

A Mixed MAP/MLSE Receiver for Convolutional Coded Signals Transmitted Over a Fading Channel

Langford B. White, *Senior Member, IEEE*, and Robert J. Elliott

Abstract—This paper addresses the problem of estimating a rapidly fading convolutionally coded signal such as might be found in a wireless telephony or data network. We model both the channel gain and the convolutionally coded signal as Markov processes and, thus, the noisy received signal as a hidden Markov process (HMP). Two now-classical methods for estimating finite-state hidden Markov processes are the Viterbi algorithm and the *a posteriori* probability (APP) filter. A hybrid recursive estimation procedure is derived whereby one hidden process (the encoder state in our application) is estimated using a Viterbi-type (i.e., sequence based) cost and the other (the fading process) using an APP-based cost such as maximum *a posteriori* probability. The paper presents the new algorithm as applied specifically to this problem but also formulates the problem in a more general setting. The algorithm is derived in this general setting using reference probability methods. Using simulations, performance of the optimal scheme is compared with a number of suboptimal techniques—decision-directed Kalman and HMP predictors and Kalman filter and HMP filter per-survivor processing techniques.

Index Terms—Convolutional decoding, fading channels, hidden Markov models, per-survivor processing.

I. INTRODUCTION

IN WIRELESS telephony and data networks, propagation characteristics of the radio channel give rise to often rapid fluctuations in the received signal power [1]. For multilevel signaling constellations such as pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM), it is necessary for the receiver to have a good estimate of the instantaneous channel power gain in order to properly demodulate the signal. For many practical channels, the channel power gain may vary so quickly that gain estimation methods based on a static model of the channel gain (e.g., adaptive methods, maximum likelihood) may not track sufficiently quickly to permit demodulation of the signal. Thus, dynamic models for the channel gain should be applied in such cases. Dynamic models will give rise to estimation structures that are designed to track more quickly and, thus, should improve performance. In this paper, we specify a finite state Markov chain to model the amplitude gain process.

Manuscript received August 14, 2000; revised February 4, 2002. This work was supported by the Defence Science and Technology Organization, Australia. R. J. Elliott was supported by NSERC. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ta-Hsin Li.

L. B. White is with the Department of Electrical and Electronic Engineering, Adelaide University, Adelaide, Australia (e-mail: Lang.White@adelaide.edu.au).

R. J. Elliott is with the Department of Mathematical Sciences, University of Alberta, Edmonton, AB, Canada T6G 2G1. He is also with the Department of Applied Mathematics, Adelaide University, Adelaide, Australia (e-mail: relliott@ualberta.ca).

Publisher Item Identifier S 1053-587X(02)03219-1.

Most wireless telecommunications signals employ forward error correction (FEC) at the physical layer to give protection against symbol errors introduced by noise on the channel. The most common type of FEC is convolutional coding [2]. Convolutional coding works by adding redundancy (linear dependence) into the transmitted symbol stream by multiple input–multiple output linear FIR filtering (modulo 2). The maximum delay in the filter is called the constraint length of the encoder. An encoder that produces n output bits for each m input bits is called a rate m/n encoder. Commonly used rates are $1/2$, $3/4$, $5/6$, and $7/8$; however, for some applications (e.g., deep space communications), rates as low as $1/128$ might be used. In this paper, we consider only rate $1/n$, $n \geq 2$ encoding.

A convolutionally encoded signal may be represented as a hidden Markov model (HMM) with state consisting of all the input bits stored in the encoder memory and observation consisting of the output symbol stream. The transition structure of the state is highly constrained. For example, for a rate $1/2$ encoder of constraint length M has 2^M states (corresponding to all possible combinations of the M stored input bits in the encoder), but there are only two possible transitions from each state, corresponding to the two possibilities for the next input bit. In such highly constrained problems, it is recognized that *maximum likelihood sequence estimation* (MLSE) should be used, leading to the well-known Viterbi algorithm (VA) [4], where it is demonstrated that MLSE yields (asymptotically) the optimal error performance.

In this paper, we model our received signal as the product of the channel gain process and the convolutionally encoded process observed in additive white Gaussian noise. Thus, we have an HMP dependent on two underlying Markov chains: one being the state of the convolutional encoder and the other being the state of the channel gain process. We derive a optimal mixed estimation algorithm, whereby we seek MLSE for the encoder state and maximum *a posteriori* probability (MAP) estimates for the channel gain process. Such an algorithm clearly involves joint estimation of both underlying Markov process states, albeit with different criteria used to determine each component. The MLSE for the encoder then allows us to extract the original input bit sequence. For motivational reasons, we initially present the algorithm as applied to this specific application (Section II), but subsequently, we derive the algorithm in general terms using reference probability methods [3] (Sections IV–VI).

As a comparison, we use two classes of suboptimal approaches. These approaches are introduced in Section III. The simplest class is a decoupled structure consisting of an estimator for the channel gain process, combined with a standard MLSE algorithm applied to estimate the encoder state. This structure

mimics in some sense the usual *automatic gain control* (AGC) commonly used in receivers. Decision feedback of delayed symbols is used to parameterize the channel gain estimator. The other suboptimal methods used are based on *per-survivor processing* (PSP) [7]. Here, a bank of amplitude estimators are used; each is associated with a surviving candidate optimal path from the MLSE. There is no requirement for feedback of delayed (or otherwise) symbols with these PSP methods.

Within each class, we investigate the performance of two types of amplitude estimators. The first class is based on an AR(1) model for the amplitude process and results in a Kalman filter-based amplitude estimator. The same AR(1) model is used to derive the Kalman filter-based PSP method similar to [8] (which also addresses the frequency-selective fading case). In each case, the second-order statistics of the Markov chain amplitude process are used to parameterize the Kalman filter(s), as detailed in Appendix A. The other type of estimator uses the finite state Markov chain model itself to derive the corresponding HMP filter(s) for the amplitude process in both decision feedback and PSP modes of operation. Performance of the optimal and the four suboptimal techniques is compared with the aid of simulated four-level PAM signals in Section VII.

II. APPLICATION

We will consider convolutionally coded signals with constraint length M . Denote by X_k the length- M binary vector with the convolutional encoder state at sample time k . This process follows a “shift-register” type behavior so that for $k \geq 0$

$$X_{k+1} = SX_k + e_1 b_{k+1}. \quad (1)$$

Here, S is the $M \times M$ shift matrix with $S_{ij} = 1$ if $i = j + 1$, and zero otherwise, and e_1 is the unit vector in \mathbb{R}^M with unity in the first position. The sequence $\{b_k\}$ denotes the input binary message stream that is independent and takes the values 0 and 1 with equal probability. Consequently, the state space of X has $2^M \equiv N^{(2)}$ binary vectors. Following Section IV, this state space can be identified with the set $\{e_1^{(2)}, \dots, e_{N^{(2)}}^{(2)}\}$ of unit vectors in $\mathbb{R}^{N^{(2)}}$

$$e_j^{(2)} = \left(0, \dots, \underbrace{1}_{j^{\text{th}} \text{ position}}, 0, \dots, 0 \right)' \in \mathbb{R}^{N^{(2)}}. \quad (2)$$

We will write $X^{(2)}$ for the version of X defined in the canonical space $\{e_1^{(2)}, \dots, e_{N^{(2)}}^{(2)}\}$. Each basis vector $e_i^{(2)}$ corresponds to one binary vector in $\{0, 1\}^M$. Each binary vector corresponds to a decimal integer; therefore, we will choose the (decimal) under i so that e_i is associated with the corresponding binary vector. Any vector X_k has only two possible successor states. The transition matrix $A^{(2)}$ for $X^{(2)}$, therefore, is sparse with elements

$$a_{ij}^{(2)} = \begin{cases} 0.5, & i = 2j \bmod N^{(2)} \\ 0.5, & i = 2j + 1 \bmod N^{(2)} \\ 0, & \text{else.} \end{cases} \quad (3)$$

The encoder operates at rate $1/P$, $P \geq 2^1$ with generator matrix $G: \{0, 1\}^M \rightarrow \{0, 1\}^P$. Suppose that $\mathcal{M}: \{0, 1\}^P \rightarrow \{q_1, \dots, q_{2^P}\}$, where q_i is real, denotes the modulation operation. Its task is to map the 2^P possible values of the encoder output onto 2^P real symbol values that may be transmitted. With minor modifications, we can also handle complex modulation types such as QAM.

The transmitted signal is then

$$x_k = \mathcal{M}(GX_k \bmod 2). \quad (4)$$

The transmitted signal is propagated through a flat fading channel, which acts on the channel as a multiplicative gain [1]. The fading process is here modeled as a finite state Markov chain taking values in the set $\{a_1, a_2, \dots, a_{N^{(1)}}\}$, where $0 = a_1 < a_2 < \dots < a_{N^{(1)}}$.² We provide some justification for the choice of such a model in Appendix A. An additional reason for such a choice is the applicability of an estimation theory based on the expectation-maximization algorithm [3], which we address in forthcoming work [6]. Write $a = (a_1, \dots, a_{N^{(1)}})^T \in \mathbb{R}^{N^{(1)}}$ and suppose the chain determining the fading dynamics is $X^{(1)} = \{X_k^{(1)}\}$, where $X_k^{(1)}$ takes values in the (canonical) set of unit vectors

$$\{e_1^{(1)}, \dots, e_{N^{(1)}}^{(1)}\} \in \mathbb{R}^{N^{(1)}} \quad (5)$$

defined analogously to (2). Then, the real value (gain) associated with the fading channel at time t is

$$\langle X_k^{(1)}, a \rangle. \quad (6)$$

We suppose the transition matrix $A^{(1)} = (a_{ji}^{(1)})$ of $X^{(1)}$ has entries

$$a_{ji}^{(1)} = \begin{cases} p, & i = 1, j = 2 \\ q, & i = 2, j = 1 \\ \lambda, & i = 2, \dots, N^{(1)} - 1, j = i + 1 \\ \mu, & j = 2, \dots, N^{(1)} - 1; i = j + 1 \end{cases} \quad (7)$$

with diagonal elements $a_{ii}^{(1)}$ chosen so that each column of $A^{(1)}$ sums to unity. The state transition structure is shown in Fig. 1. This structure is somewhat restrictive as it does not permit more general transitions between the amplitude levels. More general transition models could be used if desired, although we do not address this issue in this paper.

The received signal is given by

$$y_k = \langle X_k^{(1)}, a \rangle \langle X_k^{(2)}, d \rangle + \sigma n_k \quad (8)$$

where the $\{n_k\}$ is a sequence of independent normal $\mathcal{N}(0, 1)$ random variables, and σ^2 is the noise power. When X_k is in the state corresponding to the vector $e_i^{(2)}$, write

$$d_i = \mathcal{M}(GX_k \bmod 2)$$

and

¹More general rates can also be dealt with using a multiple input version of (1).

²The zero amplitude state is included to permit detection of the presence, or nonpresence, of the signal, if desired.

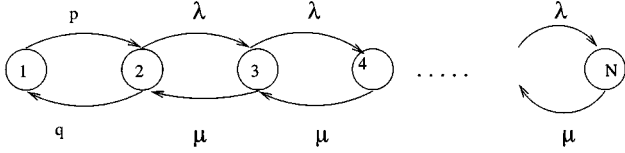


Fig. 1. State transition diagram for the fading process.

$$d = (d_1, \dots, d_{N^{(2)}})^T \quad (9)$$

so that $\mathcal{M}(GX_k \bmod 2) = d_i = \langle X_k^{(2)}, d \rangle$. We assume that all parameters $a, d, \sigma^2, p, q, \lambda$, and μ are known. Adaptive estimation is addressed in [6].

A. Optimal Demodulation

Given the observations $\mathcal{Y}_k = \{y_0, y_1, \dots, y_k\}$, we wish to obtain recursive estimates for $X_k^{(1)}$ and $X_k^{(2)}$, perhaps with some delay $\Delta \geq 0$. If one was interested in minimum variance or maximum *a posteriori* probability (MAP) estimation of both the underlying Markov chain states, one would proceed to determine a recursive update for the joint *a posteriori* probabilities

$$q_k(i, j) = \Pr\{X_{k-\Delta}^{(1)} = e_i^{(1)}, X_{k-\Delta}^{(2)} = e_j^{(2)} \mid \mathcal{Y}_k\} \quad (10)$$

and then compute the associated conditional expectations or MAP estimates. In the usual Viterbi algorithm (dynamic programming), computation of (10) is replaced by a sequential maximization over all possible sample paths of $(X_0^{(1)}, \dots, X_{k-1}^{(1)}, X_0^{(2)}, \dots, X_{k-1}^{(2)})$. The new mixed estimation procedure proposed in this paper consists of using the *a posteriori* probability estimates for the $X^{(1)}$ process, coupled with a Viterbi maximum likelihood sequence estimation criterion for $X^{(2)}$. Formally, this means considering quantities of the form

$$\begin{aligned} \tilde{q}_k(i, j) &= \max_{X_0^{(2)}, \dots, X_{k-1}^{(2)}} \\ &\times \Pr\{X_0^{(2)}, \dots, X_{k-1}^{(2)}, X_k^{(1)} = e_i^{(1)}, X_k^{(2)} = e_j^{(2)} \mid \mathcal{Y}_k\}. \end{aligned} \quad (11)$$

Candidate optimal sequences for $X_k^{(2)}$ are obtained in the usual MLSE manner, except that the each time in the backtracking phase, MAP estimates are obtained for $X_k^{(1)}$ by maximization (over i) of (11). As shown in Section VI, we have the following recursion:

$$\begin{aligned} \tilde{q}_{k+1}(i, j) \\ = \max_{1 \leq \ell \leq N^{(2)}} a_{j\ell}^{(2)} \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \tilde{q}_k(n, \ell) \frac{\phi\left(\frac{y_{k+1} - a_i d_j}{\sigma}\right)}{\sigma \phi(y_{k+1})} \end{aligned} \quad (12)$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$. Initialization at $k = 0$ is given by

$$\tilde{q}_0(i, j) = \frac{\phi\left(\frac{y_0 - a_i d_j}{\sigma}\right)}{\sigma \phi(y_0)} \pi^{(1)}(i) \pi^{(2)}(j) \quad (13)$$

where $\pi^{(1)}$ and $\pi^{(2)}$ are the initial probability distributions for $X^{(1)}$ and $X^{(2)}$, respectively.

At each time point, we keep track of the maximizing index in (12), i.e., let

$$\Psi_k(i, j) = \arg \max_{1 \leq \ell \leq N^{(2)}} a_{j\ell}^{(2)} \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \tilde{q}_k(n, \ell). \quad (14)$$

We also keep track of the maximizing value of $X^{(1)}$ for each value of $X^{(2)}$,

$$\eta_k(j) = \arg \max_{1 \leq i \leq N^{(1)}} \tilde{q}_k(i, j). \quad (15)$$

Estimates $\hat{X}_{k-\Delta|k}^{(1)}$ and $\hat{X}_{k-\Delta|k}^{(2)}$ [of $X_{k-\Delta}^{(1)}$ and $X_{k-\Delta}^{(2)}$, respectively] are produced by backtracking by a fixed number Δ samples at each time $k > \Delta$:

$$(i^*, j^*) := \arg \max_{i, j} \tilde{q}_k(i, j)$$

For $s = k - 1, \dots, k - \Delta$

$$j^* := \Psi_{s+1}(i^*, j^*)$$

$$i^* := \eta_s(j^*)$$

Then

$$\begin{aligned} \hat{X}_{k-\Delta}^{(2)} &= e_{j^*}^{(2)} \\ \hat{X}_{k-\Delta}^{(1)} &= e_{i^*}^{(1)}. \end{aligned} \quad (16)$$

The backtracking delay is necessary to enable proper construction of the maximum likelihood sequence. This delay is chosen sufficiently large that all candidate optimal sequences backtracking from time k have merged at time $k - \Delta$. Thus, in order to apply the algorithm, the quantities $\tilde{q}_k(i, j)$ are initialized at time $k = 0$ according to (13) and updated for each time $k > 0$ via (12). At each time, we also retain maximizing indices via (14) and (15). Backtracking also takes place at each time $k > \Delta$ according to (16) to extract the desired estimates.

III. REDUCED COMPLEXITY FILTERS

A standard approach in designing a suboptimal filter for the signal (8) is to decouple the gain and encoder state estimation problems. Thus, the Viterbi algorithm decoding function will employ estimates of the amplitude process, which is derived from a separate gain estimator that itself uses (possibly) delayed estimates of the encoder state. The first set of gain estimators we will use as a comparison against the optimal HMM filter are two decision-directed predictors. Prediction is required in this decoupled structure because good estimates of the encoder state sequence requires a delay, and thus, amplitude estimates for the Viterbi decoder must be predicted forward by this delay. The two types of predictors used differ in that one is based on an AR(1) model for the amplitude and results in a Kalman predictor, whereas the second is based on the Markov chain model introduced in Section II and, correspondingly, results in a HMP predictor.

There has recently been considerable interest in suboptimal sequence estimation techniques known as PSP (see e.g., [7]). As it applies to our current problem, the amplitude process $\langle X_k^{(1)}, a \rangle$ is regarded as an uncertainty in computation of the

VA likelihood function. In PSP, this uncertainty is estimated at each time using an estimator appropriate for our model of the amplitude process. However, a separate estimator for the amplitude process is used for each possible survivor path. Thus, we have a bank of filters for the amplitude process: one associated with each possible signal state value. The use of PSP in the current context will be shown to reduce the problems associated with the delay introduced in the decision-directed approach, albeit at the expense of increased computational requirements compared to the decision-directed approach. Again, two PSP estimators result: one consisting of a bank of Kalman filters and the other a bank of HMP filters.

A. Decision-Directed Kalman Predictor

Here, we use the AR(1) signal model

$$\rho_{k+1} = \lambda \rho_k + u_k \quad (17)$$

as an approximation to the second-order statistics of the process $\langle X_k^{(1)}, a \rangle$. Here, u_k is a zero-mean, white noise process with mean μ_u and variance σ_u^2 . The constant $0 < \lambda < 1$ determines the bandwidth of the process, i.e., how rapid the variations in the fading process are. Appendix A shows how to relate the parameters of this process to the true Markov amplitude process $\langle X_k^{(1)}, a \rangle$. The Kalman predictor structure is

$$\begin{aligned} \hat{\rho}_{k|k-1} &= \lambda \hat{\rho}_{k-1|k-1} + \mu_u \\ \hat{\rho}_{k|k} &= \hat{\rho}_{k|k-1} + G_k \left\{ y_k - \hat{\rho}_{k|k-1} \left\langle \hat{X}_{k|k+\Delta-1}^{(2)}, d \right\rangle \right\} \\ G_k &= \frac{\Sigma_{k|k-1} \left\langle \hat{X}_{k|k+\Delta-1}^{(2)}, d \right\rangle}{\Sigma_{k|k-1} \left\langle \hat{X}_{k|k+\Delta-1}^{(2)}, d \right\rangle^2 + \sigma^2} \end{aligned} \quad (18)$$

where $\hat{X}_{k|k+\Delta-1}^{(2)}$ is an estimate of the encoder (binary) state vector X_k written in terms of the canonical variables, given measurements $y_1, \dots, y_{k+\Delta-1}$. The predictions are given by

$$\hat{\rho}_{k+\Delta|k} = \lambda^\Delta \hat{\rho}_{k|k} + \mu_u \frac{1 - \lambda^\Delta}{1 - \lambda}. \quad (19)$$

These are used in the next step of the Viterbi algorithm processing data at time $t + \Delta$, which, in turn, results in estimates $\hat{X}_{t+\Delta|t+\Delta}$. The error ‘‘covariance’’ terms in (18) are updated by

$$\Sigma_{k+1|k} = \lambda^2 \Sigma_{k|k-1} \left\{ 1 - \left\langle \hat{X}_{k|k+\Delta-1}^{(2)}, d \right\rangle G_k \right\} + \sigma_u^2. \quad (20)$$

The computational complexity is dominated by the Viterbi part with the Kalman predictor computations being insignificant by comparison. The Viterbi algorithm needs amplitude estimates for times $k = 1, \dots, \Delta - 1$ in order to produce the estimate of $X_1^{(2)}$ and, thus, permitting the VA to commence. With such a scheme as this, suitable initialization is important. Fig. 2 depicts the structure of the predictor based algorithms.

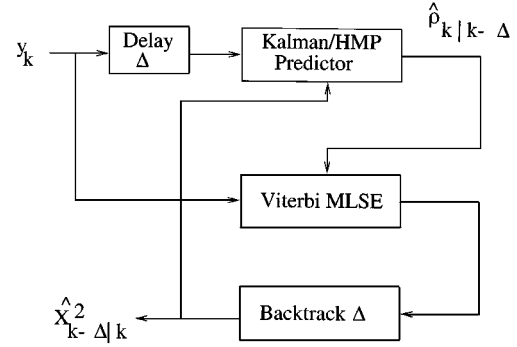


Fig. 2. Structure of the predictor-based methods.

B. Decision-Directed HMP Predictor

In this case, the Kalman predictor from Section III-A is replaced by an HMP predictor. The un-normalized filter APPs $[\pi_{k|k} \in \mathbb{R}^{N^{(1)}}]$, with $\pi_{k|k}(i) \propto \Pr\{X_k^{(1)} = e_i^{(1)} | y_1, \dots, y_k\}$ for the amplitude are updated according to [3]

$$\begin{aligned} \pi_{k|k}(i) &= \phi \left\{ \frac{y_k - a_i \left\langle \hat{X}_{k|k+\Delta-1}^{(2)}, d \right\rangle}{\sigma} \right\} \\ &\quad \times \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \pi_{k-1|k-1}(n) \end{aligned} \quad (21)$$

for $i = 1, \dots, N^{(1)}$. We then predict forward Δ samples according to

$$\pi_{k+\Delta|k} = \left(A^{(1)} \right)^\Delta \pi_{k|k} \quad (22)$$

where $\pi_{k+\Delta|k}(i) \propto \Pr\{X_{k+\Delta}^{(1)} = e_i^{(1)} | y_1, \dots, y_k\}$. The amplitude estimator used in the VA update is the amplitude state level associated with the maximizing index of $\pi_{k+\Delta|k}$.

C. PSP Kalman Filter Model

We now show details of the PSP approach, which will consist of $N^{(2)}$ Kalman filters together with the VA. Suppose we seek to compute the VA probabilities at time k , which involves maximization over each valid previous signal state at time $k - 1$ for each state at time k . Associated with each state at $k - 1$ is a survivor sequence and a sequence of Kalman filter estimates of the amplitude process that have been recursively computed using the corresponding signal state estimates. We assume that these estimates are currently labeled by the index of the state at time $k - 1$. The VA chooses the maximum likelihood predecessor signal state for each current state (which does not require the data likelihood at time k) and then updates its path probabilities by computing the data likelihood at time k , using that value of amplitude obtained by using the KF state predictions associated with the path terminating in the optimal predecessor state. That KF state is then updated and relabeled with the index of the current signal state. The process is repeated for each possible current state.

We retain the model (17) for the amplitude process and compute prediction errors associated with each possible signal state level at time k , d_i , $i = 1, \dots, d_{N^{(2)}}$ and its optimal predecessor.

For each i , let $j(i)$ denote the index of the optimal predecessor³ (from VA) for state d_i . Let

$$\epsilon_k(i) = y_k - \hat{\rho}_{k|k-1}[j(i)]d_i \quad (23)$$

where $\hat{\rho}_{k|k-1}[j]$ denotes the (one-step prediction) state associated with Kalman filter j at time $k - 1$. Filter $j(i)$ will now be relabeled according to its terminal state at time k , i.e., i . The update equations for (now) filter i are

$$\begin{aligned} \hat{\rho}_{k|k}[i] &= \hat{\rho}_{k|k-1}[j(i)] + G_k^{(i)} \epsilon_k(i) \\ G_k^{(i)} &= \frac{\Sigma_{k|k-1}[j(i)]d_i}{\Sigma_{k|k-1}[j(i)]d_i^2 + \sigma^2} \\ \hat{\rho}_{k+1|k}[i] &= \lambda \hat{\rho}_{k|k}[i] + \mu_u \\ \Sigma_{k+1|k}[i] &= \lambda^2 \Sigma_{k|k-1}[j(i)] \left\{ 1 - d_i G_k^{(i)} \right\} + \sigma_u^2. \end{aligned} \quad (24)$$

At time k , the VA uses the estimate $\hat{\rho}_{k|k-1}[j(i)]$ to compute the data likelihood corresponding to signal state d_i . Fig. 3 shows the structure of the PSP algorithms.

D. PSP HMP Filter Model

In this case, we have a bank of HMP filters similar to Section III-B, each associated with a survivor path, as in Section III-C. The survivor path HMP one-step prediction unnormalized APPs $\pi_{k|k-1}[:, i]$ with $\pi_{k|k-1}[n; i] \propto \Pr\{X_k^{(1)} = e_n^{(1)} | y_1, \dots, y_{k-1}, X_k^{(2)} = e_i^{(2)}\}$, and filter APPs $\pi_{k|k}[:, i]$ (defined analogously) associated with signal state $i = 1, \dots, N^{(2)}$ are updated according to

$$\begin{aligned} \pi_{k|k}[n; i] &= \phi \left\{ \frac{y_k - a_n d_i}{\sigma} \right\} \pi_{k|k-1}[n; j(i)] \\ \pi_{k+1|k}[n; i] &= \sum_{m=1}^{N^{(1)}} a_{nm}^{(1)} \pi_{k|k}[m; i] \end{aligned} \quad (25)$$

for $n = 1, \dots, N^{(1)}$. The amplitude state level associated with the maximizing index of $\pi_{k|k-1}[n; j(i)]$ is used to determine the VA likelihood for signal state n at time k .

E. Computational Complexity

In this section, we compare the computational requirements for the optimal, per-survivor Kalman and HMP filter techniques and the decision-directed predictor methods. For the optimal method, there is a total of $N^{(1)}N^{(2)}$ states. Using a general amplitude process Markov chain model, where the transition matrix $A^{(1)}$ is not sparse, each update (12) thus requires $O(N^{(1)2}N^{(2)})$ operations. For the sparse amplitude process model considered above, this reduces to $O(N^{(1)}N^{(2)})$ operations per sample, i.e., $N^{(1)}$ times the complexity of a standard VA. The PSP method is comprised of a VA and $N^{(2)}$ scalar Kalman filters, thus requiring $O(N^{(2)})$ operations per sample. The decision-directed Kalman predictor technique also requires $O(N^{(2)})$ operations, which is considerably fewer than the PSP

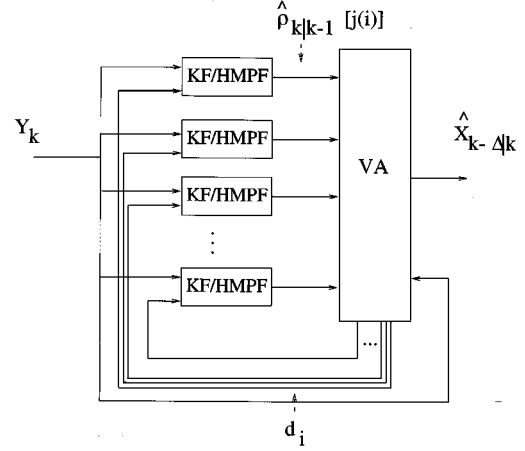


Fig. 3. Structure of the PSP methods.

technique, since only one Kalman filter is required. In the case of the decision-directed and PSP HMP methods, there are, in addition to the VA, respectively, one and $N^{(2)}$ HMP filters, each requiring $O(N^{(1)2})$ operations for the general amplitude model or $O(N^{(1)})$ for the sparse amplitude process model in Section II. It is interesting to note that the optimal technique has approximately the same computational requirements as the PSP HMP method.

IV. GENERAL SIGNAL MODEL

In this section, we describe a signal model that generalizes the application considered in Section II.

Consider two finite-state Markov chains $X^{(1)}, X^{(2)}$. Without loss of generality, the state space of $X^{(i)}$ can be taken to be the $N^{(i)}$ unit vectors $\{e_1^{(i)}, e_2^{(i)}, \dots, e_{N^{(i)}}^{(i)}\}$. The Markov chains are supposedly time homogeneous and defined on the probability space (Ω, \mathcal{F}, P) . Let $A^{(\nu)}$, $\nu = 1, 2$ denote the transition probability matrices for $X^{(\nu)}$. Then, for $k \geq 0$

$$\begin{aligned} X_{k+1}^{(1)} &= A^{(1)} X_k^{(1)} + M_{k+1}^{(1)} \\ X_{k+1}^{(2)} &= A^{(2)} X_k^{(2)} + M_{k+1}^{(2)} \end{aligned} \quad (26)$$

where $M^{(1)}$ and $M^{(2)}$ are (vector) martingale increments. That is, if

$$\mathcal{F}_k = \sigma \left\{ X_\ell^{(1)}, X_\ell^{(2)}, 0 \leq \ell \leq k \right\} \quad (27)$$

then

$$\begin{aligned} E \left[M_{k+1}^{(\nu)} \mid \mathcal{F}_k \right] &= E \left[M_{k+1}^{(\nu)} \mid X_k^{(\nu)} \right] \\ &= 0 \in \mathbb{R}^{N^{(\nu)}}. \end{aligned} \quad (28)$$

The Markov chains $X^{(1)}, X^{(2)}$ are not observed directly. Rather, there is an observation process $y = \{y_k, k \geq 0\}$ of the following form:⁴

$$y_k = h \left(X_k^{(2)} \otimes X_k^{(1)} \right) + \sigma \left(X_k^{(2)} \otimes X_k^{(1)} \right) n_k. \quad (29)$$

³We assume here that a unique optimal predecessor exists.

⁴Here, \otimes denotes Kronecker product.

We suppose the process y is scalar valued, although our discussion extends immediately to the case when y is vector valued. Here, the n_k , $k = 0, 1, \dots$ are a sequence of independent $\mathcal{N}(0, 1)$ random variables. The functions h and σ are represented by their $N^{(1)}N^{(2)}$ possible values that correspond to the $N^{(1)}N^{(2)}$ values of $X_k^{(2)} \otimes X_k^{(1)}$. Let $h_{ij} = h(e_j^{(2)} \otimes e_i^{(1)})$, and similarly define σ_{ij} . We will suppose all the values of σ_{ij} are positive.

V. REFERENCE PROBABILITY METHOD

The above dynamics for $X^{(1)}$, $X^{(2)}$ and y describe the model under a “real world” probability P . Suppose we have a different “reference” probability \bar{P} on (Ω, \mathcal{F}) such that under \bar{P} , $X^{(1)}$ and $X^{(2)}$ still have the same dynamics, that is, under \bar{P} , (26) holds, where $M^{(1)}$ and $M^{(2)}$ are now (\bar{P}, \mathcal{F}_k) martingales. However, under \bar{P} , $y = \{y_k, k \geq 0\}$ is itself a sequence of independent $\mathcal{N}(0, 1)$ random variables. We know (see [1]) that the “real world” probability P can be constructed from \bar{P} in terms of a martingale density process

$$\Lambda = \{\Lambda_k, k \geq 0\}. \quad (30)$$

Let

$$\lambda_k = \frac{\phi\left(\left(y_k - h\left(X_k^{(2)} \otimes X_k^{(1)}\right)\right) / \sigma\left(X_k^{(2)} \otimes X_k^{(1)}\right)\right)}{\sigma\left(X_k^{(2)} \otimes X_k^{(1)}\right) \phi(y_k)} \quad (31)$$

and set $\Lambda_k = \prod_{\ell=0}^k \lambda_\ell$. Write $\mathcal{G}_k = \sigma\{X_\ell^{(1)}, X_\ell^{(2)}, y_\ell: \ell \leq k\}$. If we define the probability P by setting

$$\left. \frac{dP}{d\bar{P}} \right|_{\mathcal{G}_k} = \Lambda_k \quad (32)$$

then under P , $X^{(1)}$ and $X^{(2)}$ remain Markov chains with the required transitions matrices $A^{(1)}$ and $A^{(2)}$, but the process $(y_k - h(X_k^{(2)} \otimes X_k^{(1)})) / \sigma(X_k^{(2)} \otimes X_k^{(1)})$, $k = 0, 1, \dots$ is a sequence of independent $\mathcal{N}(0, 1)$ random variables. Call them n_k , $k \geq 0$. Then, under \bar{P} , (29) holds as required. Thus, dP and $d\bar{P}$ are, respectively, the joint probability density functions of the processes $X^{(1)}$, $X^{(2)}$, and y under the original and reference probability models. The reference probability method has been shown [3] to greatly facilitate the derivation of both state and parameter estimation procedures associated with dynamic models. In general, estimation arises under \bar{P} , according to the Bayes' rule

$$E\{f(X_k^{(1)}, X_k^{(2)}) | \mathcal{Y}_k\} = \frac{\bar{E}\{\Lambda_k f(X_k^{(1)}, X_k^{(2)}) | \mathcal{Y}_k\}}{\bar{E}\{\Lambda_k | \mathcal{Y}_k\}} \quad (33)$$

for appropriately measurable functions f , where E and \bar{E} denote expectation under P and \bar{P} , respectively. Often, we work only with the numerator in (33), which is referred to as an *unnormalized* quantity; the correct value is obtained by appropriate normalization [3].

VI. STATE ESTIMATION IN THE GENERAL SETTING

Given the observations y_0, y_1, \dots, y_k , we wish to obtain recursive estimates for $X_k^{(1)}$ and $X_k^{(2)}$ or, equivalently, $X_k^{(2)} \otimes X_k^{(1)}$. If one was interested in minimum variance estimation of the underlying Markov chain states, one would proceed to determine a recursive update for the joint unnormalized *a posteriori* probabilities

$$q_k(i, j) = \bar{E}\left[\Lambda_k \left\langle X_k^{(1)}, e_i^{(1)} \right\rangle \left\langle X_k^{(2)}, e_j^{(2)} \right\rangle \middle| \mathcal{Y}_k\right] \quad (34)$$

where expectation is over both $X_k^{(1)}$ and $X_k^{(2)}$. In the usual Viterbi algorithm (dynamic programming), the expectation of (34) is replaced by a maximization over all possible sample paths of $\{X_0^{(1)}, \dots, X_{k-1}^{(1)}, X_0^{(2)}, \dots, X_{k-1}^{(2)}\}$. However, the new mixed estimation procedure proposed in this paper consists of using the *a posteriori* probability estimates for the $X^{(1)}$ process, coupled with a Viterbi maximum likelihood sequence estimation criterion for $X^{(2)}$. Formally, this means considering quantities of the form

$$\tilde{q}_k(i, j) = \bar{E} \times \left[\max_{X_0^{(2)}, \dots, X_{k-1}^{(2)}} \Lambda_k \left\langle X_k^{(1)}, e_i^{(1)} \right\rangle \left\langle X_k^{(2)}, e_j^{(2)} \right\rangle \middle| \mathcal{Y}_k\right] \quad (35)$$

i.e., we replace the conditional expectation over $X_0^{(2)}, \dots, X_{k-1}^{(2)}$ by maximization over the same quantities. We have the following recursion:

$$\begin{aligned} \tilde{q}_{k+1}(i, j) &= \max_{1 \leq \ell \leq N^{(2)}} a_{j\ell}^{(2)} \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \tilde{q}_k(n, \ell) \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})}. \end{aligned} \quad (36)$$

As the $\phi(y_{k+1})$ factor is common to all terms, it may be discarded, if so desired. Initialization at $k = 0$ is given by

$$\tilde{q}_0(i, j) = \frac{\phi\left(\frac{y_0 - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_0)} \pi^{(1)}(i) \pi^{(2)}(j) \quad (37)$$

where $\pi^{(1)}$ and $\pi^{(2)}$ are the initial probability distributions for $X^{(1)}$ and $X^{(2)}$, respectively. A proof of (36) is given in Appendix B. Backtracking to extract the estimates of $X_k^{(1)}$ and $X_k^{(2)}$ is as described in (14)–(16).

VII. SIMULATIONS

In this section, we present results of simulation experiments used to compare six demodulators applied to the fading convolutionally coded signal described above. The performance of the optimal scheme as defined by the mixed cost function (11), the Kalman and HMP PSP techniques, the Kalman and HMP predictor-based methods, and the usual MLSE with the amplitude process known to the receiver were compared. The Kalman and HMM predictor methods used $\Delta = 1$, which we argue later is the best value to choose, at least in the Kalman case. In our experiments, we did not observe any statistically significant

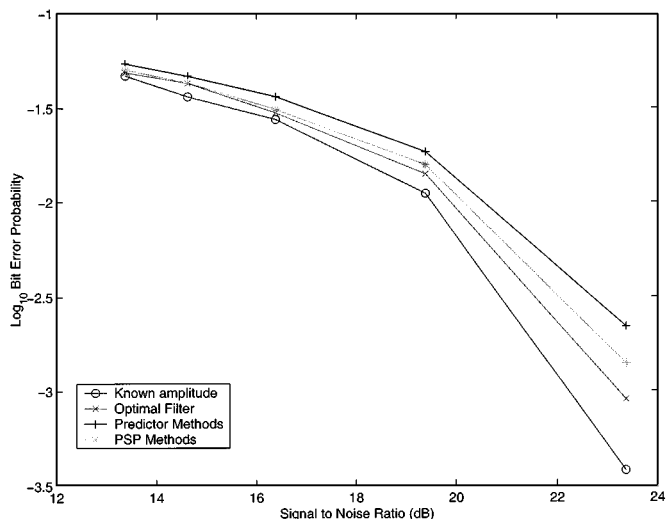


Fig. 4. BER performance for filter ($\mu = \lambda = 0.05$).

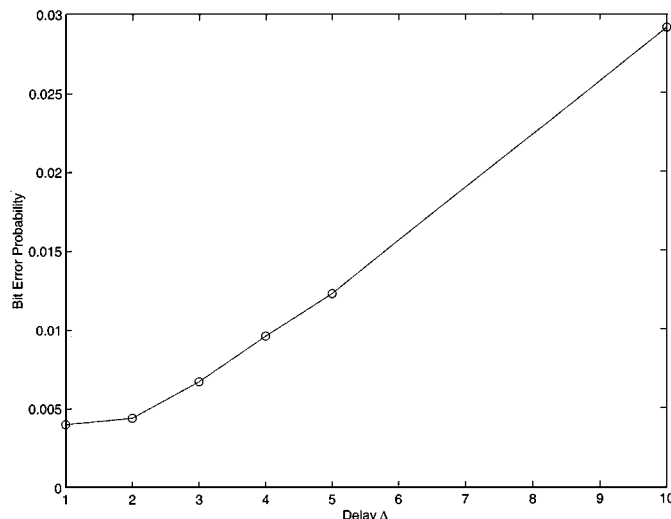


Fig. 6. Effect of parameter Δ on the Kalman predictor method.

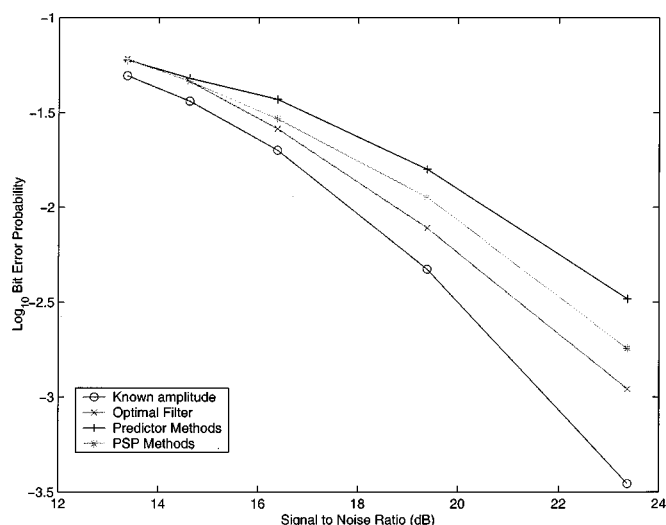


Fig. 5. BER performance for filter ($\mu = \lambda = 0.2$).

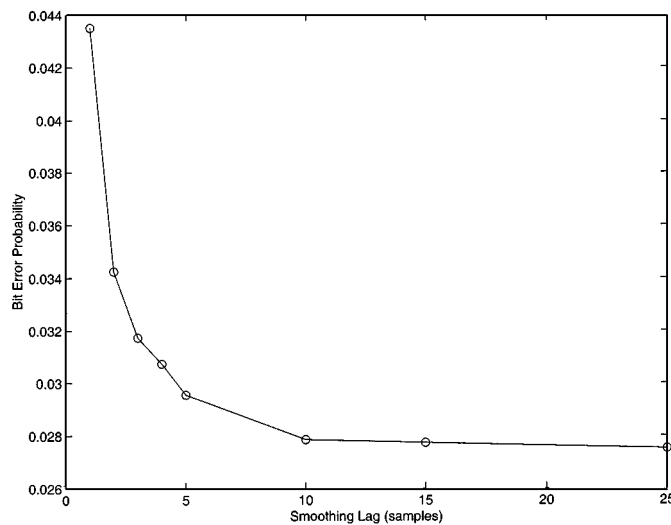


Fig. 7. Effect of smoothing lag in MLSE.

difference between the performance of the Kalman filter-based methods and the corresponding HMP-based methods, i.e., the Kalman predictor method performed similarly to the HMP predictor method and similarly for the PSP techniques. This is despite the fact that the Kalman-based algorithms use only the second-order statistics of the gain process, whereas the HMP techniques utilize the true model. Evidently, there appears little difference in terms of the resulting bit error rates (BERs) for the examples we consider.

The resulting BERs are shown in Fig. 4. Here, $\lambda = \mu = 0.05$, $p = 1$, and $q = 0$. Fig. 5 repeats, for $\lambda = \mu = 0.2$, a more rapidly varying amplitude case. Here, the average SNR is defined by

$$\text{SNR} = \frac{E\langle X_k^{(1)}, a \rangle^2 E\langle X_k^{(2)}, d \rangle^2}{\sigma^2}. \quad (38)$$

It is seen that in both cases, the predictor-based methods perform the worst, with the PSP methods yielding performance in

between that of the predictor methods and the optimal method. The optimal technique performs quite close to the case where the receiver knows the fading process exactly. The performance gain in using the optimal filter appears to increase for higher SNRs.

We also examined the error behavior of the Kalman predictor method as a function of the parameter Δ . Recall that $\Delta \geq 1$ denotes the time lag (in samples) until we make a decision about the encoder state. This value is used to predict the amplitude process (gain) value forward from the Kalman filter to the Viterbi decoder. Fig. 6 shows rather interesting behavior in that the smallest possible $\Delta = 1$ resulted in the best overall BER performance. Here, $\mu = \lambda = 0.1$, and the SNR was 29 dB. Clearly, larger smoothing lags, which one would normally expect to result in better state estimates (for the encoder process), are not resulting in better performance of the overall scheme.

In order to investigate this behavior, we examined the performance of fixed-lag smoothing when the amplitude process is exactly known at the receiver. Fig. 7 shows the reduction in the bit error probability as Δ is increased. This is the expected

type of behavior [5], with error probability reducing rapidly for a small lag but approaching a constant value as $\Delta \rightarrow \infty$. Here, $\mu = \lambda = 0.05$, and $\text{SNR} = 25.9$ dB. Thus, we can conclude that the behavior evident in Fig. 6 is due to the poor prediction performance of the Kalman method. This is to be expected since it is not generally possible to accurately predict a discrete state HMM. We conclude that some sort of joint estimation procedure (either explicit as in our optimal approach or implicit as in PSP) is really necessary to obtain reasonable performance with the model we have assumed for the fading channel amplitude process.

VIII. CONCLUSION

In this paper, we have derived the optimal filter for a hidden Markov process consisting of the product of two statistically independent underlying Markov chains observed in additive white Gaussian noise, which may have state dependent moments. We apply a mixed estimation criterion in order to formulate the filter. We seek the *maximum likelihood sequence* corresponding to one of the underlying chains and *a posteriori* probabilities (APP's) for the other underlying chain. This mixed criterion is motivated by a particular application, namely, the demodulation of a rapidly fading convolutionally coded communications signal. The signal is decoded using maximum likelihood sequence estimation (MLSE). Estimation of the fading process is performed according to the maximum *a posteriori* probability criterion, requiring computation of APPs. The performance of the optimal filter for this example is compared with a more conventional approach consisting of decoupled estimators for each underlying chain. These estimators are standard MLSE implemented via the Viterbi algorithm for the convolutionally coded part and a decision-directed predictor for the gain process. The case where the gain process is known to the receiver is used as a benchmark. We also compare performance with a per-survivor processing (PSP) technique that has computational complexity less than the optimal method but greater than the simple prediction technique. In both the prediction and PSP methods, we examined both Kalman and hidden Markov process-based approaches and found no significant difference in performance between them in each case. The PSP approach has been addressed in [8], which also considers frequency-selective fading. Simulations show that the predictor methods perform worst, but the optimal filter illustrates minimal performance degradation, as compared with the known amplitude case. The PSP technique offers performance between that of the simple prediction method and the optimal method. In this paper, we have not addressed the issue of estimating the fading process model parameters. This problem is being addressed in current work [6]. We have also not addressed frequency-selective fading here but indicate that the same idea as presented here could be applied to such cases, albeit with a substantial increase in computational requirements.

APPENDIX A

STATISTICAL PROPERTIES OF THE AMPLITUDE PROCESS

In this Appendix, we show how to parameterize the AR(1) amplitude process model (17) so that its second-order statistics

(mean and autocorrelation) match that of the amplitude process $\langle X_k^{(1)}, a \rangle$. Consider the AR(1) process in (17). The first two moments of ρ_k are (as $k \rightarrow \infty$)

$$\begin{aligned} E\{\rho_k\} &= \frac{\mu_u}{1-\lambda} \\ E\{\rho_k^{(2)}\} &= \frac{\sigma_u^2}{1-\lambda^2} + E\{\rho_k\}^2. \end{aligned} \quad (39)$$

The autocorrelation function is given by

$$E\{\rho_k \rho_{k+n}\} - E\{\rho_k\}^2 = E\{\rho_k^2\} \lambda^{|n|}. \quad (40)$$

The stationary probabilities π for the birth-death process $X_k^{(1)}$ can be easily determined via the usual formula, and thus, the first- and second-order moments can be determined. The probabilities are given by

$$\pi_n = \pi_1 \begin{cases} 1, & n = 1 \\ p/q, & n = 2 \\ p/q(\lambda/\mu)^{n-2}, & n = 3, \dots, N^{(1)} \end{cases} \quad (41)$$

where π_1 is chosen to normalize the sum of all the π_n to unity. Thus, the first two moments are (as $k \rightarrow \infty$)

$$\begin{aligned} E\{\langle X_k^{(1)}, a \rangle\} &= \sum_{n=1}^{N^{(1)}} a_n \pi_n \\ E\{\langle X_k^{(1)}, a \rangle^2\} &= \sum_{n=1}^{N^{(1)}} a_n^2 \pi_n. \end{aligned} \quad (42)$$

The autocorrelation function of the fading process is approximately given by

$$E\{\langle X_k^{(1)}, a \rangle \langle X_{k+n}^{(1)}, a \rangle\} \approx E\{\langle X_k^{(1)}, a \rangle^2\} \lambda_2^{|n|} \quad (43)$$

where λ_2 is the second largest magnitude eigenvalue of the state transition matrix $A^{(1)}$.⁵ Thus, by moment matching, the parameters λ , μ_u , and σ_u^2 can be determined. The second-order statistical properties of $\langle X_k^{(1)}, a \rangle$ and ρ_k will then be approximately the same. More details on the approximation of Markov chains and HMMs by a linear Gauss-Markov processes is given in [11].

Relationship to Standard Fading Models

In a typical mobile scenario, the channel gain is modeled as the product of three terms [1]. The first represents path loss, which we regard as constant in this study. The second models so-called shadow fading, which involves variations on the order of tens of wavelengths at the propagation frequency. This gain is modeled as a log-normal random variable. Traditionally, independent log-normal variables are chosen; however, temporal correlations can be introduced as in [10] by filtering these variables with a first-order lowpass filter. The third term models multipath fading. We consider here only flat (nonfrequency-se-

⁵This assumes that the second largest magnitude eigenvalue is distinct, and sufficiently larger than the third largest in magnitude. The result can be generalized.

lective) fading. This gain is modeled as a Ricean random variable. Again, filtering can be employed to model temporal correlations in the gain.

Let U_k and V_k denote, respectively, an iid log-normal process and an iid Ricean process with the same second-order statistics that are mutually independent. Suppose these sequences are filtered with single pole filters (as suggested in [10]), resulting in

$$\begin{aligned} W_k &= (1 - \beta) \sum_{n=0}^{\infty} \beta^n U_{k-n} \\ Z_k &= (1 - \gamma) \sum_{n=0}^{\infty} \gamma^n V_{k-n} \end{aligned} \quad (44)$$

where $\beta, \gamma \in (0, 1)$ are the filter poles. The processes W_k and Z_k will have complicated probability distributions due to the filtering; however, they have autocorrelation functions of the form

$$\begin{aligned} R_w(n) &= C_1 \beta^{|n|} \\ R_z(n) &= C_2 \gamma^{|n|} \end{aligned} \quad (45)$$

where C_1 and C_2 are constants. By the independence of U_k and V_k , we have that the autocorrelation function of the *product* $W_k Z_k$ is

$$R_{wz}(n) = C_3 (\beta\gamma)^{|n|} \quad (46)$$

i.e., the fading process is a first-order Markov process, albeit one with a complicated probability distribution form. We approximate the process by discretizing the amplitude range and applying the procedure above resulting in a HMM. This is equivalent to approximating the above density by step functions, and in principle, such an approximation may be made arbitrarily accurate by taking the number of states of the Markov process to be sufficiently large.

APPENDIX B

PROOF OF RECURSION FOR THE OPTIMAL FILTER

In this Appendix, we provide a proof for the recursion (36).

Theorem: The ‘‘pseudo-probabilities’’ of (35) are computed recursively via

$$\tilde{q}_{k+1}(i, j) = \max_{1 \leq \ell \leq N^{(2)}} a_{j\ell}^{(2)} \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \tilde{q}_k(n, \ell) \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})}. \quad (47)$$

Proof: Consider

$$\begin{aligned} \tilde{q}_{k+1}(i, j) &= \bar{E} \left[\max_{X_0^{(2)}, \dots, X_k^{(2)}} \Lambda_{k+1} \left\langle X_{k+1}^{(2)}, e_j^{(2)} \right\rangle \right. \\ &\quad \left. \times \left\langle X_{k+1}^{(1)}, e_i^{(1)} \right\rangle \middle| \mathcal{Y}_{k+1} \right] \end{aligned}$$

$$\begin{aligned} &= \bar{E} \max_{X_0^{(2)}, \dots, X_k^{(2)}} \left[\Lambda_k \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})} \left\langle A^{(2)} X_k^{(2)}, e_j^{(2)} \right\rangle \right. \\ &\quad \left. \times \left\langle A^{(1)} X_k^{(1)}, e_i^{(1)} \right\rangle \middle| \mathcal{Y}_k \right] \\ &= \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})} \bar{E} \max_{X_0^{(2)}, \dots, X_{k-1}^{(2)}} \max_{X_k^{(2)}} \\ &\quad \times \left[\Lambda_k \left\langle A^{(2)} X_k^{(2)}, e_j^{(2)} \right\rangle \times \left\langle A^{(1)} X_k^{(1)}, e_i^{(1)} \right\rangle \middle| \mathcal{Y}_k \right] \\ &= \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})} \bar{E} \max_{X_0^{(2)}, \dots, X_{k-1}^{(2)}} \max_{X_k^{(2)}} \\ &\quad \times \left[\Lambda_k \left(\sum_{n=1}^{N^{(1)}} \left\langle X_k^{(1)}, e_n^{(1)} \right\rangle \right) \times \left(\sum_{\ell=1}^{N^{(2)}} \left\langle X_k^{(2)}, e_\ell^{(2)} \right\rangle \right) \right. \\ &\quad \left. \times \left\langle A^{(2)} X_k^{(2)}, e_j^{(2)} \right\rangle \left\langle A^{(1)} X_k^{(1)}, e_i^{(1)} \right\rangle \middle| \mathcal{Y}_k \right] \\ &= \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})} \max_{1 \leq \ell \leq N^{(2)}} \sum_{n=1}^{N^{(1)}} a_{j\ell}^{(2)} a_{in}^{(1)} \bar{E} \\ &\quad \times \left[\max_{X_0^{(2)}, \dots, X_{k-1}^{(2)}} \Lambda_k \times \left\langle X_k^{(1)}, e_n^{(1)} \right\rangle \left\langle X_k^{(2)}, e_\ell^{(2)} \right\rangle \middle| \mathcal{Y}_k \right] \\ &= \frac{\phi\left(\frac{y_{k+1} - h_{ij}}{\sigma_{ij}}\right)}{\sigma_{ij} \phi(y_{k+1})} \max_{1 \leq \ell \leq N^{(2)}} a_{j\ell}^{(2)} \sum_{n=1}^{N^{(1)}} a_{in}^{(1)} \tilde{q}_k(n, \ell) \quad (48) \end{aligned}$$

as required. \square

The steps in the proof rely use of the martingale property of $X_k^{(1)}$ and $X_k^{(2)}$ (lines 1–2) together with the independence of y_k under \bar{P} (line 3). The remainder consists of algebraic manipulations that rely on the unit basis vectors $e_n^{(1)}$ and $e_\ell^{(2)}$.

ACKNOWLEDGMENT

R. J. Elliott gratefully acknowledges the hospitality of the Department of Applied Mathematics, Adelaide University. The authors thank the referees for their helpful suggestions.

REFERENCES

- [1] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [2] B. Sklar, *Digital Communications, Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [3] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. New York: Springer-Verlag, 1995.
- [4] A. J. Viterbi, ‘‘Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,’’ *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, 1967.
- [5] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [6] L. B. White and R. J. Elliott, *Adaptive Parameter Estimation for Rapidly Fading Convolutional Coded Signals*, to be published.

- [7] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 42, pp. 354–364, Feb.–Apr. 1995.
- [8] M. E. Rollins and S. J. Simmons, "Simplified per-survivor Kalman processing in fast frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 544–553, May 1997.
- [9] B. H. Juang and L. R. Rabiner, "The segmented k-means algorithm for estimating the parameters of a hidden Markov model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1639–1641, Sept. 1990.
- [10] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, 1991.
- [11] L. B. White, "A comparison of optimal and Kalman filtering for a hidden Markov process," *IEEE Signal Processing Lett.*, vol. 5, pp. 124–126, May 1998.



Langford B. White (SM'00) received the B.Sc., B.E., and Ph.D. degrees in mathematics and electrical engineering in 1984, 1985, and 1989, respectively.

From 1985 to 1999, he worked with the Defence Science and Technology Organization, Salisbury, Australia. Since 1999, he has been Professor of electrical engineering at Adelaide University, Adelaide, Australia. His research interests include signal processing, control, and internet engineering.



Robert J. Elliott received the B.E. and M.E. degrees from Oxford University, Oxford, U.K., and the Ph.D. and D.Sc. degrees from Cambridge University, Cambridge, U.K.

He has held positions at the University of Newcastle, Newcastle-upon-Tyne, U.K., Yale University, New Haven, CT, Oxford, Warwick University, Warwick, U.K., Hull University, Hull, U.K., and the University of Alberta, Edmonton, AB, Canada, and visiting positions at the University of Toronto, Toronto, ON, Canada, Northwestern University, Evanston, IL, University of Kentucky, Louisville, Brown University, Providence, RI, and universities in Paris, France, Denmark, Hong Kong, and Australia. Currently, he is the Royal Bank Professor of Finance at the University of Alberta. He has authored six books, including (with L. Aggoun and J. Moore) *Hidden Markov Models: Estimation and Control* (New York: Springer-Verlag, 1995) and (with P. E. Kopp) *Mathematics of Financial Markets* (New York: Springer, 1999), which is in its third printing, and over 250 papers. His work in recent years has investigated stochastic processes in engineering and finance.