fading process. Denote $\boldsymbol{x} \leftarrow \boldsymbol{x}[1:\kappa]$ and $\boldsymbol{\phi}_t^T \leftarrow \boldsymbol{\phi}_t^T[1:\kappa]$. Then, the fading coefficient at time t is approximated by $\alpha_t \cong \boldsymbol{\phi}_t^T\boldsymbol{x}$. Hence, the flat-fading channel model (1) can be rewritten as

$$y_t = s_t\phi_t^T\boldsymbol{x} + \nu_t, \quad t = 1, \ldots, K_0. \tag{26}$$

Denote $\boldsymbol{Y}_t = (y_1, \cdots, y_t)$ and $\boldsymbol{S}_t = (s_1, \cdots, s_t)$. We apply the SMC method to the problem of online estimation of the *a posteriori* probability of the symbol s_t based on the received signals up to time t, without knowing the wavelet coefficients $\boldsymbol{x}$. That is, at time t, we need to estimate

$$p(s_t = a_i|\boldsymbol{Y}_t), \quad a_i \in \mathcal{A}. \tag{27}$$

Then, a hard maximum *a posteriori* (MAP) decision on symbol s_t is given by

$$\hat{s}_t = \arg\max_{a_i \in \mathcal{A}} p(s_t = a_i|\boldsymbol{Y}_t). \tag{28}$$

In order to implement the SMC, we need to obtain a set of Monte Carlo samples of the transmitted symbols $\{(\boldsymbol{S}_t^{(j)}, w_t^{(j)})\}_{j=1}^m$, properly weighted with respect to $p(\boldsymbol{S}_t|\boldsymbol{Y}_t)$. Then, the *a posteriori* symbol probability in (27) is approximated by

$$p(s_t = a_i|\boldsymbol{Y}_t) \cong \frac{1}{W_t}\sum_{j=1}^{m} \mathbf{1}\left(s_t^{(j)} = a_i\right) w_t^{(j)}, \quad a_i \in \mathcal{A} \tag{29}$$

with $W_t \triangleq \sum_{j=1}^{m} w_t^{(j)}$. Following [4] and [24], an efficient trial sampling distribution at time t is

$$q\left(s_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right) = p\left(s_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right). \tag{30}$$

For this trial distribution, the importance weight is updated by

$$\begin{aligned} w_t^{(j)} &= w_{t-1}^{(j)} \frac{p\left(\boldsymbol{S}_t^{(j)} \left| \boldsymbol{Y}_t \right.\right)}{p\left(\boldsymbol{S}_{t-1}^{(j)} \left| \boldsymbol{Y}_{t-1} \right.\right) p\left(s_t^{(j)} \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right)} \\ &\propto w_{t-1}^{(j)} \cdot p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1} \right.\right). \end{aligned} \tag{31}$$

Note that $p(y_t|\boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1})$ can be computed by

$$\begin{aligned} &p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1} \right.\right) \\ &\quad \propto \sum_{a_i \in \mathcal{A}} \underbrace{p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)} s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) \cdot p(s_t = a_i)}_{\triangleq \gamma_{t,i}^{(j)}} \end{aligned} \tag{32}$$

where

$$\begin{aligned} p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) &= \int p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1}, \boldsymbol{x} \right.\right) \\ &\quad \cdot p\left(\boldsymbol{x} \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1} \right.\right) \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{33}$$

If we assume a Gaussian prior distribution on the wavelet coefficients, i.e., $\boldsymbol{x} \sim \mathcal{N}_c(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, then we have

$$p\left(\boldsymbol{x} \left| \boldsymbol{S}_t^{(j)}, \boldsymbol{Y}_t \right.\right) \propto p\left(\boldsymbol{Y}_t \left| \boldsymbol{S}_t^{(j)}, \boldsymbol{x} \right.\right) p(\boldsymbol{x}) \sim \mathcal{N}_c\left(\boldsymbol{\mu}_t^{(j)}, \boldsymbol{\Sigma}_t^{(j)}\right) \tag{34}$$

where

$$\boldsymbol{\Sigma}_t^{(j)} \triangleq \left[\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \sum_{i=1}^{t} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T\right]^{-1} \tag{35}$$

$$\text{and } \boldsymbol{\mu}_t^{(j)} \triangleq \boldsymbol{\Sigma}_t^{(j)} \left[\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{i=1}^{t} y_i s_i^{(j)*} \boldsymbol{\phi}_i\right]. \tag{36}$$

Substituting (34) into (33), we obtain

$$p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) \sim \mathcal{N}_c\left(\mu_{t,i}^{(j)}, \sigma_{t,i}^{2(j)}\right) \tag{37}$$

with mean and variance given, respectively, by

$$\mu_{t,i}^{(j)} = \boldsymbol{\phi}_t^T \boldsymbol{\mu}_{t-1}^{(j)} s_t \Big|_{s_t = a_i} = a_i \boldsymbol{\phi}_t^T \boldsymbol{\mu}_{t-1}^{(j)} \tag{38}$$

$$\begin{aligned} \text{and } \quad \sigma_{t,i}^{2(j)} &= \sigma^2 + s_t \boldsymbol{\phi}_t^T \boldsymbol{\Sigma}_{t-1}^{(j)} \boldsymbol{\phi}_t s_t^* \Big|_{s_t = a_i} \\ &= \sigma^2 + \boldsymbol{\phi}_t^T \boldsymbol{\Sigma}_{t-1}^{(j)} \boldsymbol{\phi}_t. \end{aligned} \tag{39}$$

Therefore, $\gamma_{t,i}^{(j)}$ in (32) can be computed by

$$\gamma_{t,i}^{(j)} = \frac{1}{\pi \sigma_{t,i}^{2(j)}} \exp\left(-\frac{\left\| y_t - \mu_{t,i}^{(j)} \right\|^2}{\sigma_{t,i}^{2(j)}}\right) \cdot p(s_t = a_i). \tag{40}$$

The trial distribution in (30) can be computed as follows:

$$\begin{aligned} p\left(s_t = a_i \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right) &\propto p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) \\ &\quad \cdot p(s_t = a_i) = \gamma_{t,i}^{(j)}. \end{aligned} \tag{41}$$

Note that the *a posteriori* covariance and mean of the wavelet coefficients in (35) and (36) can be updated recursively as follows. Let $\mu_t^{(j)}$ and $\sigma_t^{2(j)}$ be the quantities computed by (38) and (39) for the imputed symbol $s_t^{(j)}$. Using the matrix inversion lemma, (35) and (36) become

$$\boldsymbol{\mu}_t^{(j)} = \boldsymbol{\mu}_{t-1}^{(j)} + \frac{y_t - \mu_t^{(j)}}{\sigma_t^{2(j)}} \boldsymbol{\xi}^{(j)} \tag{42}$$

$$\boldsymbol{\Sigma}_t^{(j)} = \boldsymbol{\Sigma}_{t-1}^{(j)} - \frac{1}{\sigma_t^{2(j)}} \boldsymbol{\xi}^{(j)} \boldsymbol{\xi}^{(j)H} \tag{43}$$

$$\text{with } \quad \boldsymbol{\xi}^{(j)} \triangleq \boldsymbol{\Sigma}_{t-1}^{(j)} \boldsymbol{\phi}_t s_t^{(j)*}. \tag{44}$$

C. Resampling Procedure

The importance sampling weight $w_t^{(j)}$ measures the "quality" of the corresponding imputed signal sequence $\boldsymbol{S}_t^{(j)}$. A relatively small weight implies that the sample is drawn far from the main body of the posterior distribution and has a small contribution in the final estimation. Such a sample is said to be ineffective. If there are too many ineffective samples, the Monte Carlo procedure becomes inefficient. To avoid the degeneracy, a useful resampling procedure, which was suggested in [4] and [24], may be used. Roughly speaking, resampling is multiplying the streams with the larger importance weights while eliminating the ones with small importance weights. A simple resampling procedure consists of the following two steps.

1) Sample a new set of streams $\{\hat{\boldsymbol{S}}_t^{(j)}, \hat{\boldsymbol{\mu}}_t^{(j)}, \hat{\boldsymbol{\Sigma}}_t^{(j)}\}_{j=1}^m$ from $\{\boldsymbol{S}_t^{(j)}, \boldsymbol{\mu}_t^{(j)}\}$, $\boldsymbol{\Sigma}_t^{(j)}\}_{j=1}^m$ with probability proportional to the importance weights $\{w_t^{(j)}\}_{j=1}^m$.
2) Assign equal weight to each stream in $\{\hat{\boldsymbol{S}}_t^{(j)}, \hat{\boldsymbol{\mu}}_t^{(j)}, \hat{\boldsymbol{\Sigma}}_t^{(j)}\}_{j=1}^m$, i.e., $\hat{w}_t^{(j)} = (1/m)$, $j = 1, \ldots, m$.

Resampling can be done at every fixed-length time interval (say, every five steps), or it can be conducted dynamically. The *effective sample size* can be used to monitor the variation of the importance weights of the sample streams and to decide when to resample as the system evolves. The effective sample size is defined as [5]

$$\bar{m}_t \triangleq \frac{m}{1 + \upsilon_t^2} \tag{45}$$

where υ_t, which is the coefficient of variation, is given by

$$\upsilon_t^2 = \frac{1}{m} \sum_{j=1}^{m} \left(\frac{w_t^{(j)}}{\bar{w}_t} - 1\right)^2 \tag{46}$$

with $\bar{w}_t = \sum_{j=1}^{m} w_t^{(j)}/m$. In dynamic resampling, a resampling step is performed once the effective sample size $\bar{m}_t$ is below a certain threshold.

Heuristically, resampling can provide chances for good sample streams to amplify themselves and hence "rejuvenate" the sampler to produce a better result for future states as system evolves. It can be shown that the samples drawn by the above resampling procedure are also indeed properly weighted with respect to $p(\boldsymbol{S}_t|\boldsymbol{Y}_t)$, provided that m is sufficiently large. In practice, when small to modest m is used (we use $m = 100$ in this paper), the resampling procedure can be seen as a tradeoff

between the bias and the variance. That is, the new samples with their weights resulting from the resampling procedure are only approximately proper, which introduces small bias in the Monte Carlo estimation. On the other hand, however, resampling significantly reduces the Monte Carlo variance for future samples.

D. Resampling-Based Adaptive Wavelet Shrinkage

In the previous subsection, we have discussed the SMC algorithm for wavelet-based adaptive detection in fading channels, which assumes that the first κ wavelet coefficients are nonzeros, where κ is some fixed number set *a priori*, whereas the other wavelet coefficients are shrunk to zeros. However, as illustrated in Section II-B.3, fading processes with different Doppler values need a different number of wavelet coefficients to be approximated. We next propose a method that dynamically chooses the number of wavelet coefficients during the SMC sampling procedure.

The basic idea is as follows. At time t, associated with each sample stream j, in addition to the symbol stream sample $\boldsymbol{S}_t^{(j)}$, the posterior channel distribution parameters ($\boldsymbol{\mu}_t^{(j)}$, $\boldsymbol{\Sigma}_t^{(j)}$), and the importance weight $w_t^{(j)}$, we also have a quantity $\kappa_t^{(j)}$, which indicates the number of the wavelet coefficients used by the jth sample stream at time t. Accordingly, the wavelet basis vector and the wavelet coefficients vector at time t become

$$\boldsymbol{\phi}_t^{(j)T} \triangleq \boldsymbol{\phi}_t^T\left[1:\kappa_t^{(j)}\right] \text{ and } \boldsymbol{x}_t^{(j)} \triangleq \boldsymbol{x}_t\left[1:\kappa_t^{(j)}\right].$$

Let $\kappa_{\min}$ and $\kappa_{\max}$ be, respectively, the minimum and maximum numbers of the wavelet coefficients used in the algorithm. Initially, $\kappa_0^{(j)}$ is drawn according to some *a priori* distribution from the set $\{\kappa_{\min}, \kappa_{\min}+1, \ldots, \kappa_{\max}\}$ for each $j = 1, 2, \ldots, m$.

Assume that we have properly weighted samples $\{\boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{\mu}_{t-1}^{(j)}, \boldsymbol{\Sigma}_{t-1}^{(j)}, \kappa_{t-1}^{(j)}, w_{t-1}^{(j)}\}_{j=1}^m$ at time $(t-1)$. At time t, we first let $\kappa_t^{(j)} = \kappa_{t-1}^{(j)}$ and update the samples to obtain $\{\boldsymbol{S}_t^{(j)}, \boldsymbol{\mu}_t^{(j)}, \boldsymbol{\Sigma}_t^{(j)}, \kappa_t^{(j)}, w_t^{(j)}\}_{j=1}^m$, as described above. We then check the resampling condition, and if the effective sample size is below a specified threshold, then resampling is conducted, and we get a new set of samples $\{\hat{\boldsymbol{S}}_t^{(j)}, \hat{\boldsymbol{\mu}}_t^{(j)}, \hat{\boldsymbol{\Sigma}}_t^{(j)}, \hat{\kappa}_t^{(j)}, \hat{w}_t^{(j)} = (1/m)\}_{j=1}^m$. Hence, in this way, adaptive wavelet shrinkage is achieved through resampling. That is, samples employing a proper number of wavelet coefficients that well fit the received signals are rejuvenated, whereas those employing an improper number of wavelet coefficients are eliminated. Moreover, it is evident that the fading process is effectively modeled by a mixture of wavelet expansions with different orders, and the mixture distribution dynamically evolves during the SMC procedure.

Finally, we summarize the wavelet-based SMC blind receiver algorithm in flat-fading channels as follows.

0) Initialization: For each $j = 1, 2, \ldots, m$, do the following.
   - Sample $\kappa_0^{(j)}$ uniformly from $[\kappa_{\min}, \kappa_{\max}]$.
   - Set $\boldsymbol{\Sigma}_0^{(j)} = 1000\boldsymbol{I}_{\kappa_0^{(j)}}$. Set $\boldsymbol{\mu}_0^{(j)} \sim \mathcal{N}_c(0, 1000\boldsymbol{I}_{\kappa_0^{(j)}})$.
   - Set $w_0^{(j)} = (1/m)$.

   The following steps are implemented at time $t (t = 1, \ldots, K_0)$ to update each weighted sample. For $j = 1, \cdots, m$, do the following.
1) Let $\kappa_t^{(j)} = \kappa_{t-1}^{(j)}$. Set $\boldsymbol{\phi}_t^{(j)T} \triangleq \boldsymbol{\phi}_t^T[1:\kappa_t^{(j)}]$.
2) For each $a_i \in \mathcal{A}$, compute $\mu_{t,i}^{(j)}$, $\sigma_{t,i}^{2(j)}$ and $\gamma_{t,i}^{(j)}$ given, respectively, by (38)–(40).
3) Draw $s_t^{(j)}$ from $\mathcal{A}$ with probability
$$p\left(s_t = a_i \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t\right.\right) \propto \gamma_{t,i}^{(j)}, \quad a_i \in \mathcal{A}.$$
   Append $s_t^{(j)}$ to $\boldsymbol{S}_{t-1}^{(j)}$ to obtain $\boldsymbol{S}_t^{(j)}$.
4) Compute the importance weight
$$w_t^{(j)} \propto w_{t-1}^{(j)} \sum_{a_i \in \mathcal{A}} \gamma_{t,i}^{(j)}.$$
5) Suppose in Step 3 that the imputed sample $s_t^{(j)} = a_i$. Then, let $\mu_t^{(j)} = \mu_{t,i}^{(j)}$ and $\sigma_t^{2(j)} = \sigma_{t,i}^{2(j)}$. Update $\boldsymbol{\mu}_t^{(j)}$ and $\boldsymbol{\Sigma}_t^{(j)}$ according to (42) and (43).
6) Compute the effective sample size $\bar{m}_t$ given by (70). If $\bar{m}_t \le \lambda m$ (in this paper, $\lambda = 0.1$), then perform the following resampling steps to obtain a new set of sample streams.
   a) Sample a new set of streams $\{\hat{\boldsymbol{S}}_t^{(j)}, \hat{\boldsymbol{\mu}}_t^{(j)}, \hat{\boldsymbol{\Sigma}}_t^{(j)}, \hat{\kappa}_t^{(j)}\}_{j=1}^m$ from $\{\boldsymbol{S}_t^{(j)}, \boldsymbol{\mu}_t^{(j)}, \boldsymbol{\Sigma}_t^{(j)}, \kappa_t^{(j)}\}_{j=1}^m$ with probability proportional to the importance weights $\{w_t^{(j)}\}_{j=1}^m$.
   b) Assign equal weight to each stream in $\{\hat{\boldsymbol{S}}_t^{(j)}, \hat{\boldsymbol{\mu}}_t^{(j)}, \hat{\boldsymbol{\Sigma}}_t^{(j)}, \hat{\kappa}_t^{(j)}\}_{j=1}^m$, i.e., $\hat{w}_t^{(j)} = (1/m)$, $j = 1, \ldots, m$.

Complexity: The dominant computation in the above algorithm consists of the m matrix-vector products $\boldsymbol{\Sigma}_{t-1}^{(j)}\boldsymbol{\phi}_t$ in (39) and (43) and the m vector out-products $\boldsymbol{\xi}\boldsymbol{\xi}^H$ in (44), whose computation is determined by the number of wavelet coefficients $\kappa_t^{(j)}$ in every Markov stream. Therefore, the total computation mainly includes mK_0 matrix-vector products and mK_0 vector outproducts. Note that the parametric SMC receiver in [5] needs nearly the same number of matrix-vector and vector-out products.

E. Delayed-Weight Estimation

From the recursive procedure described in Section III-D, we get the samples $\{(\boldsymbol{S}_{t+\delta}^{(j)}, w_{t+\delta}^{(j)})\}_{j=1}^m$ at time $(t+\delta)$, $\delta > 0$, which are properly weighted with respect to $p(\boldsymbol{S}_{t+\delta}|\boldsymbol{Y}_{t+\delta})$. Hence, focusing on $\boldsymbol{S}_t$ at time $(t+\delta)$, we obtain a delayed estimation of the symbol

$$p(s_t = a_i|\boldsymbol{Y}_{t+\delta}) \cong \frac{1}{W_{t+\delta}} \sum_{j=1}^m 1\left(s_t^{(j)} = a_i\right) w_{t+\delta}^{(j)}, \quad a_i \in \mathcal{A} \tag{47}$$

with $W_{t+\delta} \triangleq \sum_{j=1}^m w_{t+\delta}^{(j)}$. Since the weights $\{w_{t+\delta}^{(j)}\}_{j=1}^m$ contain information about the signals $(y_{t+1}, \cdots, y_{t+\delta})$, the estimation in (47) is usually more accurate. Note that such a delayed estimation method incurs no additional computational cost (i.e., cpu time), but it requires some extra memory for storing $\{s_{t+1}^{(j)}, \ldots, s_{t+\delta}^{(j)}\}_{j=1}^m$.

IV. SMC Receiver for Frequency-Selective Fading Channels

A. Wavelet Model for Frequency-Selective Fading Channel

The discrete-time input-output relationship of a frequency-selective fading channel is given by [30]

$$y_t = \sum_{l=0}^{L-1} \alpha_{t,l} s_{t-l} + \nu_t \tag{48}$$

where $\{y_t\}$, $\{s_t\}$, and $\{\nu_t\}$ are, respectively, the received signal, the transmitted symbol, and the complex Gaussian noise sample at time t; L is the number of the resolvable paths of the channel; and $\alpha_{t,l}$ is the fading coefficient of the lth path at time t. It is assumed that the fading processes of different paths $\{\alpha_{t,l}\}_t$, $l = 0,1,\ldots,L-1$ are independent and that they have the same Doppler value. Following the discussion in Section II-B, the fading process of the lth path can be represented by a wavelet basis as follows:

$$\alpha_{t,l} = \boldsymbol{\phi}_t^T \boldsymbol{x}_l, \quad t = 1,2,\ldots,K_0 \tag{49}$$

where $\boldsymbol{\phi}_t^T$ is the tth row of $\boldsymbol{\Phi}$ given by (17), and $\boldsymbol{x}_l$ is the wavelet coefficient vector for the lth path. As before, we assume that the first κ wavelet coefficients are used to approximate the fading process. Denoting $\boldsymbol{x}_l \triangleq \boldsymbol{x}_l[1:\kappa]$ and $\boldsymbol{\phi}_t^T \triangleq \boldsymbol{\phi}_t^T[1:\kappa]$, then we have $\alpha_{t,l} \cong \boldsymbol{\phi}_t^T \boldsymbol{x}_l$. Substituting this into (48), we get

$$y_t = \boldsymbol{\phi}_t^T \sum_{l=0}^{L-1} \boldsymbol{x}_l s_{t-l} + \nu_t, \quad t = 1,\ldots,K_0. \tag{50}$$

Denote

$$\begin{aligned} \boldsymbol{x} &\triangleq [\boldsymbol{x}_0^T, \boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{L-1}^T]^T \\ \boldsymbol{s}_t &\triangleq [s_t, s_{t-1}, \ldots, s_{t-L+1}]^H \end{aligned}$$

and define an $(L \times L\kappa)$ block diagonal matrix $\boldsymbol{\Psi}_t = \text{diag}[\boldsymbol{\phi}_t^T, \boldsymbol{\phi}_t^T, \ldots, \boldsymbol{\phi}_t^T]$; then, (50) can be rewritten as

$$y_t = \boldsymbol{s}_t^H \boldsymbol{\Psi}_t \boldsymbol{x} + \nu_t, \quad t = 1,2,\ldots,K_0. \tag{51}$$

B. SMC Blind Receiver

As before, if we are to make inference with respect to $p(\boldsymbol{S}_t|\boldsymbol{Y}_t)$, we need to draw samples from the trial distribution

$$\begin{aligned} q\left(s_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right) &= p\left(s_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_t \right.\right) \\ &\propto p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t, \boldsymbol{Y}_{t-1} \right.\right) p(s_t) \\ &= p(s_t) \int p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t, \boldsymbol{Y}_{t-1}, \boldsymbol{x} \right.\right) \\ &\quad \cdot p\left(\boldsymbol{x} \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1} \right.\right) \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{52}$$

Assuming a Gaussian prior on the wavelet coefficients, i.e., $\boldsymbol{x} \sim \mathcal{N}_c(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, we have

$$p\left(\boldsymbol{x} \left| \boldsymbol{S}_t^{(j)}, \boldsymbol{Y}_t \right.\right) \propto p\left(\boldsymbol{Y}_t \left| \boldsymbol{S}_t^{(j)}, \boldsymbol{x} \right.\right) p(\boldsymbol{x}) \sim \mathcal{N}_c\left(\boldsymbol{\mu}_t^{(j)}, \boldsymbol{\Sigma}_t^{(j)}\right) \tag{53}$$

with

$$\boldsymbol{\Sigma}_t^{(j)} \triangleq \left[\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \sum_{i=1}^{t} \Psi_i^T \boldsymbol{s}_i^{(j)} \boldsymbol{s}_i^{(j)H} \boldsymbol{\Psi}_i\right]^{-1} \tag{54}$$

$$\boldsymbol{\mu}_t^{(j)} \triangleq \boldsymbol{\Sigma}_t^{(j)} \left[\Sigma_0^{-1} \boldsymbol{\mu}_0 + \frac{1}{\sigma^2} \sum_{i=1}^{t} y_i \boldsymbol{\Psi}_i^T \boldsymbol{s}_i^{(j)}\right]. \tag{55}$$

Hence

$$p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) \sim \mathcal{N}_c\left(\mu_{t,i}^{(j)}, \sigma_{t,i}^{2(j)}\right) \tag{56}$$

with

$$\mu_{t,i}^{(j)} = \boldsymbol{s}_t^{(j)H} \boldsymbol{\Psi}_t \boldsymbol{\mu}_{t-1}^{(j)} \Big|_{s_t = a_i} \tag{57}$$

$$\text{and} \quad \sigma_{t,i}^{2(j)} = \sigma^2 + \boldsymbol{s}_t^{(j)H} \boldsymbol{\Psi}_t \boldsymbol{\Sigma}_{t-1}^{(j)} \Psi_t^T \boldsymbol{s}_t^{(j)} \Big|_{s_t = a_i}. \tag{58}$$

Denote the quantities $\mu_t^{(j)}$ and $\sigma_t^{2(j)}$ corresponding to the imputed sample $s_t^{(j)}$; then, $\boldsymbol{\mu}_t^{(j)}$ and $\boldsymbol{\Sigma}_t^{(j)}$ in (55) and (54) can be recursively calculated as

$$\boldsymbol{\mu}_t^{(j)} = \boldsymbol{\mu}_{t-1}^{(j)} + \frac{y_t - \mu_t^{(j)}}{\sigma_t^{2(j)}} \boldsymbol{\xi} \tag{59}$$

$$\boldsymbol{\Sigma}_t^{(j)} = \boldsymbol{\Sigma}_{t-1}^{(j)} - \frac{1}{\sigma_t^{2(j)}} \boldsymbol{\xi}\boldsymbol{\xi}^H \tag{60}$$

$$\text{with} \quad \boldsymbol{\xi} \triangleq \boldsymbol{\Sigma}_{t-1}^{(j)} \Psi_t^T \boldsymbol{s}_t^{(j)}. \tag{61}$$

Finally, the importance weight is updated according to

$$\begin{aligned} w_t^{(j)} &\propto w_{t-1}^{(j)} p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t-1} \right.\right) \\ &= w_{t-1}^{(j)} \sum_{a_i \in \mathcal{A}} \left[p\left(y_t \left| \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \boldsymbol{Y}_{t-1} \right.\right) p(s_t = a_i)\right]. \end{aligned} \tag{62}$$

The frequency-selective fading channel exhibits a strong memory, and the "future" observations $\{y_i\}_{i>t}$ contain information about the current data symbol s_t. The delayed-weight method discussed in Section III-E is not sufficient to exploit the channel memory effects. A more efficient delayed-sample method proposed in [5] is described next.

C. Delayed-Sample Estimation

In delayed-sample estimation, we generate both the samples and the weights $\{\boldsymbol{S}_t^{(j)}, w_t^{(j)}\}_{j=1}^m$ based on the signals $\boldsymbol{Y}_{t+\Delta}$, hence making $p(\boldsymbol{S}_t|\boldsymbol{Y}_{t+\Delta})$ the target distribution at time $(t+\Delta)$. The procedure will provide better Monte Carlo samples since it utilizes the future observation $(y_{t+1},\ldots,y_{t+\Delta})$ in generating the current samples of s_t. The basic idea is to marginalize out $\{s_{t+\tau}\}_{\tau=1}^{\Delta}$ in the sampling procedure.

Following [4], a useful trial distribution in this case is given by

$$\begin{aligned} \rho_{t,i}^{(j)} &\triangleq q\left(s_t = a_i \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t+\Delta} \right.\right) \\ &= p\left(s_t = a_i \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t+\Delta} \right.\right), \quad i = 1,2,\ldots,|\mathcal{A}|. \end{aligned} \tag{63}$$

Denote $\underline{s}_a^b \triangleq (s_a, s_{a+1}, \ldots, s_b)$. The trial distribution can be computed by

$$\rho_{t,i}^{(j)} = \sum_{\underline{s}_{t+1}^{t+\Delta} \in \mathcal{A}^\Delta} \left[\prod_{\tau=0}^{\Delta} \underbrace{p\left(y_{t+\tau} \left| \boldsymbol{Y}_{t+\tau-1}, \boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\tau}\right.\right)}_{\triangleq \beta_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\tau}\right)} \times p(s_t = a_i) \prod_{\tau=1}^{\Delta} p(s_{t+\tau}) \right], i = 1, \ldots, |\mathcal{A}|. \quad (64)$$

Assuming the imputed sample $s_{t-1}^{(j)} = a_k$, the importance weight is updated according to [5]

$$w_t^{(j)} \propto \frac{w_{t-1}^{(j)}}{\rho_{(t-1),k}^{(j)}} p\left(y_{t-1} \left| \boldsymbol{Y}_{t-2}, \boldsymbol{S}_{t-1}^{(j)}\right.\right) p(s_{t-1} = a_k) \sum_{i=1}^{|\mathcal{A}|} \rho_{t,i}^{(j)}. \quad (65)$$

The conditional distribution $p(y_{t+\tau}|\boldsymbol{Y}_{t+\tau-1}, \boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau})$, $\tau = 0, 1, \ldots, \Delta$ in (64) are Gaussian

$$p\left(y_{t+\tau} \left| \boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t+\tau-1}, \underline{s}_t^{t+\tau}\right.\right) \sim \mathcal{N}_c\left(\mu_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) \sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)\right), \tau = 0, 1, \ldots, \Delta \quad (66)$$

with

$$\mu_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) = \boldsymbol{s}_{t+\tau}^{(j)H} \boldsymbol{\Psi}_{t+\tau} \times \boldsymbol{\mu}_{t+\tau-1}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1}\right) \quad (67)$$

and

$$\sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) = \sigma^2 + \boldsymbol{s}_{t+\tau}^{(j)H} \boldsymbol{\Psi}_{t+\tau} \times \boldsymbol{\Sigma}_{t+\tau-1}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1}\right) \times \boldsymbol{\Psi}_{t+\tau}^T \boldsymbol{s}_{t+\tau}^{(j)} \quad (68)$$

where $\boldsymbol{s}_{t+\tau}^{(j)} \triangleq [s_{t+\tau}^{(j)}, s_{t+\tau-1}^{(j)}, \ldots, s_{t+\tau-L+1}^{(j)}]^T$. Hence, $\beta_{t+\tau}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau})$ in (64) is given by

$$\beta_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) \triangleq \frac{1}{\pi \sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)} \times \exp\left(- \frac{\left\| y_{t+\tau} - \mu_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)\right\|^2}{\sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)} \right). \quad (69)$$

The terms $\boldsymbol{\mu}_{t+\tau-1}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1})$ and $\boldsymbol{\Sigma}_{t+\tau-1}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1})$ in (67) and (68) can be computed recursively as follows. Let $\mu_{t+\tau}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau})$ and $\sigma_{t+\tau}^{2(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau})$ be the quantities computed by (67) and (68); therefore, we have

$$\boldsymbol{\mu}_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) = \boldsymbol{\mu}_{t+\tau-1}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1}\right) + \frac{y_{t+\tau} - \mu_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)}{\sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)} \boldsymbol{\xi}, \quad (70)$$

and

$$\boldsymbol{\Sigma}_{t+\tau}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right) = \boldsymbol{\Sigma}_{t+\tau-1}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1}\right) - \frac{1}{\sigma_{t+\tau}^{2(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}\right)} \boldsymbol{\xi}\boldsymbol{\xi}^H, \quad (71)$$

with $\boldsymbol{\xi} \triangleq \boldsymbol{\Sigma}_{t+\tau-1}^{(j)}\left(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau-1}\right) \boldsymbol{\Psi}_{t+\tau}^T \boldsymbol{s}_{t+\tau}^{(j)}$. (72)

Hence, for each possible "future" (relative to time $t-1$) symbol sequence at time $(t+\Delta-1)$, i.e., $(s_t, s_{t+1}, \ldots, s_{t+\Delta-1}) \in \mathcal{A}^\Delta$, we keep the values of Δ-step recursive updates $\{\boldsymbol{\mu}_{t+\Delta-1}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\Delta-1}), \boldsymbol{\Sigma}_{t+\Delta-1}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}\underline{s}_t^{t+\Delta-1}), \beta_{t+\tau}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau}), \tau = 0, \ldots, \Delta-1\}_{j=1}^m$. Then, the delayed-sample algorithm with resampling-based adaptive wavelet shrinkage is summarized as follows.

0) Initialization: For each $j = 1, 2, \ldots, m$, do the following.
   - Sample $\kappa_0^{(j)}$ uniformly from $[\kappa_{\min}, \kappa_{\max}]$. Set $\boldsymbol{\phi}_{\tau,l}^{(j)T} = \boldsymbol{\phi}_{\tau,l}^T[1 : \kappa_0^{(j)}]$, and obtain the corresponding $\boldsymbol{\Psi}_\tau^{(j)}$, $\tau = 1, \ldots, \Delta$.
   - Sample $\underline{s}_{(1-L)}^{0(j)}$ from $\mathcal{A}^L$ with equal probability, and obtain $\boldsymbol{s}_0^{(j)}$
   - Set $\boldsymbol{\Sigma}_0^{(j)} = 1000\boldsymbol{I}_{\kappa_0^{(j)}}$. Draw $\boldsymbol{\mu}_0^{(j)} \sim \mathcal{N}_c(0, 1000\boldsymbol{I}_{\kappa_0^{(j)}})$.
   - Set $w_0^{(j)} = (1/m)$.
   - Compute $\{\boldsymbol{\mu}_\tau^{(j)}(\underline{s}_1^\tau)\}_{j=1}^m$, $\{\boldsymbol{\Sigma}_\tau^{(j)}(\underline{s}_1^\tau)\}_{j=1}^m$, and $\{\beta_\tau^{(j)}(\underline{s}_1^\tau)\}_{j=1}^m$, $\tau = 1, 2, \ldots \Delta$, by recursively using the (69)–(71) for each value $\underline{s}_1^\Delta \in \mathcal{A}^\Delta$ at time $t = \Delta$.

   The following steps are implemented at time $(t+\Delta)(t = 1, \ldots, K-\Delta)$ to update each weighted sample. For $j = 1, \cdots, m$, do the following.
1) Let $\kappa_{t+\Delta}^{(j)} = \kappa_{t+\Delta-1}^{(j)} = \ldots = \kappa_{t-1}^{(j)}$. Set $\boldsymbol{\phi}_{(t+\tau),l}^{(j)T} = \boldsymbol{\phi}_{(t+\tau),l}^T[1 : \kappa_t^{(j)}]$, and obtain the corresponding $\boldsymbol{\Psi}_{t+\tau}^{(j)}$, $\tau = 1, \ldots, \Delta$.
2) For each $s_{t+\Delta} = a_i \in \mathcal{A}$ and $\underline{s}_t^{t+\Delta-1} \in \mathcal{A}^\Delta$, compute $\mu_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\Delta-1}, s_{t+\Delta} = a_i)$, $\sigma_{t+\Delta}^{2(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\Delta-1}, s_{t+\Delta} = a_i)$ and $\beta_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\Delta-1}, s_{t+\Delta} = a_i)$ according to (67)–(69).
3) Draw $s_t^{(j)}$ from $\mathcal{A}$ with probability $p(s_t = a_i|\boldsymbol{S}_{t-1}^{(j)}, \boldsymbol{Y}_{t+\Delta}) = \rho_{t,i}^{(j)}$, where $\rho_{t,i}^{(j)}$ is given by (64). Append $s_t^{(j)}$ to $\boldsymbol{S}_{t-1}^{(j)}$ to obtain $\boldsymbol{S}_t^{(j)}$.
4) Update $w_t^{(j)}$ according to (65).
5) Suppose, in Step 3, that the imputed sample $s_t^{(j)} = a_i$. Then, for each value in $\underline{s}_{t+1}^{t+\Delta} \in \mathcal{A}^\Delta$, update $\boldsymbol{\mu}_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\Delta})$ and $\boldsymbol{\Sigma}_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\Delta})$ according to (70) and (71), and keep the values $\{\boldsymbol{\mu}_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\Delta})\}_{j=1}^m$, $\{\boldsymbol{\Sigma}_{t+\Delta}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\Delta})\}_{j=1}^m$, and $\{\beta_{t+\tau}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, s_t = a_i, \underline{s}_{t+1}^{t+\tau})\}_{j=1}^m$, $\tau = 1, 2, \ldots \Delta$.
6) Compute the effective sample size $\bar{m}_t$ given by (70). If $\bar{m}_t \le \lambda m$, then perform the resampling described in Section III-D.

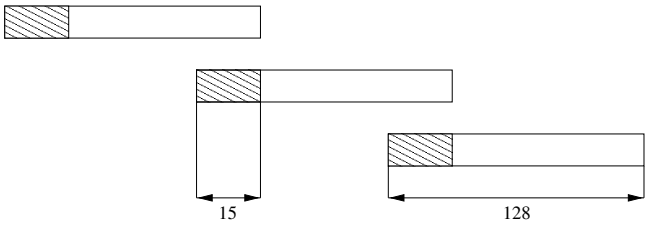

Fig. 2. Frame structure for the wavelet-based SMC receivers. The block size is 128, and the adjacent blocks overlap by 15 symbols.

Complexity: The dominant computation in the above algorithm involves the $m\mathcal{A}^{\Delta}$ matrix-vector products $\boldsymbol{\Sigma}_{t+\tau-1}^{(j)}(\boldsymbol{S}_{t-1}^{(j)}, \underline{s}_t^{t+\tau})\boldsymbol{\Psi}_{t+\tau}$ in (68) and (72) and the $m\mathcal{A}^{\Delta}$ vector-out products $\boldsymbol{\xi}\boldsymbol{\xi}^H$ in (71), whose computations are determined by the number of wavelet coefficients $\kappa_t^{(j)}$ in every Markov stream. Therefore, the total computation mainly includes $mK_0\mathcal{A}^{\Delta}$ matrix-vector products and $mK_0\mathcal{A}^{\Delta}$ vector out-products.

Finally, as noted in [5], we can use the above delayed-sample method in conjunction with the delayed-weight method. For example, using the delayed-sample method, we generate delayed samples and weights $\{(\boldsymbol{S}_t^{(j)}, w_t^{(j)})\}_{j=1}^m$ based on observations $\boldsymbol{Y}_{t+\Delta}$. Then, with an additional delay δ, we can use the following delayed weight method to approximate the a posterior probability of symbol s_t

V. Simulation Results

In this section, we present some simulation examples to illustrate the performance of the proposed wavelet-based sequential Monte Carlo receivers in both flat-fading and frequency-selective fading channels. Of particular interest is the resampling-based wavelet shrinkage technique, which dynamically chooses the number of wavelet coefficients to best fit the fading process. We first outline the simulation setup.

The BPSK modulation is employed in the simulation, i.e., the transmitted symbols $\{s_t\}$ take values from ± 1. The characteristics of the fading channels are described in Section II. The Jakes' fading process is generated using the frequency spectrum method [29]. Note that in [5], the fading process is generated according to an ARMA process. Such a low-order ARMA model cannot fit the Jakes' fading spectrum (3) well. In the decomposition of the fading process, the Daubechies wavelet filter with order 2 is used to construct the reconstruction matrix given by (17). Our simulations show that little performance improvement can be achieved with Daubechies filters of order higher than 2, whereas with Daubechies filter of order 1 (i.e, Harr wavelet), the performance degradation is significant. The reason for the performance degradation when using the Harr wavelet is the zig-zag approximation error caused by the special structure of Harr wavelet transform.

The signal frame structure is shown in Fig. 2. Each data block contains $K_0 = 128$ symbols. Adjacent blocks overlap by 15 symbols to allow the SMC filter to reach the steady state. To speed up the convergence, for each data block, the values of the mean and the covariance of wavelet coefficients $\{\boldsymbol{\mu}_0^{(j)}, \boldsymbol{\Sigma}_0^{(j)}\}_{j=1}^m$ are initialized as the corresponding values $\{\boldsymbol{\mu}_{K_0}^{(j)}, \boldsymbol{\Sigma}_{K_0}^{(j)}\}_{j=1}^m$ at the end of the previous block.

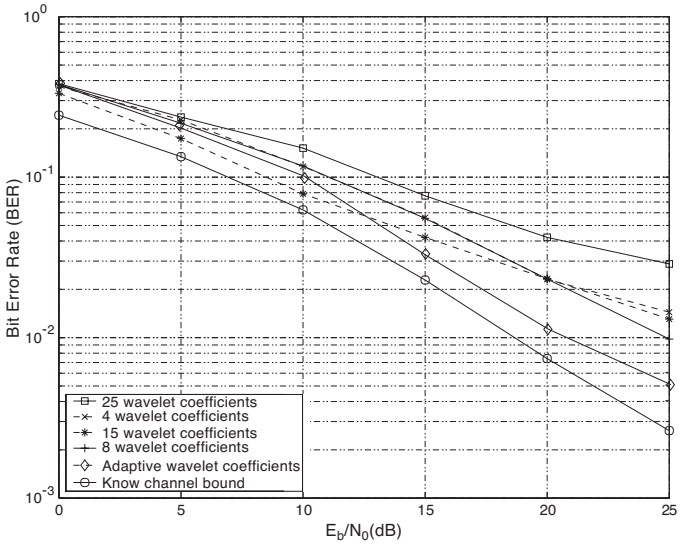

Fig. 3. BER performance of the SMC receiver using fixed number (e.g., 4, 8, 15, and 25) of wavelet coefficients and the adaptive wavelet-based SMC receiver. The channel is flat-fading with $f_dT = 0.005$. The delayed-weight method is used with $\delta = 6$.

The performance of the proposed SMC blind receivers is compared with that of the receivers with perfect channel state information (CSI). In flat-fading channels, the receiver with CSI makes a decision on symbol s_t according to $\hat{s}_t = \text{sign}(\Re\{\alpha_t^* y_t\})$, whereas in frequency-selective fading channels, the receiver with CSI makes use of the Viterbi Algorithm to estimate the transmitted symbol sequence [30]. We call the performance of the receiver with CSI the "known channel bound."

A. Performance of Wavelet-Based SMC in Flat Fading Channels

First, we consider the performance of the SMC blind receiver with different number of wavelet coefficients in a flat-fading channel with normalized Doppler $f_dT = 0.005$. The SMC receiver discussed in Section III is implemented with different but fixed number of wavelet coefficients. The number of the Monte Carlo samples drawn at each time is empirically set as $m = 100$. The resampling procedure is employed in the SMC, and the threshold of effective sample size is $\bar{m}_t = m/10$. The delayed-weight method is used with $\delta = 6$. The bit error rate (BER) versus the signal-to-noise ratio is plotted in Fig. 3 for different fixed number of wavelet coefficients. In Fig. 3, we also draw the BER performance achieved by the resampling-based wavelet shrinkage method. It is seen that in this case, the best performance (close to the known channel bound) is achieved using eight wavelet coefficients. With four wavelet coefficients, the performance is a bit worse, and the performance is substantially degraded if the number of wavelet coefficients is very large (e.g, 15, 25, and 32). The BER performance of the SMC blind receiver in a flat-fading channel with normalized Doppler $f_dT = 0.01$ is shown in Fig. 4. It is seen that in this case, when

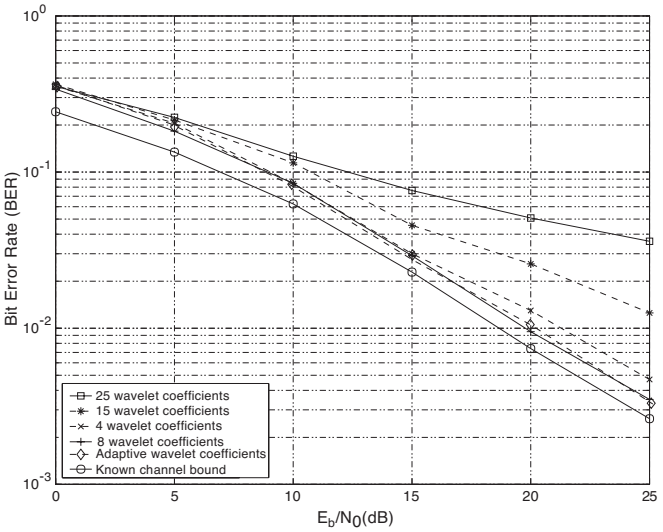

Fig. 4. BER performance of the SMC receiver using fixed number (e.g., 4, 8, 15, and 25) of wavelet coefficients and the adaptive wavelet-based SMC receiver. The channel is flat-fading with $f_dT = 0.01$. The delayed-weight method is used with $\delta = 6$.

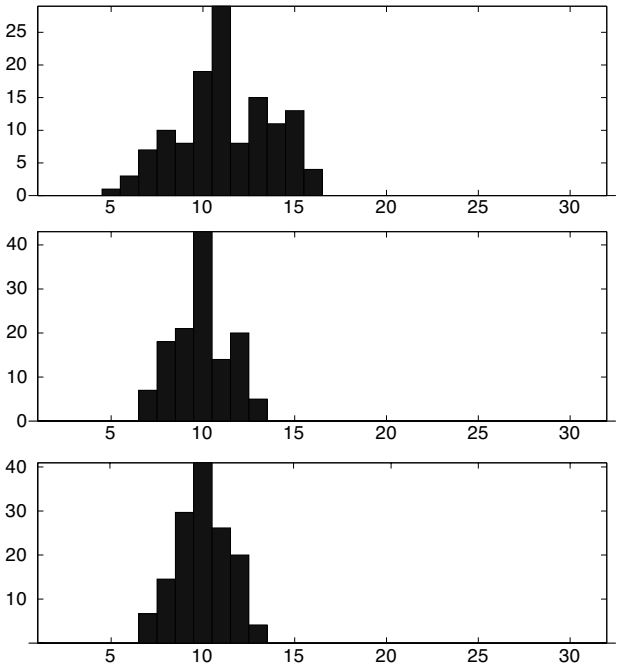

Fig. 5. Histogram for the number of the wavelet coefficients associated with the sample streams at times $t = 20$ (top), $t = 40$ (middle), and $t = 64$ (bottom) for $f_dT = 0.005$ and $E_b/N_0 = 10$ dB. The result is the average over 100 simulations.

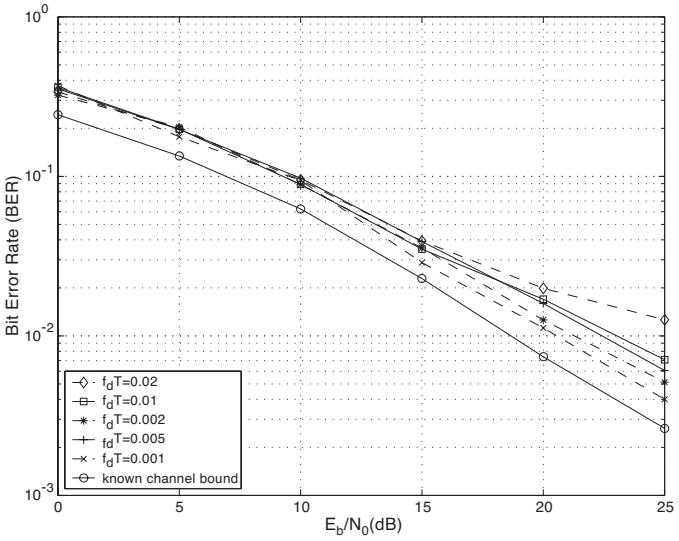

Fig. 6. BER performance of the adaptive shrinkage SMC receiver. The channel is flat fading under different Dopplers. The delayed-weight method is used with $\delta = 6$.

$E_b/N_0 < 20$ dB, the best number of wavelet coefficients is 15; otherwise, the best number is 8.

However, simulation results seem to be contrary to what is shown in Fig. 1, where more coefficients give better channel fading representation accuracy. The reason for the BER degradation using more wavelet coefficients is roughly summarized as follows. First, the approximation error in Fig. 1 is achieved without considering the effects of noise. Since the noise most often occurs in high frequency, the extra wavelet coefficients will bring noise into the receiver, which degrade the performance of the receiver. Furthermore, to track a larger number of wavelet coefficients would require a much longer burn-in period and less accuracy than to track fewer of them.

We next show the performance of the SMC receiver employing the resampling-based wavelet shrinkage in flat-fading channels. The possible number of wavelet coefficients for $\{\kappa_t^{(j)}\}_{j=1}^m$ is chosen from the set $\{1, 2, \ldots, 32\}$. Fig. 5 shows the histogram of the number of wavelet coefficients associated with each sample stream at times $t = 20$, $t = 40$, and $t = 64$ for $f_dT = 0.005$ and $E_b/N_0 = 10$ dB. The result is the average over 100 simulations. It is seen that the numbers of wavelet coefficients focus on the range of (5,16) at time $t = 20$ and the range of (7,13) at time $t = 40$ and $t = 64$, respectively. Fig. 6 illustrates the BER performance of the adaptive shrinkage SMC receiver in flat-fading channels with different Doppler values. It is seen that the proposed blind receiver achieves very good performance (close to the known channel bound) if the Doppler is not very high (e.g., $f_dT < 0.01$). However, for very fast fading channels (e.g., $f_dT \geq 0.01$), the performance is substantially degraded due to the difficulty in tracking the fast variation of the channel.

Finally, we compare the performance of the wavelet-based nonparametric SMC receiver with the parametric SMC receiver proposed in [5]. To employ the parametric SMC receiver, we need to obtain an ARMA representation of the fading process based on its second-order statistics. To that end, a least square fitting is used to estimate ARMA model parameters (as in [5], both the AR and MA parts of the model have order 3). The fading process is generated based on Jakes's fading model [32]. In Fig. 7, the BER performance is shown for both methods under different fading coefficients (e.g., $f_dT = 0.01$ or 0.001). It is seen that in this case, much better performance (nearly 10 dB better) is achieved using the wavelet-based SMC receiver. Moreover, an error floor is seen at high SNR for the parametric SMC receiver because of the mismatch between the ARMA model and real fading process.

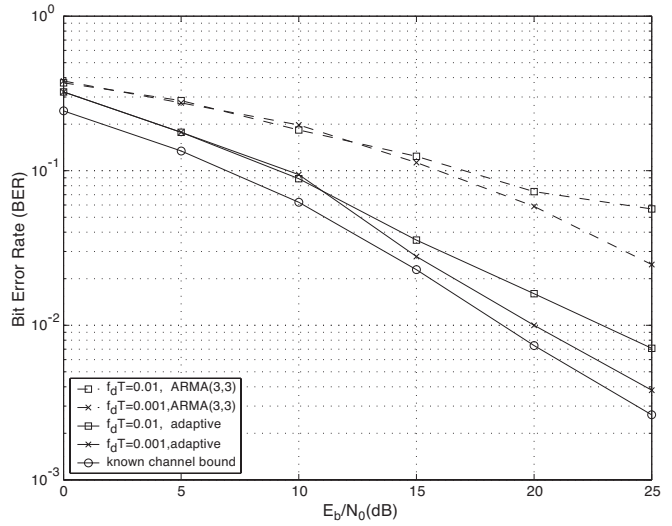

Fig. 7. BER performance of the adaptive shrinkage SMC receiver and the mixture Kalman filter with statistically approximated ARMA models. The channel is flat fading under different Dopplers. The delayed-weight method is used with $\delta = 6$.

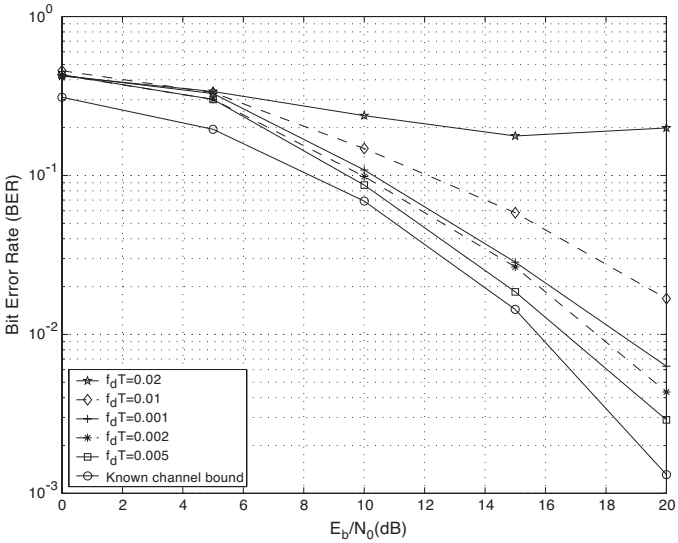

Fig. 9. BER performance of the adaptive shrinkage SMC receiver. The channel is frequency-selective fading with $L = 2$. The delayed-sample/delayed-weight methods are used with $\Delta = 2$ and $\delta = 6$.

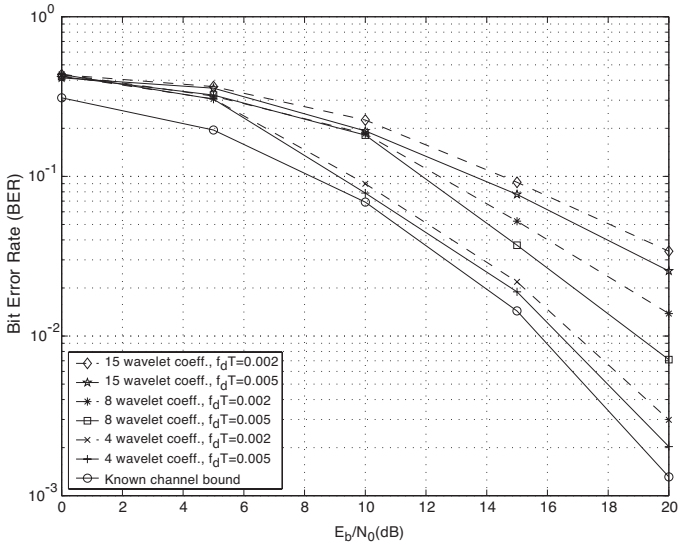

Fig. 8. BER performance of the SMC receiver using fixed number (e.g., 4, 8, and 15) of wavelet coefficients under different Dopplers. The channel is frequency-selective fading with $L = 2$. The delayed-sample method is used with $\Delta = 2$.

B. Performance of Wavelet-Based SMC in Frequency-Selective Fading Channels

We show the BER performance of the proposed SMC receiver in frequency-selective fading channels. The number of paths is $L = 2$. The delayed-sample method is used with $\Delta = 2$. In Fig. 8, the BER performance is plotted for different number of wavelet coefficients and different Dopplers. It is seen that in this case, the best performance is achieved using four wavelet coefficients whereas a loss of nearly 5 and 10 dB, respectively, is incurred when eight and 15 wavelet coefficients are used. The BER performance of the resampling-based adaptive wavelet shrinkage SMC receiver with delayed-sample and delayed-weight estimation is plotted in Fig. 9. As in flat-fading channels, the SMC blind receiver achieves close-to-bound performance in channels with moderate fading rates (e.g., $f_dT < 0.01$), whereas the performance degrades substantially in very fast fading channels (e.g., $f_dT \geq 0.01$). However, it is seen that the BER performance (see Fig. 8) of the adaptive receiver is little bit worse (about 0.4 dB) than the BER performance (see Fig. 9) of the receiver using fixed four wavelet coefficients with very slow fading channels (e.g., $f_dT = 0.002$ or 0.005). The reason for this is that the slow fading channels can be characterized by only a few wavelet coefficients, in particular, about four wavelet coefficients for channels with normalized Doppler frequency $f_dT = 0.002$ or 0.005. Hence, a better BER performance shown in Fig. 8 is achieved with four wavelet coefficients. On the other hand, the resampling-based adaptive wavelet shrinkage SMC receiver takes a short time to dynamically choose the suitable number of wavelet coefficients, which results in a little bit worse BER performance. However, it is clearly seen that the close-to-bound performance is usually achieved by the adaptive SMC blind receiver without the prior information of the channels. In summary, the number of coefficients requirements is unknown *a priori*, and the receiver must be able to estimate it. For this particular example, by varying the number of coefficients, we found that four coefficients gives the best performance, and the adaptive shrinkage SMC receiver achieves a similar performance. However, again, the order is not known *a priori* and must be estimated.

VI. Conclusions

We have developed a new nonparametric Bayesian receiver technique for blind detection in fading channels with unknown channel statistics. It is based on the wavelet modeling of the fading process and the sequential Monte Carlo method for online Bayesian inference. A novel resampling-based wavelet shrinkage technique is developed to dynamically choose the number of the wavelet coefficients to best fit the fading

process. Blind adaptive receivers for both flat-fading channels and frequency-selective fading channels are developed, and their performance is demonstrated via computer simulations. It is seen that for moderate fading rate (e.g., $f_dT < 0.01$), the performance of the proposed wavelet based SMC receivers is close to that of the receiver with perfect channel state information in both flat-fading and frequency-selective fading channels. Our simulation results also show that the performance of the nonparametric SMC receiver is nearly the same for higher order Daubechies filter. Nonparametric Bayesian blind adaptive receivers for very fast fading channels is still an open problem and will be investigated in the future. Finally, we note that the proposed nonparametric SMC receiver can be extended to coded systems in the same way as described in [5] and [36].

REFERENCES

[1] C. Andrieu, J. F. G. deFreitas, and A. Doucet, "Robust full Bayesian learning for radial basis networks," *Neural Comput.*, vol. 13, no. 10, pp. 2359–2407, Oct. 2001.

[2] E. Baccarelli and R. Cusani, "Combined channel estimation and data detection using soft statistics for frequency-selective fast fading digital links," *IEEE Trans. Commun.*, vol. 46, pp. 424–427, Apr. 1998.

[3] E. Beadle and P. Djuric, "A fast-weighted bayesian bootstrap filter for nonlinear model state estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 338–343, Jan. 1997.

[4] R. Chen and J. S. Liu, "Mixture Kalman filters," *J. R. Stat. Soc. B*, vol. 62, pp. 493–509, 2000.

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

[6] C.-M.Chao-Ming Cho and P. Djuric, "Bayesian detection and estimation of cisoids in colored noise," *IEEE Trans. Signal Processing*, vol. 43, pp. 2943–2952, Dec. 1995.

[7] M. Clyde and E. George, "Flexible empirical Bayes estimation for wavelets," *J. R. Stat. Soc. B*, vol. 62, pp. 681–698, 2000.

[8] I. B. Collings and J. B. Moore, "An adaptive hidden Markov model approach to FM and M-ary DPSK demodulation in noisy fading channels," *Signal Process.*, vol. 47, pp. 71–84, 1995.

[9] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency-selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 941–950, Feb./Mar./Apr. 1994.

[10] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[11] L. M. Davis, I. B. Collings, and R. J. Evans, "Coupled estimators for equalization of fast-fading mobile channels," *IEEE Trans. Commun.*, vol. 46, pp. 1262–1265, Oct. 1998.

[12] P. M. Djuric and J. Chun, "An MCMC sampling approach to estimation of nonstationary hidden markov models," *IEEE Trans. Signal Processing*, vol. 50, pp. 1113–1123, May 2002.

[13] D. Donoho and I. Hohnstone, "Multiple shrinkage and subset selection in wavelets," *Biometrica*, vol. 81, pp. 425–455, 1994.

[14] A. Doucet, J. F. G. deFreitas, and N. Gordon, *Sequential Monte Carlo in Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[15] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential simulation-based methods for Bayesian filtering," *Stat. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.

[16] W. Fong, S. J. Godsill, M. West, and A. Doucet, "Monte carlo smoothing with application to audio signal enhancement," *IEEE Trans. Signal Processing*, vol. 50, pp. 438–448, Feb. 2002.

[17] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, pp. 300–308, Mar. 1997.

[18] G. B. Giannakis and C. Tepedelenliglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.

[19] R. Haeb and H. Meyr, "A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels," *IEEE Trans. Commun.*, vol. 37, pp. 748–754, July 1989.

[20] Y. Huang and P. Djuric, "Multiuser detection of synchronous code-division multiple-access signals by perfect sampling," *IEEE Trans. Signal Processing*, vol. 50, pp. 1724–1734, July 2002.

[21] G. T. Irvine and P. J. McLane, "Symbol-aided plus decision-directed reception for PSK/TCM modulation on shadowed mobile satellite fading channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 1289–1299, Oct. 1992.

[22] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Statist. Assoc.*, vol. 89, pp. 278–288, 1994.

[23] H. Kong and E. Shwedyk, "On channel estimation and sequence detection of interleaved coded signals over frequency nonselective Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 558–565, May 1998.

[24] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Stat. Assoc.*, vol. 93, pp. 1032–1044, 1998.

[25] Y. Liu and S. D. Blostein, "Identification of frequency nonselective fading channels using decision feedback and adaptive linear prediction," *IEEE Trans. Commun.*, vol. 43, pp. 1484–1492, Feb./Mar./Apr. 1995.

[26] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN: A combined envelop, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, pp. 63–75, Jan. 1994.

[27] M. Martone, "Optimally regularized channel tracking techniques for sequence estimation based on cross-validated subspace signal processing," *IEEE Trans. Commun.*, vol. 48, pp. 95–105, Jan. 2000.

[28] ——, "Wavelet-based separating kernels for array processing of cellular DS/CDMA signals in fast fading," *IEEE Trans. Commun.*, vol. 48, pp. 979–995, June 2000.

[29] S. Prakriya and D. Hatzinakos, "Blind identification of LTI-ZMNL-LTI nonlinear channel models," *IEEE Trans. Signal Processing*, vol. 43, pp. 3007–3013, Dec. 1995.

[30] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[31] N. Shephard and M. K. Pitt, "Likelihood analysis of nongaussian measurement time series," *Biometrika*, vol. 84, pp. 653–667, 1997.

[32] G. L. Stüber, *Priniciples of Mobile Communication*. Boston, MA: Kluwer, 2000.

[33] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Process.*, vol. 53, no. 2–3, pp. 211–229, Sept. 1996.

[34] T. Vaidis and C. L. Weber, "Block adaptive techniques for channel identification and data demodulation over band-limited channels," *IEEE Trans. Commun.*, vol. 46, pp. 232–243, Feb. 1998.

[35] G. M. Vitetta and D. P. Taylor, "Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 2750–2758, Nov. 1995.

[36] X. Wang, R. Chen, and D. Guo, "Delayed-pilot sampling for mixture Kalman filter with application in fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 241–254, Feb. 2002.

Dong Guo received the B.S. degree in geophysics and computer science from China University of Mining and Technology (CUMT), Xuzhou, China, in 1993, the M.S. degree in geophysics from the Graduate School of Research Institute of Petroleum Exploration and Development (RIPED), Beijing, China, in 1996. In 1999, he received the Ph.D. degree in applied mathematics from Beijing University. He is now pursuing a second Ph.D. degree with the Department of Electrical Engineering, Columbia University, New York, NY.

His research interests are in the area of statistical signal processing and communications.

Xiaodong Wang (M'98) received the B.S. degree in electrical engineering and applied mathematics (with the highest honor) from Shanghai Jiao Tong University, Shanghai, China, in 1992, the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1995, and the Ph.D degree in electrical engineering from Princeton University, Princeton, NJ, in 1998.

From July 1998 to December 2001, he was an Assistant Professor with the Department of Electrical Engineering, Texas A&M University, College Station. In January 2002, he joined the Department of Electrical Engineering, Columbia University, New York, NY, as an Assistant Professor. His research interests fall in the general areas of computing, signal processing, and communications. He has worked in the areas of digital communications, digital signal processing, parallel and distributed computing, nanoelectronics, and bioinformatics and has published extensively in these areas. His current research interests include wireless communications, Monte Carlo-based statistical signal processing, and genomic signal processing.

Dr. Wang received the 1999 NSF CAREER Award and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *EURASIP Journal of Applied Signal Processing*.

Rong Chen received the B.S. degree in mathematics from Peking University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University, Pittsburgh, PA in 1987 and 1990, respectively.

He is a Professor with the Department of Information and Decision Sciences, College of Business Administration, University of Illinois (UIC), Chicago. Before joining UIC in 1999, he was with Department of Statistics, Texas A&M University, College Station. His main research interests are in time series analysis, statistical computing and Monte Carlo methods in dynamic systems, and statistical applications in engineering and business. He is an Associate Editor for the *Journal of American Statistical Association*, the *Journal of Business and Economic Statistics*, *Statistica Sinica*, and *Computational Statistics*.