

Joint Estimation and Decoding of Space-Time Trellis Codes

Jianqiu Zhang

Department of Electrical and Computer Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794-2350, USA
Email: jizzhang@ece.sunysb.edu

Petar M. Djurić

Department of Electrical and Computer Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794-2350, USA
Email: djuric@ece.sunysb.edu

Received 31 May 2001 and in revised form 22 January 2002

We explore the possibility of using an emerging tool in statistical signal processing, sequential importance sampling (SIS), for joint estimation and decoding of space-time trellis codes (STTC). First, we provide background on SIS, and then we discuss its application to space-time trellis code (STTC) systems. It is shown through simulations that SIS is suitable for joint estimation and decoding of STTC with time-varying flat-fading channels when phase ambiguity is avoided. We used a design criterion for STTCs and temporally correlated channels that combats phase ambiguity without pilot signaling. We have shown by simulations that the design is valid.

Keywords and phrases: space-time coding, sequential importance sampling, Kalman filtering, phase ambiguity, design criterion.

1. INTRODUCTION

Space-time coding (STC) originally introduced by Foschini [1] and further developed by Tarokh et al. in [2] provides a framework for exploiting spatial and temporal diversity to increase data rate in wireless communications. A general introduction of space-time coding can be found in [3]. Among families of space-time codes, we are interested in STTCs because of their many advantages over block STCs as pointed out in [2].

It is generally assumed that STCs will be used in fading environments, and therefore in decoding it is necessary to obtain channel state information (CSI), that is, estimate of the fading coefficients of the channel. In the literature, most of the time it is assumed that CSI becomes available through sending pilot sequences periodically from the transmit to the receive side. When CSI is not available, we have to estimate the channel, and that presents many challenges because the wireless channel may vary with time or frequency or both, and in systems with multiple transmit and receive antennas, many channels need to be estimated. Some design efforts have been directed to the construction of STCs which allow for circumventing channel estimation such as the unitary ST-modulation scheme discussed in [4] or differential space-time modulation considered in [5]. An interesting Kalman

filter space-time algorithm for joint estimation and decoding of Alamouti's block ST code [6] was proposed in [7], where Kalman filtering is used for tracking the channel. We have not seen similar work for STTCs. Here we consider the problem of joint channel estimation and decoding of STTCs when no pilot signal is available and the channel is modeled as time-varying flat-fading channel. Because STTCs use complicated modulation schemes, the observation at the receive side may not be a linear function of the transmitted data. Also, the CSI is hidden in the observation, and it is hard to use conventional methods like the Kalman filter to solve this problem. In this paper, we address joint estimation and detection of STTC by the sequential importance sampling (SIS) method combined with Kalman filtering.

The following text is a brief summary of SIS. The statistical property of many parameters in communication problems, like the fading coefficients in flat-fading channels, channel impulse response in frequency selective fading channels, and the trellis coded user data can be characterized by hidden Markov processes (HMPs). Consequently, many systems, including STTC systems, can be described adequately by dynamic state space models (DSSMs) [8, 9], which use state equations for describing the dynamics of the HMPs $\{\mathbf{x}_t\}$ of interest, and observation equations for representing

the observations $\{\mathbf{y}_t\}$ as functions of $\{\mathbf{x}_t\}$. In the Bayesian framework, all information about the unknowns $\{\mathbf{x}_t\}$ is included in the posterior distribution, $p(\mathbf{x}_0, \dots, \mathbf{x}_T | \mathbf{y}_0, \dots, \mathbf{y}_T)$, up to time T , from which the marginalized posterior distribution of interest, $p(\mathbf{x}_t | \mathbf{y}_0, \dots, \mathbf{y}_T)$ for all $t \leq T$, can be derived. However, the posterior density is usually high dimensional, and it is hardly ever possible to find the analytical expression for the density function, let alone to evaluate it. Due to the properties of the HMP, the posterior distribution, $p(\mathbf{x}_0, \dots, \mathbf{x}_T | \mathbf{y}_0, \dots, \mathbf{y}_T)$, can be factored in many ways and each term after the factorization may be evaluated individually. For example, if the posterior density is written as

$$p(\mathbf{x}_0, \dots, \mathbf{x}_T | \mathbf{y}_0, \dots, \mathbf{y}_T) \propto p(\mathbf{x}_0, \dots, \mathbf{x}_{T-1} | \mathbf{y}_0, \dots, \mathbf{y}_{T-1}) p(\mathbf{y}_T | \mathbf{x}_{T-1}), \quad (1)$$

this factorization provides a way to evaluate the desired posterior density recursively. However, except in the few special cases like in the linear Gaussian DSSM when Kalman filtering can be applied, there is usually no analytical solution to evaluate it recursively. Recently, a variety of Monte Carlo sampling methods, like importance sampling, Markov Chain Monte Carlo (MCMC) including Gibbs sampling [10, 11, 12], have revived Bayesian signal processing with their ability to sample from highly complicated probability density functions and thereby allowing for the computation of complex high dimensional integrals. In the literature, SIS [13, 14, 15], was introduced for solving problems that can be described by DSSMs without the strict restrictions imposed by the Kalman filter. In communications, SIS methods have been used for blind detection of user data [16, 17, 18]. Because there are no mathematical approximations of the model, SIS methods are found to perform better than traditional methods like the Extended Kalman Filter (EKF) [19, 20] or the Gaussian Sum Filter (GSF) [21]. It should be noted, however, that the sampling based methods use approximations too in that they approximate relevant densities with samples from these densities. The good performance, versatility, and the potential for implementation in VLSI make SIS methods very attractive.

This paper is organized as follows: Section 2 introduces the basic concept of SIS, where many aspects of SIS are discussed. Section 3 elaborates the application of SIS to the joint estimation and decoding of STTCs. In Section 4, we analyze the problem of phase ambiguity and find that when the channel is temporally correlated, it is possible to design ST coding schemes that alleviate phase ambiguity. There we propose an ad hoc space-time coding scheme, which is effective in combating phase ambiguity. Simulations are presented in Section 5, and in Section 6, conclusions are drawn.

2. SEQUENTIAL IMPORTANCE SAMPLING

In this section, we provide some background of SIS and the important special case of mixture SIS and Kalman filtering, which is suitable for joint estimation and decoding of STTCs in a time-varying flat-fading environment.

The SIS algorithm is applicable to systems that can be described by DSSMs which consist of state and observation

equations, that is,

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t), \quad \mathbf{y}_t = g(\mathbf{x}_t, \mathbf{v}_t), \quad (2)$$

where $f(\cdot)$ is the state transition function, which defines how the state vector \mathbf{x}_t evolves with time based on the previous state vector \mathbf{x}_{t-1} and the input noise vector \mathbf{u}_t at time t , and $g(\cdot)$ is an observation function of \mathbf{x}_t and the observation noise vector \mathbf{v}_t . We can see that the state transition function defines a Markov process because, given the previous state vector \mathbf{x}_{t-1} , the current state vector is no longer dependent on other previous state vectors or observations. Note that the state vector is not observed directly, and therefore the model describes an HMP.

From Bayesian perspective, all information about the state vectors is contained in the full posterior density $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ from which the filtering density $p(\mathbf{x}_t | \mathbf{y}_{0:t})$ and the smoothing density $p(\mathbf{x}_t | \mathbf{y}_{0:T})$ for all $t < T$ can be obtained. Here, $\mathbf{x}_{0:t} = \{\mathbf{x}_0, \dots, \mathbf{x}_t\}$ is the set of all state vectors from time 0 to time t , and $\mathbf{y}_{0:t}$ is defined similarly. When the full posterior density is available, the expected value of any function of the state vectors, $\xi(\mathbf{x}_{0:t})$ can be evaluated by

$$E_{\mathbf{x}_{0:t} | \mathbf{y}_{0:t}}(\xi(\mathbf{x}_{0:t})) = \int \xi(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t}. \quad (3)$$

If a DSSM is linear and Gaussian, that is, both the state transition and the observation functions are linear and both noise vectors \mathbf{u}_t and \mathbf{v}_t are Gaussian, the full posterior density of interest, $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$, can be obtained using the Kalman filter. In other cases, if the DSSM is not linear and Gaussian, approximations to the model must be made before applying the Kalman filter. Because of these approximations, the performance of the Kalman filter is limited. On the other hand, SIS can be used in such cases as discussed in [13, 16]. With SIS, the goal is to generate samples from $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ that are used for computation of statistical expectations as in (3). However, it is generally hard to draw samples directly from the posterior distribution, $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$, and instead, the importance sampling method is employed. Samples, $\{\mathbf{x}_{0:t}^{(j)}\}_{j=1}^J$, where $j = 1, 2, \dots, J$ is a sample index, are generated from a proposal density function $\pi(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$, and subsequently they are weighted according to the real posterior distribution $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$. The weight of a particular sample is given by

$$w_t^{(j)} = \frac{p(\mathbf{x}_{0:t}^{(j)} | \mathbf{y}_{0:t})}{\pi(\mathbf{x}_{0:t}^{(j)} | \mathbf{y}_{0:t})}. \quad (4)$$

The expectation in (3) is then approximated using Monte Carlo integration, that is,

$$\hat{E}_{\mathbf{x}_{0:t} | \mathbf{y}_{0:t}}(\xi(\mathbf{x}_{0:t})) = \frac{1}{W_t} \sum_{j=1}^J \xi(\mathbf{x}_{0:t}^{(j)}) w_t^{(j)}, \quad (5)$$

where $W_t = \sum_{j=1}^J w_t^{(j)}$ is the sum of the unnormalized weights. In the literature, the weighted sample $(\mathbf{x}_{0:t}^{(j)}, w_t^{(j)})$ is also called a particle, and sometimes SIS is referred to as

particle filtering. Note that in this paper, $\mathbf{x}_t^{(j)}$ is called a sample element.

An interesting characteristic of the full posterior density function at time t is that it can be factored as a product of the full posterior density of the previous time instant $t - 1$ and an incremental term,

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) \propto p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}) \times p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}). \quad (6)$$

If the incremental term $p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})$ can be evaluated at each time instant, we can obtain the full posterior density recursively. If the importance function is chosen in the form of

$$\pi(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) = \pi(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})\pi(\mathbf{x}_t), \quad (7)$$

we can derive an SIS algorithm which allows us to draw samples and evaluate their weights recursively. Suppose that we have obtained the sample set $\{\mathbf{x}_{0:t-1}^{(j)}\}_{j=1}^J$ properly weighted with respect to $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})$ from previous recursions. Then at time t for the j th sample, we

- (1) draw a new sample element $\mathbf{x}_t^{(j)}$ from the incremental proposal density $\pi(\mathbf{x}_t)$,
- (2) append the new sample element $\mathbf{x}_t^{(j)}$ to $\mathbf{x}_{0:t-1}^{(j)}$ and obtain the new sample $\mathbf{x}_{0:t}^{(j)}$, and
- (3) evaluate the weight for the new sample $\mathbf{x}_{0:t}^{(j)}$ according to (4), that is,

$$\begin{aligned} w_t^{(j)} &= \frac{p(\mathbf{x}_{0:t}^{(j)}|\mathbf{y}_{0:t})}{\pi(\mathbf{x}_{0:t}^{(j)}|\mathbf{y}_{0:t})} \\ &= \frac{p(\mathbf{x}_{0:t-1}^{(j)}|\mathbf{y}_{0:t-1})p(\mathbf{y}_t|\mathbf{x}_t^{(j)})p(\mathbf{x}_t^{(j)}|\mathbf{x}_{0:t-1}^{(j)})}{\pi(\mathbf{x}_{0:t-1}^{(j)}|\mathbf{y}_{0:t-1})\pi(\mathbf{x}_t^{(j)})} \\ &= w_{t-1}^{(j)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(j)})p(\mathbf{x}_t^{(j)}|\mathbf{x}_{t-1}^{(j)})}{\pi(\mathbf{x}_t^{(j)})}. \end{aligned} \quad (8)$$

The same procedure is repeated for all J samples. The path that each sample takes is also called a trajectory and so there are totally J trajectories when running the SIS algorithm. Note that the incremental proposal density function $\pi(\mathbf{x}_t)$ is simply referred to as the proposal density function or importance function in the literature.

Next question of interest is the selection of the importance function, $\pi(\mathbf{x}_t)$. First of all, $\pi(\mathbf{x}_t)$ must have appropriate support. Second, we should be able to draw samples from $\pi(\mathbf{x}_t)$, and third, $\pi(\mathbf{x}_t)$ should be easily computable. There are many choices, two of which are the prior density function $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and the so-called ‘‘optimal’’ proposal function $p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})$. Intuitively, good proposal functions take into account all the available observations and past state vectors.

The optimal importance function is proportional to the product of two density functions,

$$p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (9)$$

and after obtaining the new sample element $\mathbf{x}_t^{(j)}$, its weight $w_{0:t}^{(j)}$ is assigned using (8)

$$\begin{aligned} w_t^{(j)} &= w_{t-1}^{(j)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(j)})p(\mathbf{x}_t^{(j)}|\mathbf{x}_{t-1}^{(j)})}{p(\mathbf{x}_t^{(j)}|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})} \\ &= w_{t-1}^{(j)} p(\mathbf{y}_t|\mathbf{x}_t^{(j)}). \end{aligned} \quad (10)$$

We can see that to compute $w_t^{(j)}$ from $w_{t-1}^{(j)}$, we need to evaluate an integral because

$$p(\mathbf{y}_t|\mathbf{x}_{t-1}^{(j)}) = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(j)})d\mathbf{x}_t. \quad (11)$$

If \mathbf{x}_t is defined on a finite discrete space, there would be no problem in computing (11) when the size of the space is small because the integration simply becomes a sum of small number of proposal densities. When \mathbf{x}_t is defined on an infinite discrete space or continuous space, there is in general no analytical solution to (11).

2.1. Mixture SIS and Kalman filtering

An important special case of SIS is when the state vector can be partitioned into two parts, where one of them is conditionally linear and Gaussian given the other part. For example, in wireless communications with flat fading, the observation is the product of the sent user signal and the fading coefficients embedded in additive noise. The state vector \mathbf{x}_t contains two parts $(\mathbf{h}_t, \mathbf{s}_t)$, where \mathbf{h}_t is the state vector of the fading coefficients (channel state vector) and \mathbf{s}_t is the user state vector. The user state vector and channel state vector can all be modeled as HMPs, that is,

$$\mathbf{s}_t = z(\mathbf{s}_{t-1}, d_t), \quad (12)$$

$$\mathbf{h}_t = \mathbf{F}\mathbf{h}_{t-1} + \mathbf{E}\mathbf{u}_t, \quad (13)$$

where d_t is the user data input at time t , $z(\cdot)$ is the state transition function of the user state vector, and \mathbf{E} is a known matrix. The observation equation can be written as

$$\mathbf{y}_t = \mathbf{C}(\mathbf{s}_t)\mathbf{R}\mathbf{h}_t + \mathbf{v}_t, \quad (14)$$

where $\mathbf{C}(\cdot)$ is a matrix coding and modulation function of the communication system and \mathbf{R} is another known matrix of model coefficients (their meaning will be explained in the sequel). Given \mathbf{s}_t , the DSSM is linear and Gaussian in \mathbf{h}_t , that is, both noise vectors \mathbf{u}_t and \mathbf{v}_t are complex Gaussian and the state transition and observation functions are linear and are described by the state transition matrix \mathbf{F} and matrix \mathbf{R} . Note that in most cases, one can find the unique inverse function of the state transition function of the user state vector, $d_t = z^{-1}(\mathbf{s}_t, \mathbf{s}_{t-1})$, and then the inference of $\mathbf{d}_{0:t}$ can be derived from $\mathbf{s}_{0:t}$. Here the posterior density of interest is $p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$, and when applying the SIS algorithm, the optimal proposal density for the j th sample is $p(\mathbf{s}_t|\mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})$, which can be expanded according to

$$\begin{aligned}
p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}) &\propto p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}) p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}) \\
&\propto \left[\int p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{h}_t) p(\mathbf{h}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}) d\mathbf{h}_t \right] \\
&\quad \times p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}).
\end{aligned} \tag{15}$$

Given all the past user data $\mathbf{s}_{0:t-1}^{(j)}$, the DSSM defined by (13) and (14) becomes linear Gaussian for the channel state vector \mathbf{h}_t , and therefore the density function $p(\mathbf{h}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ can be evaluated using the predictive step of Kalman filtering in terms of its mean and variance. In turn, because $p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{h}_t)$ is a Gaussian function, the integration in (15) can be carried out analytically. In essence, when the state vector is composed of two parts, we integrate out the part that we are not interested in. This method is referred to as Rao-Blackwellization in [22].

Based on this line of reasoning, a mixture of SIS and Kalman filtering algorithm was developed in [16]. Suppose that at time instant t for the j th trajectory, we have obtained from previous recursions via Kalman filtering and SIS the posterior estimate of the channel state vector $p(\mathbf{h}_{t-1} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ and $\mathbf{s}_{0:t-1}^{(j)}$ weighted according to the posterior distribution $p(\mathbf{s}_{0:t-1} | \mathbf{y}_{0:t-1})$. The following steps are then performed:

- (1) predict the current channel state vector using the Kalman filter based on the previous sample and all available observations, that is, find $p(\mathbf{h}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ based on $p(\mathbf{h}_{t-1} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$;
- (2) sample from the posterior distribution $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t})$ using importance sampling. The optimal importance function $p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})$ is given by (15);
- (3) the weight of the drawn samples is computed by (11), which involves integration. Note that, the transmitted user signals belong to a finite signal set \mathcal{A} and the integration becomes a summation, or

$$w_t^{(j)} = w_{t-1}^{(j)} \sum_{\mathbf{s}_t} p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}); \tag{16}$$

- (4) compute the posterior probability of the channel state vector $p(\mathbf{h}_t | \mathbf{s}_{0:t}^{(j)}, \mathbf{y}_{0:t})$ using the Kalman filter.

2.2. Delayed estimates

When there exists channel coding or Inter Symbol Interference (ISI), future observations, too, contain information about the current state vector. In such cases, a delayed estimate will be more accurate. There are two methods pertaining to delayed estimation, one of which is the delayed weight method and the other, the delayed sample method [16]. In the following discussion, first we assume that the only unknown portion of the state vector is the user state vector, which belongs to a finite discrete space, that is, $\mathbf{s}_t \in \mathcal{A}$.

The distribution function of interest of the delayed weight method is $p(\mathbf{s}_t | \mathbf{y}_{0:t+q})$. It can be obtained through

marginalizing the full delayed posterior distribution, $p(\mathbf{s}_{0:t+q} | \mathbf{y}_{0:t+q})$. If at time $t + q$, we have obtained samples $\{\mathbf{s}_{0:t+q}^{(j)}\}_{j=1}^J$ appropriately weighted with weights $\{w_{t+q}^{(j)}\}_{j=1}^J$, the full posterior distribution at $t + q$ is approximated by,

$$\hat{p}(\mathbf{s}_{0:t+q} | \mathbf{y}_{0:t+q}) = \sum_{j=1}^J w_{t+q}^{(j)} I_{\mathbf{s}_{0:t+q}^{(j)}}(\mathbf{s}_{0:t+q}), \tag{17}$$

where $I(\cdot)$ is an indicator function which equals to 1 when $\mathbf{s}_{0:t+q}^{(j)} = \mathbf{s}_{0:t+q}$. Similarly, the density function of interest can be approximated as

$$\hat{p}(\mathbf{s}_t | \mathbf{y}_{0:t+q}) = \sum_{j=1}^J w_{t+q}^{(j)} I_{\mathbf{s}_{0:t+q}^{(j)}}(\mathbf{s}_t). \tag{18}$$

The second method of delayed estimation comes from the observation that we can improve sampling efficiency by incorporating future observations to the proposal distribution, which becomes $p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q})$. This proposal distribution can be expanded according to

$$\begin{aligned}
p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q}) &\propto p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}) p(\mathbf{y}_{t:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{s}_t, \mathbf{y}_{0:t-1}) \\
&= p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}) \sum_{\mathbf{s}_{t+1:t+q} \in \mathcal{A}^q} p(\mathbf{y}_{t:t+q} | \mathbf{s}_{t+1:t+q}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}) \\
&\quad \times p(\mathbf{s}_{t+1:t+q} | \mathbf{s}_t) \\
&= p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}) p(\mathbf{y}_t | \mathbf{s}_t) \\
&\quad \times \sum_{\mathbf{s}_{t+1:t+q} \in \mathcal{A}^q} \left[\prod_{l=1}^q p(\mathbf{y}_{t+l} | \mathbf{s}_{t+l}) p(\mathbf{s}_{t+l} | \mathbf{s}_{t+l-1}) \right].
\end{aligned} \tag{19}$$

The evaluation of this proposal density is quite complex because in essence, the algorithm is propagated for all possible future state vectors from $t + 1$ to $t + q$ and then marginalized for \mathbf{s}_t . If the state vectors are continuous, the summation will become integration in (19) and in most cases there will be no analytical solution.

The delayed proposal distribution also needs to be weighted according to the desired posterior distribution $p(\mathbf{s}_t | \mathbf{y}_{0:t+q})$, or

$$\begin{aligned}
w_t^{(j)} &= \frac{p(\mathbf{s}_{0:t}^{(j)} | \mathbf{y}_{0:t+q})}{p(\mathbf{s}_{0:t-1}^{(j)} | \mathbf{y}_{0:t+q-1}) p(\mathbf{s}_t^{(j)} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q})}
\end{aligned} \tag{17}$$

$$\begin{aligned}
&= w_{t-1}^{(j)} \frac{\sum_{\mathbf{s}_{t:t+q} \in \mathcal{A}^{q+1}} p(\mathbf{y}_{t:t+q} | \mathbf{s}_{t:t+q}) p(\mathbf{s}_{t:t+q} | \mathbf{s}_{0:t-1}^{(j)})}{\sum_{\mathbf{s}_{t:t+q-1} \in \mathcal{A}^q} p(\mathbf{y}_{t:t+q-1} | \mathbf{s}_{t:t+q-1}) p(\mathbf{s}_{t:t+q-1} | \mathbf{s}_{0:t-1}^{(j)})},
\end{aligned} \tag{18}$$

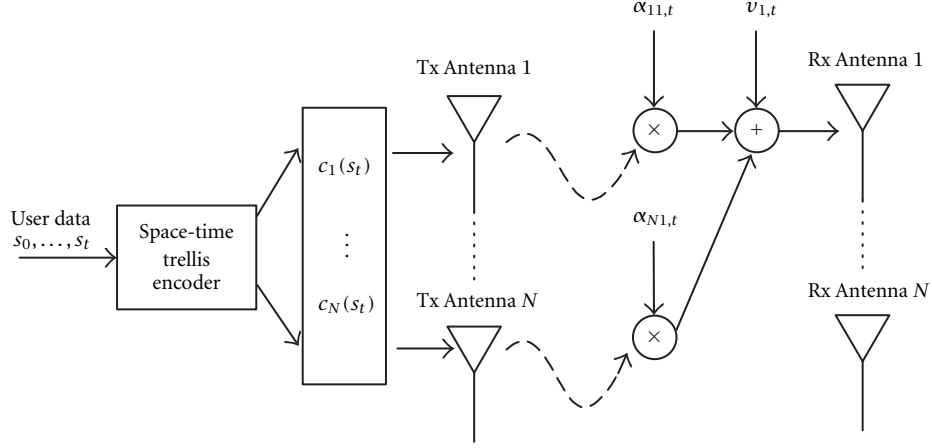


FIGURE 1: STTC system.

where the denominator can be expanded as

$$\begin{aligned} & \sum_{\mathbf{s}_{t:t+q-1} \in \mathcal{A}^q} p(\mathbf{y}_{t:t+q-1} | \mathbf{s}_{0:t+q-1}) p(\mathbf{s}_{t:t+q-1} | \mathbf{s}_{0:t-1}^{(j)}) \\ &= \sum_{\mathbf{s}_{t:t+q-1} \in \mathcal{A}^q} \left[\prod_{l=1}^{q-1} p(\mathbf{y}_{t+l} | \mathbf{s}_{t+l}) p(\mathbf{s}_{t+l} | \mathbf{s}_{t+l-1}) \right] \\ & \quad \times p(\mathbf{y}_t | \mathbf{s}_t) p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}). \end{aligned} \quad (21)$$

Now we assume that the channel state vector is also unknown and the two parts of the state vector are conditionally Gaussian. Then the proposal distribution for the delayed sampling method can be derived similarly as the one of the non-delay case, and it results in

$$p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q}) = \sum_{\mathbf{s}_{t+1:t+q} \in \mathcal{A}^q} p(\mathbf{s}_t, \mathbf{s}_{t+1:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q}). \quad (22)$$

2 The term inside the summation in (22) can be further derived as

$$\begin{aligned} & p(\mathbf{s}_t, \mathbf{s}_{t+1:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q}) \\ & \propto p(\mathbf{y}_{t+q} | \mathbf{s}_{t+1:t+q}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q-1}) \\ & \quad \times p(\mathbf{s}_t, \mathbf{s}_{t+1:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q-1}) \\ &= \left[\int p(\mathbf{y}_{t+q} | \mathbf{s}_{t+q}, \mathbf{h}_{t+q}) \right. \\ & \quad \times p(\mathbf{h}_{t+q} | \mathbf{s}_{t+1:t+q}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q-1}) d\mathbf{h}_{t+q} \left. \right] \\ & \quad \times p(\mathbf{s}_t, \mathbf{s}_{t+1:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+q-1}) \\ &= \prod_{l=0}^q \left[\int p(\mathbf{y}_{t+l} | \mathbf{s}_{t+l}, \mathbf{h}_{t+l}) \right. \\ & \quad \times p(\mathbf{h}_{t+l} | \mathbf{s}_{t+1:t+l}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+l-1}) d\mathbf{h}_{t+l} \left. \right] \\ & \quad \times p(\mathbf{s}_t, \mathbf{s}_{t+1:t+q} | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}). \end{aligned} \quad (23)$$

The probability density function of the channel state

$p(\mathbf{h}_{t+l} | \mathbf{s}_{t+1:t+l}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+l-1})$ can be obtained using the prediction step of Kalman filtering if the channel state is Gaussian. The prediction can give the mean and covariance matrix of the channel state, and the integral

$$\int p(\mathbf{y}_{t+l} | \mathbf{s}_{t+l}, \mathbf{h}_{t+l}) p(\mathbf{h}_{t+l} | \mathbf{s}_{t+1:t+l}, \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t+l-1}) d\mathbf{h}_{t+l} \quad (24)$$

can be readily evaluated. The weights of the samples are obtained similarly as in (20). Based on these derivations, an algorithm was developed for the online estimation of user data using the delayed sample method when the state vector can be partitioned into two parts. The details of the algorithm can be found in [16]. The complexity of the algorithm is proportional to the size of the set \mathcal{A}^q . If the size of the set is moderate, the algorithm is still practical; otherwise, the algorithm is just too computationally intensive to be applied.

3. APPLICATION OF SIS ON JOINT ESTIMATION AND DECODING OF STTCs

In Section 2, we showed that the SIS algorithm can be used to systems described by DSSMs, and when the state vector consists of two parts, that the mixture SIS and Kalman filtering algorithm can be applied. Therefore, to use the SIS algorithm, first we try to represent the STTC system as a DSSM.

Suppose that a communication system employs N transmit and M receive antennas as in Figure 1. A sequence of user data symbols, s_0, \dots, s_t , where $s_t \in \mathcal{A}$, is put through a trellis space-time encoder. The new state vector of the trellis space-time encoder at time t is determined according to the state transition equation $\mathbf{s}_t = \mathbf{z}(\mathbf{s}_{t-1}, s_t)$, where \mathbf{s}_{t-1} is the previous state, and \mathbf{s}_t is the new user state. For example, in the case of delayed diversity STTC, the state vector simply consists of the last two user symbols, $\mathbf{s}_t^T = [s_t, s_{t-1}]$. Based on the current state vector, the space-time encoder then generates a set of N symbols, $\mathbf{c}^T(\mathbf{s}_t) = [c_1(\mathbf{s}_t), \dots, c_N(\mathbf{s}_t)]$, to be transmitted by N antennas where $c_i(\cdot)$ denotes the code and modulation function of the i th antenna. Suppose $\alpha_{nm,t}$ is the fading coefficient from the n th transmit antenna to the m th

receive antenna at time t . When assuming ideal timing and frequency information, the received signal in the flat fading environment at the m th antenna can be written as

$$y_{m,t} = \sum_{n=1}^N \alpha_{nm,t} c_n(\mathbf{s}_t) + v_{m,t}, \quad m = 1, 2, \dots, M, \quad (25)$$

where $v_{m,t} \sim \mathcal{N}_c(0, \sigma_{v_m}^2)$ is a complex Gaussian process present at the m th receive antenna. Let $\boldsymbol{\alpha}_{m,t} = [\alpha_{1m,t} \cdots \alpha_{Nm,t}]^T$ represent the set of channel states from all transmit antennas to the m th receive antenna. If we arrange all the channel states at time t into a single $NM \times 1$ vector $\boldsymbol{\alpha}_t = [\boldsymbol{\alpha}_{1,t} \cdots \boldsymbol{\alpha}_{M,t}]$, all the received signals at time t can be written in the vector form as

$$\mathbf{y}_t = \mathbf{C}(\mathbf{s}_t) \boldsymbol{\alpha}_t + \mathbf{v}_t, \quad (26)$$

where $\mathbf{y}_t = [y_{1,t} \cdots y_{M,t}]^T$ is the received signal vector, and $\mathbf{v}_t = [v_{1,t} \cdots v_{M,t}]^T$ is the noise vector. The code and modulation matrix is an $M \times NM$ matrix of the form

$$\mathbf{C}^T(\mathbf{s}_t) = \begin{bmatrix} \mathbf{c}(\mathbf{s}_t) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}(\mathbf{s}_t) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{c}(\mathbf{s}_t) \end{bmatrix}. \quad (27)$$

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

The fading coefficients from the n th to the m th antenna can be modeled as an ARMA process to match the power spectral density of the channel [16]. An ARMA(r_1, r_2) process can be represented as

$$\begin{aligned} \alpha_{nm,t} + \phi_1 \alpha_{nm,t-1} + \cdots + \phi_{r_1} \alpha_{nm,t-r_1} \\ = \rho_0 u_{nm,t} + \cdots + \rho_{r_2} u_{nm,t-r_2}, \end{aligned} \quad (28)$$

where $u_{nm,t}$ is a unit variance i.i.d. complex Gaussian process that drives the ARMA process, and $\{\phi_i\}$ and $\{\rho_i\}$ are known auto-regressive (AR) and moving-average (MA) coefficients. In order to represent the ARMA process as a DSSM, we assume equal orders of the AR and MA parts of the model, that is $r_1 = r_2 = r$. Otherwise, we make the orders equal with zero padding. The ARMA process can be described as a DSSM if we introduce $\mathbf{h}_{nm,t} = [h_{nm,t}, \dots, h_{nm,t-r}]^T$ as the state vector of dimension $r+1$ for the AR part of the ARMA model (see the Appendix for more detailed information). The state transition equation is in the form of

$$\mathbf{h}_{nm,t} = \boldsymbol{\Phi} \mathbf{h}_{nm,t-1} + \mathbf{e} u_{nm,t}, \quad (29)$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix} -\phi_1 & -\phi_2 & \cdots & -\phi_r & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (30)$$

The observation equation of the ARMA DSSM is $\alpha_{nm,t} = \boldsymbol{\rho}^T \mathbf{h}_{nm,t}$, where $\boldsymbol{\rho}$ is a vector of known MA coefficients, that is, $\boldsymbol{\rho}^T = [\rho_0, \rho_1, \dots, \rho_r]$.

Considering all the fading coefficients involved, one can represent the whole system in a compact DSSM form given by

$$\mathbf{h}_t = \mathbf{F} \mathbf{h}_{t-1} + \mathbf{E} \mathbf{u}_t, \quad \boldsymbol{\alpha}_t = \mathbf{R} \mathbf{h}_t, \quad (31)$$

where $\mathbf{h}_t = [\mathbf{h}_{11,t}, \dots, \mathbf{h}_{N1,t}, \dots, \mathbf{h}_{1M,t}, \dots, \mathbf{h}_{NM,t}]^T$ is the extended state vector. The other symbols, \mathbf{R} , \mathbf{F} , \mathbf{E} , in the above equations are defined similarly (see the Appendix).

Now the received signal vector can be written as

$$\mathbf{y}_t = \mathbf{C}(\mathbf{s}_t) \mathbf{R} \mathbf{h}_t + \mathbf{v}_t. \quad (32)$$

3.1. Mixture Kalman filtering and sequential importance sampling for STTC

We have represented the STTC system by a DSSM, and we notice that the state vector consists of two parts. One part is the user state vector \mathbf{s}_t , and the other is the channel state vector \mathbf{h}_t , which is linear Gaussian given \mathbf{s}_t . Consequently, the mixture SIS and Kalman filtering algorithm can be applied. Define the set of all received signals at time t as $\mathbf{y}_{0:t} = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t\}$, and define $\mathbf{s}_{0:t}$ and $\mathbf{h}_{0:t}$ similarly. Our goal is to sample from the posterior distribution $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t})$. The algorithm for doing it is summarized as follows: for the recursion of the j th sample at the t th time instant:

- (1) sample from the optimal proposal distribution following (15). In case of the STTC, the proposal distribution for $\mathbf{s}_t = \mathbf{a}_i \in \mathcal{A}$ or equivalently $\mathbf{s}_t = z(\mathbf{a}_i, \mathbf{s}_{t-1}^{(j)})$ can be obtained by

$$\begin{aligned} \pi_{t, \mathbf{s}_t = z(\mathbf{a}_i, \mathbf{s}_{t-1}^{(j)})}^{(j)} &= p(\mathbf{s}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}) \\ &\propto \int p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{h}_t) p(\mathbf{h}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}) d\mathbf{h}_t \\ &\quad \times p(\mathbf{s}_t | \mathbf{s}_{t-1}^{(j)}). \end{aligned} \quad (33)$$

In the above equation, the term $p(\mathbf{h}_t | \mathbf{s}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}) \sim \mathcal{N}_c(\boldsymbol{\mu}_{\mathbf{h}_t}, \boldsymbol{\Sigma}_{\mathbf{h}_t})$ is calculated via the prediction step of Kalman filtering given the previous elements of the j th sample $\mathbf{s}_{0:t-1}$. Then we can find that the likelihood $p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{h}_t) \propto \mathcal{N}_c(\boldsymbol{\mu}_{\mathbf{y}_t}^{(j)}, \boldsymbol{\Sigma}_{\mathbf{y}_t}^{(j)})$ in terms of its mean and covariance matrix

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{y}_t}^{(j)} &= \mathbf{C}(\mathbf{s}_t) \mathbf{R} \boldsymbol{\mu}_{\mathbf{h}_t}, \\ \boldsymbol{\Sigma}_{\mathbf{y}_t}^{(j)} &= \mathbf{C}(\mathbf{s}_t) \mathbf{R} \boldsymbol{\Sigma}_{\mathbf{h}_t} \mathbf{R}^H \mathbf{C}(\mathbf{s}_t)^H + \mathbf{Q}_v. \end{aligned} \quad (34)$$

Here, \mathbf{Q}_v is the covariance matrix of the received noise vector of dimension $M \times M$. At the end of this step, we have the j th sample as $\mathbf{s}_{0:t}^{(j)}$ and $\pi_{t, \mathbf{s}_t = z(\mathbf{a}_i, \mathbf{s}_{t-1}^{(j)})}^{(j)}$ for $\mathbf{a}_i \in \mathcal{A}$;

- (2) compute the weight of the j th sample according to (16);

- (3) update the estimate of the channel state vector probability given the new observation \mathbf{y}_t and the new sample $\mathbf{s}_{0:t}^{(j)}$ to prepare for the next round of estimation using the Kalman filter. It is easy to verify that the posterior density of the channel state is proportional to the product of the likelihood and the predictive channel density

$$p(\mathbf{h}_t | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}) \propto p(\mathbf{y}_t | \mathbf{h}_t, \mathbf{s}_t) p(\mathbf{h}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}). \quad (35)$$

The predictive density of the channel state vector is Gaussian, that is, $\mathcal{N}_c(\boldsymbol{\mu}_{\mathbf{h}_t}^{(j)}, \boldsymbol{\Sigma}_{\mathbf{h}_t}^{(j)})$. The likelihood function is proportional to a Gaussian with mean and covariance given by (34), and the posterior density of the channel state vector is a complex Gaussian distribution with mean and covariance

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{h}_t}(\mathbf{s}_{0:t}^{(j)}) &= \boldsymbol{\mu}_{\mathbf{h}_t}^{(j)} + \mathbf{K}[\mathbf{y}_t - \mathbf{C}(\mathbf{s}_t^{(j)})\mathbf{R}\boldsymbol{\mu}_{\mathbf{h}_t}^{(j)}], \\ \boldsymbol{\Sigma}_{\mathbf{h}_t}(\mathbf{s}_{0:t}^{(j)}) &= \boldsymbol{\Sigma}_{\mathbf{h}_t}^{(j)} - \mathbf{K}\mathbf{C}(\mathbf{s}_t^{(j)})\mathbf{R}\boldsymbol{\Sigma}_{\mathbf{h}_t}^{(j)}, \end{aligned} \quad (36)$$

where

$$\begin{aligned} \mathbf{K} &= \boldsymbol{\Sigma}_{\mathbf{h}_t}^{(j)} \mathbf{R}^H \mathbf{C}(\mathbf{s}_t^{(j)})^H \\ &\times [\mathbf{C}(\mathbf{s}_t^{(j)})\mathbf{R}\boldsymbol{\Sigma}_{\mathbf{h}_t}^{(j)}\mathbf{R}^H \mathbf{C}(\mathbf{s}_t^{(j)})^H + \mathbf{Q}_v]^{-1} \end{aligned} \quad (37)$$

and $(\cdot)^H$ stands for Hermitian transpose. Note that $\mathbf{C}(\mathbf{s}_t^{(j)})$ need not be full rank to ensure the existence of the matrix inversion in (37). Equation (37) is equivalent to the Kalman filter equations and \mathbf{K} is analogous to the Kalman filter gain when the code matrix $\mathbf{C}(\mathbf{s}_t^{(j)})$ is known;

- (4) predict the channel state vector for the next time instant and calculate the mean and covariance matrix of the prediction. The mean vector and covariance matrix are

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{h}_{t+1}}(\mathbf{s}_{0:t}^{(j)}) &= \mathbf{F}\boldsymbol{\mu}_{\mathbf{h}_t}(\mathbf{s}_{0:t}^{(j)}), \\ \boldsymbol{\Sigma}_{\mathbf{h}_{t+1}}(\mathbf{s}_{0:t}^{(j)}) &= \mathbf{F}\boldsymbol{\Sigma}_{\mathbf{h}_t}(\mathbf{s}_{0:t}^{(j)})\mathbf{F}^H + \mathbf{E}\mathbf{Q}_u\mathbf{E}^H. \end{aligned} \quad (38)$$

Here \mathbf{Q}_u is the $NM \times NM$ covariance matrix of the complex Gaussian noise process that drives the ARMA process of the fading coefficients;

- (5) draw inference about the transmitted data. According to [16], the a posteriori symbol probability is estimated by

$$p(s_t = a_i | \mathbf{y}_{0:t}) \cong \frac{1}{W_t} \sum_j I_{s_t^{(j)}}(a_i) w_t^{(j)}, \quad a_i \in \mathcal{A}. \quad (39)$$

Note that in the algorithm described above, it is not required that the fading coefficients or the noise vector at the receive antennas be independent as required in most other algorithms.

Since the space-time code is trellis coded, besides the current and previous received signals $\mathbf{y}_{0:t}$, future observations

hold information about the current user state, hence it is appropriate to use the delayed importance function as well as the delay weight method in evaluating the posterior density function. The number of delays can be chosen according to the constraint length of the trellis code.

4. PHASE AMBIGUITY

The fading coefficients and the modulated user data are the unknowns to be estimated, and there may be multiple pairs of solutions to the same observation. For example, if $(\mathbf{C}(\mathbf{s}_t), \boldsymbol{\alpha}_t)$ is the true solution for the observation equation

$$\mathbf{y}_t = \mathbf{C}(\mathbf{s}_t)\boldsymbol{\alpha}_t + \mathbf{v}_t \quad (40)$$

a phase shifted version $(\tilde{\mathbf{C}}(\mathbf{s}_t), \tilde{\boldsymbol{\alpha}}_t) = (\mathbf{C}(\mathbf{s}_t)\boldsymbol{\Theta}, \boldsymbol{\Theta}^{-1}\boldsymbol{\alpha}_t)$ will be an equally likely solution, where $\boldsymbol{\Theta}$ is a phase shifting matrix. Such unfavorable condition is called phase ambiguity.

Let $\mathbf{C}_{0:T} = \{\mathbf{C}(\mathbf{s}_0), \dots, \mathbf{C}(\mathbf{s}_T)\}$ represent a sequence of ST coded vectors and let $\boldsymbol{\alpha}_{0:T} = \{\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_T\}$ be the sequence of channel state vectors from time 0 to T . Denote a phase shifted but legitimate code vector sequence as

$$\tilde{\mathbf{C}}_{0:T} = \{\mathbf{C}(\mathbf{s}_0)\boldsymbol{\Theta}_0, \dots, \mathbf{C}(\mathbf{s}_T)\boldsymbol{\Theta}_T\}. \quad (41)$$

Define $\boldsymbol{\Theta}_{0:T} = \{\boldsymbol{\Theta}_0, \dots, \boldsymbol{\Theta}_T\}$ as the phase difference sequence between the pair $(\mathbf{C}_{0:T}, \tilde{\mathbf{C}}_{0:T})$. Suppose the phase shifted code vector sequence and the accompanying phase shifted channel estimates $\tilde{\boldsymbol{\alpha}}_{0:T} = \{\boldsymbol{\Theta}_0^{-1}\boldsymbol{\alpha}_0, \dots, \boldsymbol{\Theta}_T^{-1}\boldsymbol{\alpha}_T\}$ have the same likelihood as the true code vector sequence and channel vector sequence, that is,

$$p(\mathbf{y}_{0:T} | \mathbf{C}_{0:T}, \boldsymbol{\alpha}_{0:T}) = p(\mathbf{y}_{0:T} | \tilde{\mathbf{C}}_{0:T}, \tilde{\boldsymbol{\alpha}}_{0:T}). \quad (42)$$

This obviously is a major problem that needs consideration. In the literature, unless unitary STC is employed, pilot signal is used for channel estimation. However, unitary STC or pilot signal are not necessarily the best strategies. To reduce phase ambiguity while providing better coding gain, here we opt to use an STTC designed using a new criterion proposed in [23] and summarized here.

It is desirable to design a set of code vector sequences so that with optimal joint estimation and decoding for any pair $(\mathbf{C}_{0:T}, \tilde{\mathbf{C}}_{0:T})$, the condition $p(\mathbf{C}_{0:T}, \boldsymbol{\alpha}_{0:T} | \mathbf{y}_{0:T}) > p(\tilde{\mathbf{C}}_{0:T}, \tilde{\boldsymbol{\alpha}}_{0:T} | \mathbf{y}_{0:T})$ is met. With the assumption in (42) and because the code vector sequences are independent of the channel vector sequences, the above condition can be simplified as

$$\begin{aligned} p(\mathbf{y}_{0:T} | \mathbf{C}_{0:T}, \boldsymbol{\alpha}_{0:T}) p(\mathbf{C}_{0:T}, \boldsymbol{\alpha}_{0:T}) &> p(\mathbf{y}_{0:T} | \tilde{\mathbf{C}}_{0:T}, \tilde{\boldsymbol{\alpha}}_{0:T}) p(\tilde{\mathbf{C}}_{0:T}, \tilde{\boldsymbol{\alpha}}_{0:T}) \\ \Leftrightarrow p(\mathbf{C}_{0:T}) p(\boldsymbol{\alpha}_{0:T}) &> p(\tilde{\mathbf{C}}_{0:T}) p(\tilde{\boldsymbol{\alpha}}_{0:T}) \\ \Leftrightarrow p(\boldsymbol{\alpha}_{0:T}) &> p(\tilde{\boldsymbol{\alpha}}_{0:T}). \end{aligned} \quad (43)$$

The last step of the derivation is based on the assumption that the user data sequences are i.i.d. and therefore the prior probability for each code vector sequence is the same. If the

channel state vectors are i.i.d. in time, then the condition in (43) cannot be met, the two sides are always equal to one another and there is nothing we can do in terms of ST design to solve the problem. On the other hand, if the channel state vector is represented by an HMP as we have described, then the condition can be further simplified using the chain rule,

$$\begin{aligned} \prod_{t=1}^T p(\mathbf{\alpha}_t | \mathbf{\alpha}_{t-1}) &> \prod_{t=1}^T p(\tilde{\mathbf{\alpha}}_t | \tilde{\mathbf{\alpha}}_{t-1}) \\ \Leftrightarrow \sum_{t=1}^T (\mathbf{h}_t - \mathbf{F}\mathbf{h}_{t-1})^H \mathbf{Q}_u (\mathbf{h}_t - \mathbf{F}\mathbf{h}_{t-1}) \\ &< \sum_{t=1}^T (\Theta_t^{-1} \mathbf{h}_t - \mathbf{F}\Theta_{t-1}^{-1} \mathbf{h}_{t-1})^H \mathbf{Q}_u (\Theta_t^{-1} \mathbf{h}_t - \mathbf{F}\Theta_{t-1}^{-1} \mathbf{h}_{t-1}). \end{aligned} \quad (44)$$

It can be verified that the only case when the above condition is not satisfied is when the phase difference sequence is homogeneous, that is, $\Theta_0 = \dots = \Theta_T$. Therefore, in designing the code vector sequence set, we should avoid the situation when the phase difference sequences are homogeneous between any pair of code vector sequences. If this cannot be avoided completely, one should try to alleviate frequent appearance of code vector sequences with homogeneous phase difference sequences.

Consider, for example, the 8PSK delay diversity STTC with two transmit antennas as described in [2]. We refer to this code as Tarokh STTC. The constraint length of the trellis is two and the codes for the two transmit antennas are

$$c_1(\mathbf{s}_t) = e^{-j(3\pi/4)s_t}, \quad c_2(\mathbf{s}_t) = e^{j(\pi/4)s_{t-1}}, \quad (45)$$

where $\mathbf{s}_t = [s_t, s_{t-1}]^T$, and each user data symbol $s_t \in \mathcal{A} = \{0, 1, \dots, 7\}$. The first two constellations in Figure 4 illustrate the two codes represented by (45). The phase shifting matrix in this case has two elements and it is given by $\Theta = \text{diag}\{\theta_1, \theta_2\}$. A phase shifted version of the channel state estimate will produce erroneous estimate of the transmitted user data. For example, if the two channel estimates over two consecutive time slots are phase shifted by π , then the estimate of s_t at t is $\hat{s}_{t,1} = c_1^{-1}(c_1(s_t) * e^{-j\pi})$, and at $t+1$, it is $\hat{s}_{t,2} = c_2^{-1}(c_2(s_t) * e^{-j\pi})$. It can be verified that for the code and modulation function $c_1(\cdot)$ and $c_2(\cdot)$ in (45), the two estimates of s_t are equal for all $s_t \in \mathcal{A}$. Therefore, the phase shifted channel estimate and \hat{s}_t are equally likely estimates of s_t . In summary, for this STTC the condition for homogeneous phase difference sequences to occur is equivalent to the following:

$$c_1^{-1}(c_1(s_t)e^{j\theta_1}) = c_2^{-1}(c_2(s_t)e^{j\theta_2}), \quad \forall \theta_1, \theta_2. \quad (46)$$

To address the problem of phase ambiguity, one may design space-time codes such that the occurrence of the condition specified in (46) is minimized given a certain coding gain. Define the degree of phase regularity as the number of pairs of user data which have the same phase difference among all the modulation constellations. For example, the phase difference of $s_t = 2$ and $s_t = 4$, is the same in constellations

c_1 and c_2 . We can find 12 pairs of such user data and the degree of phase regularity is 12. To reduce the occurrence of (46) is equivalent to reduce the degree of phase regularity. We can reduce phase regularity by redesigning the constellation and/or by increasing the number of constellations at the expense of coding, spatial, or temporal efficiency. The design of temporal/spatial varying modulation constellations that best explore the spatial and temporal diversities while balancing the need of reducing phase regularity is a very challenging task itself.

We consider the usage of pilot signal as a special case of the general idea for reducing phase regularity by using different constellations across different antennas or time slots. The use of pilot signal can be viewed as a special constellation, where all user data are modulated at one point, and the pilot signal is transmitted periodically.

5. SIMULATION

We simulated a two transmit and one receive antenna STTC system. The ARMA model for the fading coefficient chosen is as described in [16], and it is of order (3, 3). This model corresponds to a fast fading scenario with a normalized Doppler frequency (with respect to the symbol rate $1/T$) $f_d T = 0.05$.

We applied the mixture SIS and Kalman filtering algorithm to the Tarokh STTC system described in Section 4. Due to phase ambiguity, the algorithm invariably breaks down. Figure 2 displays the tracking of the complex fading coefficients where the dotted lines represent the true channel, and the solid lines show the estimated channel. Consequently, the estimated user symbols are erroneous under the phase shifted channel estimates. To alleviate phase ambiguity, first we tried the traditional way of sending pilot signals. To accommodate pilot signals, each time-slot was divided into two sub-slots (an increase in bandwidth), and it was considered that fading coefficients did not change between the sub-slots. The first set of sub-slots were assigned to transmit user signals coded by Tarokh STTC, and the second set of sub-slots were used to send pilot signals, that is, both transmit antennas sent known signals to the receiver. In our scheme, the pilot signals used the same amount of energy as the user signals. Pilot signals were inserted every other symbol because of the considered fast fading channel. The tracking of the channel is shown in Figure 3. We can see that phase ambiguity no longer exists. However, the performance of the algorithm in terms of symbol error rate (SER) is not satisfactory which is shown in Figure 6. In Section 4 we pointed out that if we design the code vector sequences so that the phase difference sequences are not homogeneous, phase ambiguity can be avoided. Here we propose 4 ad-hoc designed 8PSK constellations to be used when each time-slot is divided into two sub-slots as in the case with pilot signals. The received signal at time slot t can be described as

$$\begin{bmatrix} y_t^1 \\ y_t^2 \end{bmatrix} = \begin{bmatrix} c_1(s_t) & c_2(s_{t-1}) \\ c_3(s_{t-1}) & c_4(s_t) \end{bmatrix} \begin{bmatrix} \alpha_{11,t} \\ \alpha_{12,t} \end{bmatrix} + \begin{bmatrix} v_t^1 \\ v_t^2 \end{bmatrix}. \quad (47)$$

In the above equation, y_t^1 and y_t^2 represent the received

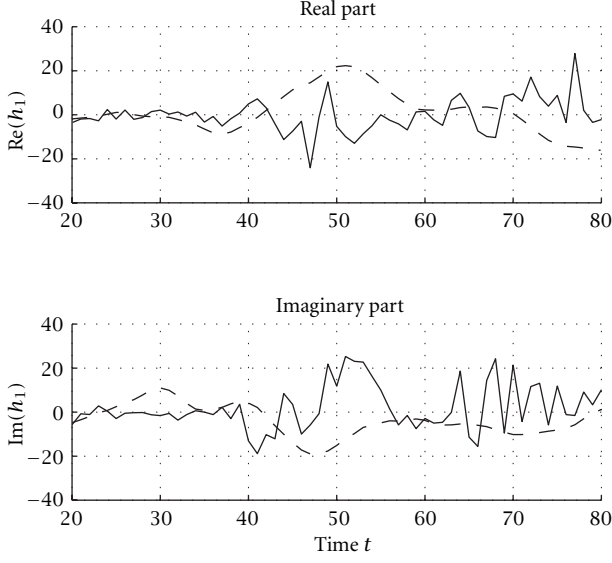


FIGURE 2: Example of phase ambiguity.

signals at two sub-time slots, where c_1 and c_2 have the same definition as in (45), and c_3 and c_4 are defined as

$$\begin{aligned} c_3 &= [1, e^{(-j\pi/4)}, -j, e^{(-j\pi 3/4)}, e^{(j\pi/4)}, j, e^{(j\pi 3/4)}, -1], \\ c_4 &= [1, e^{(j\pi 3/4)}, -j, e^{(j\pi/4)}, e^{(-j\pi 3/4)}, j, e^{(-j\pi/4)}, -1]. \end{aligned} \quad (48)$$

The constellations of c_1 to c_4 are shown in Figure 4. It can be verified that only two data pairs, (4, 6) and (5, 7), have the same phase difference across all four constellations. The degree of phase regularities is reduced to 2 as opposed to 12 in the original Tarokh STTC. Unless a long string of user data composed of only 4–7 occurs, which has very low probability when the user data symbols are i.i.d., phase ambiguity can be avoided. The tracking of the channel estimates is shown in Figure 5, which verifies our analysis of phase ambiguity. From Figure 6, we can see that the proposed system performs much better than the system using pilot signals although the two systems have the same time and frequency efficiency. In general, we have found that using pilot signals is not necessarily as efficient as using carefully designed space-time coding schemes.

It is well known that in the implementation of the SIS method, as the recursions proceed, the variance of the sample weights increases. This means that only a few samples have significant weights leading to poor estimates of the unknowns. This problem can be addressed by resampling [14], which is a process of redrawing a new set of samples from the old set where the variance of the new weights is smaller than the variance of the old weights. During our simulations, we found that we had to apply resampling frequently, which is essential for good performance of the algorithm. We used the same residual resampling process as described in [16] and resampling was performed every 5 steps. Because the constraint length is one in the STTC, the number of delayed weights and

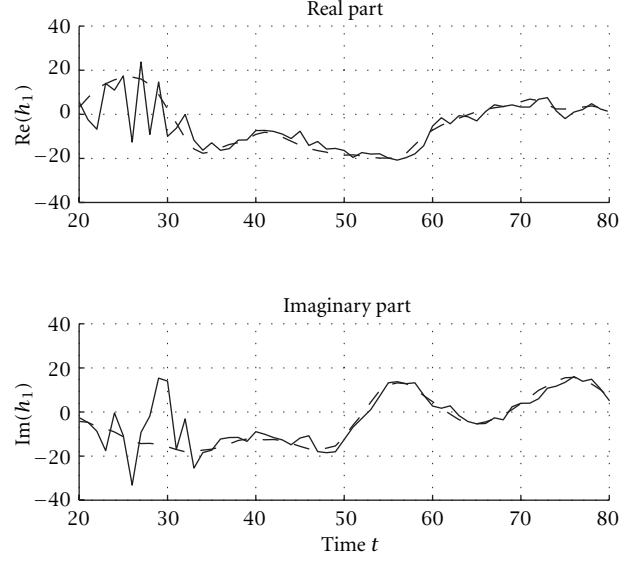


FIGURE 3: Channel tracking with pilot signals.

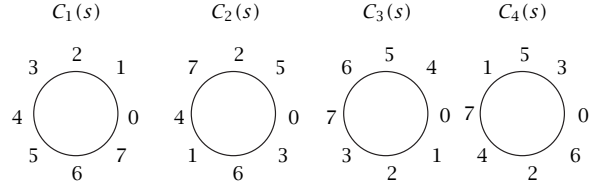


FIGURE 4: Constellations of the STTCs.

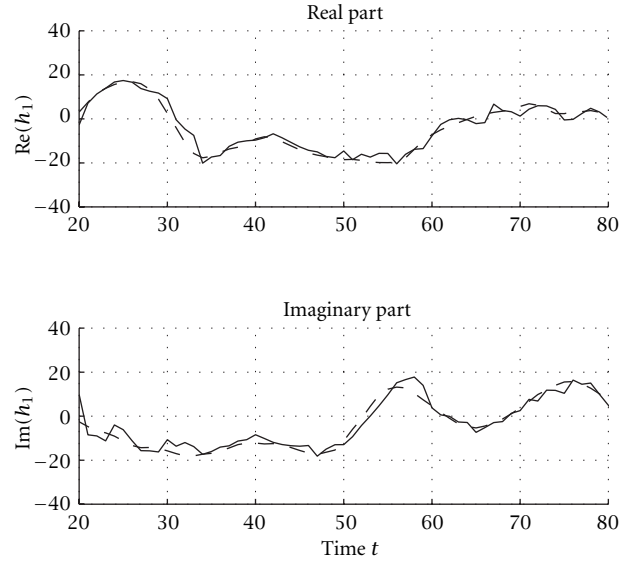


FIGURE 5: Channel tracking of the proposed STTC.

delayed samples was all selected as one. As a lower bound, the results were compared with the Genie aided case when a separated stream of known user data was space-time coded in the same proposed scheme and sent through the same channel for

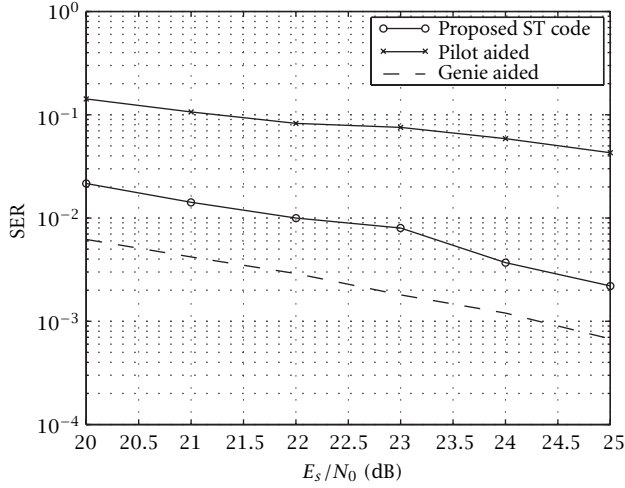


FIGURE 6: Simulation result.

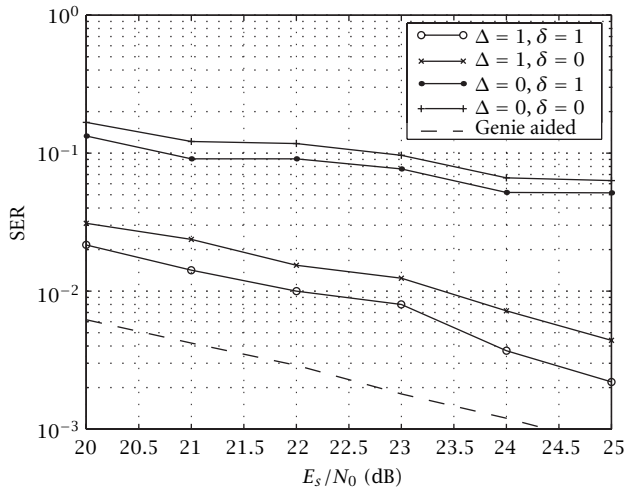


FIGURE 7: Comparison of delayed samples and delayed weights.

the estimation of the channel. The true user data were then decoded using the additional transmitted data. As we can see from Figure 6, there is a 3 dB performance gap between the two cases. If we take into account the total amount of energy used for channel estimation in the Genie aided case, the performance loss of 3 dB can be explained. For every simulated point, at least 100 symbol errors were accumulated.

If the delayed weight and delayed sample methods are not used, the performance deteriorates considerably. In Figure 7, we compared the performance when different number of delayed samples $\Delta = 0, 1$ and delayed weights $\delta = 0, 1$ were used in the algorithm. The results in Figure 7 are consistent with the ones in [16].

6. CONCLUSIONS

In this paper, we proposed and showed the viability of mixture SIS and Kalman filtering algorithm for joint estimation

and decoding of STTCs. The channels in the system were modeled as ARMA processes of known orders and model coefficients, and the whole system was represented by a DSSM. Good results can only be obtained if the problem of phase ambiguity is addressed appropriately. Four ad-hoc designed constellations were used in our simulations and the results showed that the phase ambiguity was avoided.

Future research could extend to the design of STTCs that can best explore the spatial/temporal diversity and reduce phase ambiguity when using the SIS algorithm. Another direction is to apply more computationally efficient delayed sampling algorithms based on SIS for the joint estimation and decoding of STTC.

APPENDIX

A.1. The construction of channel state vector $\mathbf{h}_{nm,t}$

The fading coefficients $\alpha_{nm,t}$ when modeled as an ARMA (r, r) process, are represented as

$$\begin{aligned} \alpha_{nm,t} + \phi_1 \alpha_{nm,t-1} + \dots + \phi_r \alpha_{nm,t-r} \\ = \rho_0 u_{nm,t} + \dots + \rho_r u_{nm,t-r}, \end{aligned} \quad (\text{A.1})$$

where $\{\rho_i\}$ and $\{\phi_i\}$ are the AR and MA coefficients, respectively. In the Z domain, (A.1) becomes $\alpha_{nm}(Z) = (\rho(Z)/\phi(Z))u_{nm}(Z)$. If we introduce a new entity, $h_{nm}(Z) = u(Z)_{nm}/\phi(Z)$, we obtain $\alpha_{nm}(Z) = \rho(Z)h_{nm}(Z)$. In the time domain, we have

$$\begin{aligned} h_{nm,t} &= -\phi_1 h_{nm,t-1} - \dots - \phi_r h_{nm,t-r} + u_{nm,t}, \\ \alpha_{nm,t} &= \rho_0 h_{nm,t} + \dots + \rho_r h_{nm,t-r}. \end{aligned} \quad (\text{A.2})$$

If we let $\mathbf{h}_{nm,t} = [h_{nm,t}, \dots, h_{nm,t-r}]^T$ be the channel state vector, its dynamics in time is described by (29), and the fading coefficient becomes $\alpha_{nm,t} = \boldsymbol{\rho}^T \mathbf{h}_{nm,t}$.

A.2. The construction of extended matrices

In this part we list the construction of \mathbf{R} , \mathbf{E} , and \mathbf{F} . First, the $NM(r+1) \times NM$ matrix \mathbf{R} is represented as

$$\mathbf{R}^T = \begin{bmatrix} \boldsymbol{\rho} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\rho} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\rho} \end{bmatrix}, \quad (\text{A.3})$$

where $\boldsymbol{\rho}$ is the $(r+1) \times 1$ vector as defined before, and $\mathbf{0}$ is an all-zero vector of the same dimension. \mathbf{E} is of dimension $NM(r+1) \times NM$ and it is written as

$$\mathbf{E} = \begin{bmatrix} \mathbf{e} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{e} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{e} \end{bmatrix}, \quad (\text{A.4})$$

where $\mathbf{0}$ is also an $(r+1) \times 1$ all-zero vector. Finally, the square matrix \mathbf{F} of dimension $NM(r+1) \times NM(r+1)$ is constructed

according to

$$\mathbf{F} = \begin{bmatrix} \Phi & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Phi & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Phi \end{bmatrix}, \quad (\text{A.5})$$

10 where $\mathbf{0}$ is an $(r + 1) \times (r + 1)$ all-zero matrix.

11 ACKNOWLEDGEMENTS

12 We thank Joaquín Míguez Arenas, Tadesse Ghirmai, and Yufei Huang for their helpful comments and suggestions.

REFERENCES

- [1] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [3] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, pp. 77–91, May 2000.
- [4] B. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.
- [5] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.
- [6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [7] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for diversity transmissions through time-selective fading channels," December 2000, submitted to *IEEE Trans. Communications*.
- [8] R. S. Kennedy, *Fading in Dispersive Communication Channels*, Wiley-Interscience, New York, USA, 1969.
- [9] M. West, *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York, USA, 1997.
- [10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [11] D. Gamerman, *Markov Chain Monte Carlo*, Chapman and Hall, New York, USA, 1997.
- [12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, USA, 1996.
- [13] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.
- [14] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Statist. Assoc.*, vol. 93, no. 443, pp. 1032–1044, 1998.
- [15] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Statist. Assoc.*, vol. 89, no. 425, pp. 278–288, 1994.
- [16] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering,"

IEEE Transactions on Information Theory, vol. 46, no. 6, pp. 2079–2094, 2000.

- [17] J. H. Kotecha and P. M. Djurić, "Sequential Monte Carlo sampling detector for Rayleigh fast-fading channels," in *Proc. IEEE ICASSP*, Istanbul, Turkey, 2000.
- [18] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 567–576, 1995.
- [19] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, USA, 1979.
- [20] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Communications*, vol. 38, no. 6, pp. 787–794, 1990.
- [21] D. L. Aspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximation," *IEEE Trans. Automat. Contr.*, vol. 17, no. 4, pp. 439–448, 1972.
- [22] G. Casella and C. P. Robert, "Rao-blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.
- [23] J. Zhang and P. M. Djurić, "A new design criterion for space-time trellis codes that combat phase ambiguity," Technical Report 794, State University of New York at Stony Brook, NY, USA, 2002.

Jianqiu Zhang received her B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, and Zhongshan University, Guangzhou, China, in 1992 and 1995, respectively. From 1995 to 1997, she was a software engineer with the Research and Development of Guangdong Nortel, Guangzhou, China. Since 1997 she has been studying for her Ph.D. in the Department of Electrical and Computer Engineering at the State University of New York at Stony Brook, USA, where she is a currently research assistant. Her research interests include statistical signal processing, turbo coding and decoding, space-time coding, multiuser detection, and joint channel estimation and detection. Ms. Zhang is a student member of the IEEE.



Petar M. Djurić received his B.S. and M.S. degrees in electrical engineering from the University of Belgrade, Yugoslavia, in 1981 and 1986, respectively, and his Ph.D. degree in electrical engineering from the University of Rhode Island, USA, in 1990. From 1981 to 1986 he was Research Associate with the Institute of Nuclear Sciences, Vinca, Belgrade, Yugoslavia. Since 1990 he has been with the State University of New York at Stony Brook, USA, where he is a Professor in 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 telecommunications, biomedicine, and power engineering. Prof. Djurić has served on numerous Technical Committees for the IEEE and SPIE, and has been invited to lecture at universities in USA and overseas. He was Associate Editor of the *IEEE Transactions on Signal Processing* and currently he is the Treasurer of the IEEE Signal Processing Conference Board. He is also Vice Chair of the IEEE Signal Processing Society Committee on Signal Processing—Theory and Methods, and a Member of the American Statistical Association and the International Society for Bayesian Analysis.



COMPOSITION COMMENTS

- 1 We changed "the above equation" to "(19)." Please check.
- 2 We changed "the above equation" to "(22)." Please check.
- 3 We changed "the previous section" to "Section 2." Please check.
- 4 In Figure 1, We changed the label " s_0, \dots, t " to " s_0, \dots, s_t ". Please check.
- 5 In Figure 1, Do you mean by "Rx Antenna" and "Tx Antenna" Receive and transmit Antennas, respectively? If so, Is it ok to denote them by "R Antenna" and "T Antenna"?
- 6 We added "+" sign to equation "(28)." Please check.
- 7 We changed "The above equation" to "Equation (37)." Please check.
- 8 We changed "the previous section" to "Section 4." Please check.
- 9 We changed "the above equation" to "(A.1)." Please check.
- 10 Please provide vol. and no. for reference [3].
- 11 Please update reference [7].
- 12 We updated reference [18]. Please check.