

# Deinterleaving Pulse Trains Using Discrete-Time Stochastic Dynamic-Linear Models

John B. Moore, *Fellow, IEEE*, and Vikram Krishnamurthy, *Member, IEEE*

**Abstract**—Pulse trains from a number of different sources are often received on the one communication channel. It is then of interest to identify which pulses are from which source, based on different source characteristics. This sorting task is termed deinterleaving. In this paper we next propose time-domain techniques for deinterleaving pulse trains from a finite number of periodic sources based on the time of arrival (TOA) and pulse energy, if available, of the pulses received on the one communication channel. We formulate the pulse train deinterleaving problem as a stochastic discrete-time dynamic linear model (DLM), the “discrete-time” variable  $k$  being associated with the  $k$ th received pulse. The time-varying parameters of the DLM depend on the sequence of active sources. The deinterleaving detection/estimation task can then be done optimally via linear signal processing using the Kalman filter (or recursive least squares when the source periods are constant) and tree searching. The optimal solution, however, is computationally infeasible for other than small data lengths since the number of possible sequences grow exponentially with data length. Here we propose and study two of a number of possible suboptimal solutions: 1) Forward dynamic programming with fixed look-ahead rather than total look-ahead as required for the optimal scheme; 2) a probabilistic teacher Kalman filtering for the detection/estimation task. In simulation studies we show that when the number of sources is small, the proposed suboptimal schemes yield near-optimal estimates even in the presence of relatively large jitter noise. Also, issues of robustness and generalizations of the approach to the case of missing pulses, unknown source number, and non-Gaussian jitter noise are addressed.

## NOMENCLATURE

$N$	Number of signal sources.
$T^i, \phi^i$	Period and phase of $i$ th source.
$t_k$	Noise-free TOA of $k$ th pulse.
$y_k$	Observed TOA of $k$ th pulse (2.2).
$Y_k = y_1, \dots, y_k$	
$s_k = i$	If $i$ th source generated the $k$ th pulse.
$X_k$	State of process; $X_k = e_i$ if $s_k = i$ (2.1).
$\mathcal{X}_k = X_1, \dots, X_k$	
$\tau_k^i$	Last time source $i$ was active up to and including arrival of $k$ th pulse (2.3).

$F(X_k), G(X_k), H(X_k)$	State space representation of DLM (2.7).
DLM	Dynamic linear model.
FDP	Forward dynamic programming.
KF	Kalman filter.
MAP	Maximum <i>a posteriori</i> .
PT	Probabilistic teacher.
RLS	Recursive least squares.
TOA	Time of arrival.
WGN	White Gaussian noise.

## I. INTRODUCTION

PULSE trains from a number of different sources are often received on the one communication channel. It is then of interest to identify which pulses are from which source, based on the assumption that the different sources have different characteristics. This sorting task is termed deinterleaving. It has applications in radar detection and potential applications in computer communications and neural systems. In this paper we address, in the first instance, the problem of deinterleaving time-interleaved pulse trains from a finite *known* number of *periodic* sources. We assume that observations of the time of arrival of the pulses are obtained in additive white Gaussian noise without any information of the pulse amplitudes and phases. The aim is to deinterleave the received signal, i.e., to detect which source is responsible for each received pulse. From this it is trivial to estimate the periods and phases of the periodic pulse-train sources, although the detection and estimation tasks are intimately linked.

A number of suboptimal heuristic solutions have been proposed for deinterleaving, e.g., histogramming [1], folding [2]. These techniques work well when the jitter noise is small. In addition they require prior information about the periods of the sources to select appropriate initial conditions.

In this paper, we first formulate the pulse-train deinterleaving problem as a stochastic discrete-time dynamic linear model (DLM) (see pp. 212–215 in [3], [4]). A DLM is a time-varying linear system formulated in state space form with the state matrix and observation matrix at each time instant belonging to a finite set of possible values. In the deinterleaving case, the discrete-time instants are not the pulse times of arrival but rather integers indicating the pulses number. Thus the “time” instant  $k$  indicates the arrival of the  $k$ th pulse. Then the state and observation matrices at each “time” instant  $k$ , termed here pulse instant  $k$ , depend on which source is active to generate the  $k$ th pulse. The state at each pulse instant consists of the

Manuscript received April 6, 1993; revised February 12, 1994. This work was supported by the Cooperative Research Center for Robust and Adaptive Systems. The associate editor coordinating the review of this paper and approving it for publication was Prof. Douglas Williams.

The authors are with the CRASyS, Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 2601, Australia.

IEEE Log Number 9404769.

periods of the sources and the last arrival time for each of the sources. If the pulses contain energy (e.g., amplitude) information about the sources, this information can also be incorporated in the state vector.

If the actual source sequence was known, e.g., when there is only one periodic source, then optimal estimates of the state of the DLM and hence the periods of the sources can be obtained using a Kalman filter (KF) or using recursive least-squares (RLS) parameter estimation when the pulse periods are constant. However, in general, when there is more than one source, because the actual source sequence is not known, the optimal solution involves evaluating the prediction-error cost of each source sequence and choosing the sequence with the minimum cost. The number of possible source sequences increases exponentially with the data length and so this procedure is not computationally feasible for other than short data segments with few sources (typically about 20 data points and three sources). Clearly, forward dynamic programming, in its simplest form cannot be used effectively to pick the optimal sequence because the costs at any stage of the multistage decision process are dependent on the history of the sequence (path).

In this paper we propose two suboptimal solutions to the deinterleaving problem. These suboptimal solutions can be viewed as tree-pruning algorithms that attempt to eliminate low probability paths so as to achieve a computationally feasible algorithm.

1) *Forward dynamic programming (FDP) with fixed look-ahead*: As described above, FDP in its rudimentary form cannot be used to obtain the optimal path sequence. We propose a scheme that combines the optimal full tree-search algorithm over a short segment (look-ahead interval) to reject improbable paths and FDP to update the most likely sequences and costs terminating at each source at each pulse arrival. That is, over the look-ahead interval the KF prediction error of all sequences is evaluated. For  $N$  sources with a look-ahead of  $\Delta$  the computational cost is  $O(N^{\Delta+3})$ . Typically, for a small number of sources ( $N \leq 10$ ), simulations show that for satisfactory performance, the look-ahead required is about 3, and so the computational cost is not excessive. Of course, if the look-ahead interval is the length of the observation sequence, then the algorithm is the optimal full tree-search algorithm mentioned above. The tree-pruning algorithms presented in [5, ch. 2], are very similar to FDP with look-ahead.

2) *Probabilistic teacher (PT)*: PT algorithms have been proposed for estimating DLM's in [6], [10]. If the *a priori* probabilities of the sources (related to the periods) are known exactly then the estimates of the periods using PT asymptotically tend to the optimal estimates. In fact, as shown in simulation studies for sequences of reasonable length, PT using the correct *a priori* probabilities yields estimates of the periods as good as the estimates when the true source sequence is known. When the *a priori* probabilities are not known, we compute *a posteriori* probabilities and use PT with these probabilities to obtain state estimates of the DLM. However, PT using *a posteriori* probabilities is prone to error propagation and is not robust to initial conditions.

Finally in this paper, robustness issues are studied when the assumptions on the models are relaxed. In particular, we consider the following cases: Missing or supernumery (extra) pulses; pulses with energy information. We propose modified algorithms for dealing with these cases.

Simulation studies show that both suboptimal algorithms yield useful estimates providing the initial state estimates are chosen sufficiently close to the true values.

This paper is organized as follows: In Section II we formulate the deinterleaving problem as a DLM and present the optimal solution, which is computationally feasible only for short data sequences. In Sections III and IV our two suboptimal solutions using FDP with look-ahead and PT are described. In Section V simulation examples are presented. In Section VI issues of robustness, including the presence of missing and supernumery pulses, are addressed.

## II. PROBLEM FORMULATION

In this section, we first present the pulse-train signal model in its simplest form. The estimation objectives for deinterleaving the pulse trains are then described in terms of this model. We formulate the deinterleaving problem as a DLM suitable for estimation by a KF. A parametrized DLM, which is suitable for RLS estimation, is also given. If the pulses contain energy information of the sources, we show that this information can also be incorporated in the DLM. We show how to evaluate the one-step prediction-error costs using a KF or RLS estimator. Finally, the optimal solution, which involves a full tree search and the prediction-error costs using either a KF or RLS estimator, is given.

### A. Signal Model

Consider  $N$  signal sources, each generating periodic pulse-trains with period  $T^i$ ,  $i = 1, \dots, N$ . That is,  $T^i$  is the period of the  $i$ th source. We assume that the pulses contain no amplitude or width information and consequently no information about the pulse source. The pulses are then interleaved, i.e., summed at the receiver in a single noisy communication channel. Let  $\phi^i$  denote the time of arrival (TOA) at the receiver of the first pulse from source  $i$  in the absence of measurement noise. That is,  $\phi^i$  is the phase of the  $i$ th source. Let  $t_k$  denote the noise-free TOA of the  $k$ th pulse at the receiver. Here  $k \in I^+$  where  $I^+$  denotes the set of positive integers. Let  $s_k = i$  denote that source  $i$  is active at pulse instant  $k$ . Thus  $s_k \in [1, \dots, N]$ . Let  $e_i$ ,  $i = 1, \dots, N$ , be unit column vectors in  $R^N$  with 1 in the  $i$ th position. Let  $X_k \in \{e_1, \dots, e_N\}$  denote the state of the process in that

$$X_k = e_i \quad \text{if} \quad s_k = i. \quad (2.1)$$

*Example*: Consider two periodic sources ( $N = 2$ ) with periods  $T' = [11, 17]$  and phases  $\phi' = [3, 4]$ . In Fig. 1 we show the pulse trains from the two sources. The interleaved pulse train is also shown. Fig. 2 shows the corresponding evolution with pulse number of the sequence of active sources.

Let  $y_k$  denote the *observed* TOA of the  $k$ th pulse at the receiver. The observations  $y_k$  are in fact  $t_k$  contaminated by

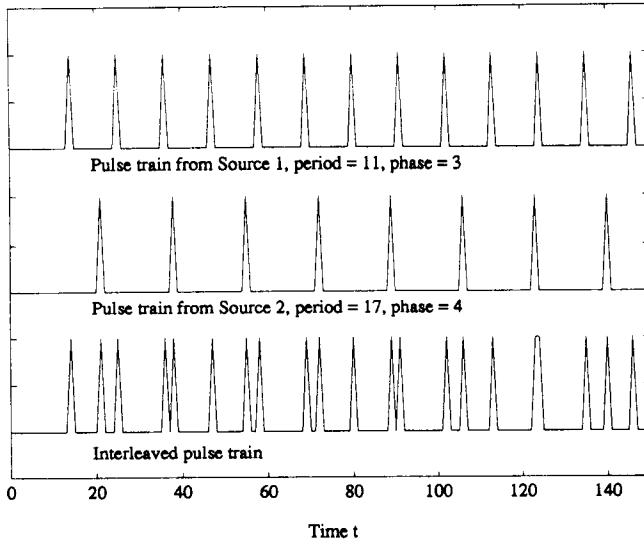


Fig. 1. Pulse trains in absence of jitter.

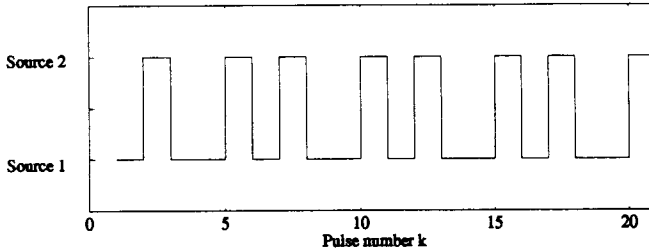


Fig. 2. Source sequencing.

additive jitter noise  $w_k$  as follows

$$y_k = t_k + w_k, \quad w_k \sim N[0, \sigma_w^2]. \quad (2.2)$$

Here  $w_k$  is zero-mean white Gaussian noise (WGN) with known variance  $\sigma_w^2$ . Let  $Y_k$  denote the sequence of observations till time  $k$ :  $Y_k = (y_1, \dots, y_k)$ . We assume that all pulses are detected, i.e., there are no missing pulses. We also assume that if  $l$  pulses arrive simultaneously at the receiver, these  $l$  pulses are detected and  $y_k = y_{k+1} = \dots = y_{k+l-1}$ . This is a practical assumption if the pulses have energies that are additive.

### B. Deinterleaving Objects

1) *Estimation*: Obtain a sequential optimal (minimum mean-square error) estimate of the source periods  $T$  and phases  $\phi$ , given the  $\kappa$ -length observation sequence  $Y_\kappa$ .

2) *Detection*: Detect the pulse-train sequence  $s_k$  given the observation sequence  $Y_k$ ,  $1 \leq k \leq \kappa$  and thereby deinterleave the pulse train. Equivalently, obtain an optimal, say maximum a posteriori (MAP), estimate of  $X_k$  denoted  $\hat{X}_k^{\text{MAP}} \in \{e_1, \dots, e_N\}$ .

### C. Formulation as Dynamic Linear Model

We now formulate the deinterleaving problem as a DLM suitable for estimation by a KF.

Let  $\mathcal{X}_\kappa$  denote the sequence  $(X_1, \dots, X_\kappa)$ . There are  $N^\kappa$  possible source sequences. Let  $S_\kappa^1, \dots, S_\kappa^{N^\kappa}$  denote all these

possible source sequences, termed paths, of length  $\kappa$ . So  $S_\kappa^p = (s_1^p, \dots, s_\kappa^p)$  with  $p \in [1, N^\kappa]$  and  $X_k^p = e_i$  if  $s_k^p = i$ .

Let  $\tau_k^i$  denote the last time source  $i$  was active up to and including the arrival of the  $k$ th pulse. (We initialize  $\tau_k^i$  by setting it to  $\phi^i$  until source  $i$  becomes active). We can express  $\tau_k^i, k \in I^+$  by the following dynamical equation

$$\tau_{k+1}^i = \begin{cases} \tau_k^i + T^i, & \text{if } k+1 \text{th pulse is due to source } i \\ \tau_k^i, & \text{otherwise} \end{cases} \quad (2.3)$$

$$\tau_1^i = \phi^i.$$

Define the  $R^N$  vectors

$$\tau_k' = (\tau_k^1, \dots, \tau_k^N), \quad T' = (T^1, \dots, T^N), \quad (2.4)$$

$$\phi' = (\phi^1, \dots, \phi^N).$$

We can write (2.3) as

$$\tau_{k+1} = \tau_k + \text{diag}(X_{k+1})T'$$

$$\tau_1 = \phi'$$

$$t_k = \tau_k' X_k \quad (2.5)$$

where for any  $R^N$  vector  $z = (z_1, \dots, z_N)$ ,  $\text{diag}(z) = \text{diag}(z_1, \dots, z_N)$  and  $T, \tau_k$  are defined in (2.4).

Define the "state vector"  $x_k$  as

$$x_k = \begin{pmatrix} T \\ \tau_k \end{pmatrix}, \quad x_1 = \begin{pmatrix} T \\ \phi \end{pmatrix}. \quad (2.6)$$

Then  $x_k$  can be expressed in standard state-space form as the following DLM

$$x_{k+1} = F(X_{k+1})x_k + G(X_k)v_k, \quad v_k \sim N[0, Q]$$

$$y_k = H'(X_k)x_k + w_k, \quad w_k \sim N[0, R]. \quad (2.7)$$

Here  $v_k$  is an independent white-Gaussian noise process with variance  $Q = \sigma_v^2, R = \sigma_w^2, X_k = e_i$  if pulse  $k$  is due to source  $i$ , and

$$F(X_{k+1}) = \begin{pmatrix} I_N & 0_{N \times N} \\ \text{diag}(X_{k+1}) & I_N \end{pmatrix},$$

$$H'(X_k) = (0_{1 \times N} \quad X_k'), \quad G(X_k) = 0_{2N \times 1}. \quad (2.8)$$

In (2.8),  $0_{M \times N}$  denotes a matrix of dimension  $M \times N$  with elements zero and  $I_N$  is the identity matrix with dimension  $N \times N$ . For a given source sequence  $\{X_k\}$ , Kalman filtering (smoothing) can be used to estimate  $x_k(x_1)$  as we shall describe in Section II-E.

A related useful model formulation is (2.7) with

$$x_k = \begin{pmatrix} T \\ \phi \\ \tau_k - \phi \end{pmatrix}, \quad x_1 = \begin{pmatrix} T \\ \phi \\ 0 \end{pmatrix} \quad (2.9)$$

$$F(X_{k+1}) = \begin{pmatrix} I_N & 0_{N \times N} & 0_{N \times N} \\ 0_{N \times N} & I_N & 0_{N \times N} \\ \text{diag}(X_{k+1}) & I_N & I_N \end{pmatrix}$$

$$H(X_k) = (0_{1 \times N} \quad X_k' \quad X_k')$$

$$G(X_k) = 0_{3N \times 1}. \quad (2.10)$$

For a given sequence, KF estimates of  $T, \phi$ , and  $\tau$  can be obtained using this model.

1) *Models with Pulse Energy Information:* If the pulses contain energy (e.g., amplitude) information of the source that generated them, then the above DLM can be augmented to include this information as follows: Let  $a_i$  denote the amplitude of pulses generated from the  $i$ th source. Define  $a' = (a_1, \dots, a_N)$ . Each observation now is a vector  $y_k$  consisting of WGN noise-corrupted measurements of the TOA and pulse amplitude. Thus the DLM can be expressed as (2.7) with  $x_k = (T', \tau', a')', G(X_k) = 0_{3N \times 1}$

$$\begin{aligned} F(X_{k+1}) &= \begin{pmatrix} I_N & 0_{N \times N} & 0_{N \times N} \\ \text{diag}(X_{k+1}) & I_N & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} & I_N \end{pmatrix} \\ H'(X_{k+1}) &= \begin{pmatrix} 0_{1 \times N} & X'_k & 0_{1 \times N} \\ 0_{1 \times N} & 0_{1 \times N} & X'_k \end{pmatrix} \end{aligned} \quad (2.11)$$

where  $R$  is diagonal.

For the rest of the paper we shall tackle that harder problem where pulse amplitude and width information are not available.

2) *Stochastic Sources:* It is also possible to consider more general DLM's in (2.7) when  $G(X_k)v_k \neq 0$ . The driving noise  $v_k$  allows modeling of sources that are active at random time instants or have randomly varying amplitudes. For example, the model allows for time-varying or stochastically varying periodic sources by replacing  $I_N$  in the  $(1, 1)$  block of  $F(X_{k+1})$  in (2.8) with  $\alpha I_N$  for some scalar  $\alpha \geq 0$  and setting  $G(X_k) \neq 0$  and  $\sigma_v^2 \neq 0$ . If  $\alpha > 1$  then the periods are larger, if  $\alpha = 1$  the periods remain constant, and if  $\alpha < 1$  the periods get smaller.

#### D. Parameterized Dynamics Linear Model

We now present an alternative formulation of the deinterleaving problem as a parameterized DLM suitable for estimation via RLS.

It is possible to re-express the DLM (2.7), (2.8) as follows

$$\begin{aligned} z_{k+1} &= z_k + \text{diag}(X_{k+1}) \\ y_k &= X'_k(z_k - I) \begin{pmatrix} T \\ \phi \end{pmatrix} + w_k \end{aligned} \quad (2.12)$$

where  $z_k \triangleq \sum_{t=2}^k \text{diag}(X_t)$ . So for a given source sequence  $\mathcal{X}_k$ , (2.12) can be expressed as

$$y_k = \eta'_k(\mathcal{X}_k)\theta + v_k \quad (2.13)$$

where  $\eta'_k(\mathcal{X}_k) = X'_k(z_k - I)$  is path dependent and corresponds to the familiar regression vector. Also  $\theta = (T' \ \phi')'$ . The measurements  $y_k$  are linear in the unknown parameter  $\theta$ . Thus recursive least squares (RLS) can be used to estimate the parameter  $\theta$  in (2.13) for a given source sequence  $\{\mathcal{X}_k\}$ ; see Section II-E. The RLS algorithm is independent of the noise variance  $\sigma_w^2$ .

#### E. Path-Dependent Estimators

We now show how to estimate the DLM states and parameters using either a KF or RLS for a given source sequence (path).

1) *Kalman Filter Estimator:* For a source sequence  $\mathcal{X}_k = S_k^p$ , denoted as path  $p$ , let  $F_k^p, G_k^p$ , and  $H_k^p$  denote the associated DLM matrices in (2.7). Given the noisy observations  $Y_k$ , let

$$\begin{aligned} x_{k+1|k}^p &= E\{x_{k+1}|Y_k, S_k^p\} \\ \Sigma_{k+1|k}^p &= E\{(x_{k+1} - x_{k+1|k}^p)(x_{k+1} - x_{k+1|k}^p)' | S_k^p\} \end{aligned} \quad (2.14)$$

be the predicted state estimate and the predicted state-covariance estimate at time  $k+1$  given the path  $p$ . Notice that the covariance estimate is independent of the observations.

Denote the one-step output prediction error at time  $k+1$  given the path  $p$  as  $e_{k+1}$ . The KF for the DLM (2.7) conditioned on the path  $p$  and observations  $Y_k$  is as follows

$$\begin{aligned} x_{k+1|k}^p &= F_{k+1}^p x_{k|k-1}^p + K_k^p e_k^p \\ \Sigma_{k+1|k}^p &= F_{k+1}^p \left( \Sigma_{k|k-1}^p - \Sigma_{k|k-1}^p H_k^p [H_k^p \Sigma_{k|k-1}^p H_k^p + R]^{-1} \right. \\ &\quad \left. \cdot H_k^p \Sigma_{k|k-1}^p \right) F_{k+1}^p G_k^p Q G_k^p \\ K_k^p &= F_{k+1}^p \Sigma_{k|k-1}^p H_k^p (H_k^p \Sigma_{k|k-1}^p H_k^p + R)^{-1} \\ e_{k+1}^p &= y_{k+1} - H_{k+1}^p x_{k+1|k}^p, \quad k \in [1, \dots, \kappa] \end{aligned} \quad (2.15)$$

initialized with *a priori* estimates  $x_{1|0}^p$  and  $\Sigma_{1|0}^p$ .

2) *Recursive Least Squares Estimator:* Let  $\eta_k^p$  (see 2.13) denote the "regression" vector associated with the path  $p$ . Let  $\theta_k^p$  and  $P_k^p$ , denote, respectively, the RLS parameter estimate of  $\theta$  and the covariance matrix based on path  $p$ . Thus

$$P_k^p = E\{(\theta_k^p - \theta)(\theta_k^p - \theta)' | S_k^p\} \quad (2.16)$$

which is independent of the observations  $Y_k$ .

Denote the one-step output prediction error at time  $k+1$  given the path  $p$  as  $e_{k+1}$ . The RLS estimator for the parameterized DLM (2.12) conditioned on the path  $p$  and observations  $Y_k$  is given by

$$\begin{aligned} \theta_{k+1}^p &= \theta_k^p + P_{k+1}^p \eta_{k+1}^p e_{k+1}^p \\ P_{k+1}^p &= P_k^p - P_k^p \eta_{k+1}^p [I + \eta_{k+1}^p P_k^p \eta_{k+1}^p]^{-1} \\ &\quad \cdot \eta_{k+1}^p P_k^p = \left[ \sum_{t=2}^{k+1} \eta_t \eta_t' + P_1^{-1} \right]^{-1} \\ e_{k+1}^p &= y_{k+1} - \eta_{k+1}^p \theta_k^p, \quad k \in [1, \dots, \kappa] \end{aligned} \quad (2.17)$$

initialized with *a priori* estimates  $\theta_1^p$  and  $P_1^p$ . In a Bayesian context, we have from Sternby [8] that if  $P_k^p \rightarrow 0$  as  $k \rightarrow \infty$ , then  $\theta_k^p \rightarrow \theta$  a.s. with  $(\theta_k^p - \theta)'(\theta_k^p - \theta) = O(P_k^p)$ , see also [9]. The parameterized DLM (2.12) gives us useful insight into the *persistence of excitation* conditions on  $\mathcal{X}_k$  required for RLS parameter estimation: For  $\sum_{t=2}^k \eta_t \eta_t'$  to have full rank we need at least two pulses from each source. Useful suboptimal but asymptotically optimal estimation can be achieved by working with block-diagonal versions of  $\sum_{t=2}^k \eta_t \eta_t'$ .

3) *Remarks:* a) The RLS estimator cannot be used to estimate stochastic sources or sources with time-varying periods without introducing forgetting factors or working with the appropriate KF formulation.

b) If the measurement jitter is modeled as colored noise using an augmented model, then an augmented KF can be applied. If the colored-noise model has unknown parameters then an extended Kalman filter (EKF) (see [7], pp. 296–301) would have to be used instead of the KF. In the least-squares case, an extended least-squares (ELS) scheme (see [7], pp. 279–286) could be used instead of RLS.

#### F. Prediction Error Cost for a Sequence

The noise sample-path dependent KF or RLS one-step prediction-error cost of the path  $p$  denoted as  $J_\kappa^p$  is calculated as

$$J_\kappa^p = \sum_{k=1}^{\kappa} (e_k^p)^2 \quad (2.18)$$

where  $e_k$  is defined in (2.15) for the KF estimator and in (2.17) for the RLS estimator. Also  $E\{J_\kappa^R\} = \sum_{k=1}^{\kappa} (H'(X_k) \Sigma_{k|k-1}^p H(X_k) + R)$ .

Of course, in the case (2.7), (2.8), the costs  $J_\kappa^p$  and  $E\{J_\kappa^p\}$  are identical to those associated with the RLS parameter estimation of the model (2.13).

#### G. Optimal Scheme

Given a sequence of observations  $Y_\kappa$ , evaluate sample-path one-step prediction-error costs of all  $N^\kappa$  possible source sequences (paths)  $S_\kappa^p, p = 1, \dots, N^\kappa$ . Pick the optimal (MAP) path  $p^*$  as

$$p^* = \arg \min_p J_\kappa^p \quad (2.19)$$

where  $J_\kappa^p$  is defined in (2.18). Thus we have the optimal sequence  $S_\kappa^{p^*}$  with  $\{\hat{X}_k^{\text{MAP}}\} = \{X_k^{p^*}\}$ .

Of course the KF on path  $p^*$  yields filtered estimates of the state and hence of  $T$ . By running a Kalman smoother on the path  $p^*$  we can then evaluate  $E\{x_1|Y_\kappa, S_\kappa^{p^*}\}$  and so obtain fixed-interval smoothed estimates of the source periods  $T$  and source phases  $\phi$ .

Similarly, the RLS estimator on path  $p^*$  yields optimal estimates of the periods and phases.

#### H. Computational Cost

1) *Kalman Filter*: Because of the structure of  $F_{k+1}^p$  and  $H_k^p$ , the computational cost for implementing (2.15) can be reduced from the usual cost of  $O(N^3)$  to  $O(N^2)$ . This is because computing  $\Sigma_{k|k-1}^p H_k^p H_k^{p'} \Sigma_{k|k-1}^{p'}$  requires  $O(N^2)$  multiplications; all the other computations in (2.15) require additions. So the total computational cost is  $O(N^{\kappa+2})$ , which makes a full-tree search impractical for large  $N$  or  $\kappa$ .

2) *RLS Estimator*: The computational cost for implementing (2.17) at each instant  $k$  is  $O(N^2)$ . So the total cost for a full tree search is  $O(N^{\kappa+2})$ .

#### I. Estimation of a Single Source

In the special case when  $N = 1$ , the actual source sequence is known:  $X_k = e_1, \forall k$ . Then the KF (2.15) or least-squares estimator (2.17) with  $p = 1$  yields optimal filtered

estimates of the source period. Also, by using a fixed-interval Kalman smoother, smoothed estimates of the phase can be obtained. We illustrate the estimation of a single source in simulation studies.

#### III. FORWARD DYNAMIC PROGRAMMING WITH LOOK-AHEAD

In this section we present our first suboptimal solution to the deinterleaving problem. It combines the optimal full-tree search algorithm proposed in Section II-D over short segments  $\Delta$  at each pulse  $k$  and then use FDP to maintain the best  $N$  paths (in terms of prediction-error costs) prior to and including pulse  $k$  where each path terminates at a different source. We choose to work with the KF prediction errors. However, the RLS prediction errors can also be used. Similar schemes have also been used for tree pruning in [5].

Let  $r_k^i$  denote the KF prediction-error cost of the best sequence (i.e., the sequence with minimum error cost among all sequences) ending in source  $i$  at time  $k$  based on a  $\Delta$  look-ahead. Similarly  $x_{k|k-1}^i, \Sigma_{k|k-1}^i$  denotes the KF predicted state and state covariance, respectively, at time  $k$  based on the best sequence ending in source  $i$  at time  $k$  with a  $\Delta$  look-ahead.

*Initialize*: Set  $\Sigma_{1|0}^i, x_{1|0}^i, r_1^i = 0.0$  for  $i = 1, \dots, N$ .

*Iterations*: Given  $\Sigma_{k|k-1}^i, x_{k|k-1}^i, r_k^i, \varphi_k^i, i = 1, \dots, N$ .

Step 1.  *$\Delta$  look ahead to eliminate unlike sequences*: For  $p = 1$  to  $N^{\Delta+1}$ , evaluate prediction-error costs  $J_{k+1, k+\Delta}^p$  for each  $\Delta + 1$  length sequence  $s_k^p, \dots, s_{k+\Delta}^p$ . Here

$$J_{k+1, k+\Delta}^p = \sum_{t=k+1}^{k+\Delta} (e_t^p)^2 \quad (3.1)$$

and is computed based on the observations  $y_k, \dots, y_{k+\Delta}$ .

Partition these costs into  $N^2$  subsets according to whether  $s_k^p = i, s_{k+1}^p = j, i = 1, \dots, N, j = 1, \dots, N$ . Let  $c_{ij}$  be the minimum cost in each partition.

Step 2. *Forward dynamic programming*: Using the  $N^2$  best costs  $c_{ij}$  to Step 1, we obtain via FDP the best sequence ending in source  $j$  at time  $k+1$  for each  $j, j = 1, \dots, N$

$$\begin{aligned} r_{k+1}^j &= \min_i c_{ij} + r_k^i; \quad i, j \in [1, \dots, N] \\ \varphi_{k+1}^j &= \arg \min_i c_{ij} + r_k^i. \end{aligned} \quad (3.2)$$

Here,  $\varphi_{k+1}^j$  denotes the active source at time  $k$  of the best sequence ending in source  $j$  at time  $k+1$ .

Step 3. *Kalman filter update*: For  $j = 1, \dots, N$  run the following  $N$  Kalman filters in parallel

$$\begin{aligned} x_{k+1|k}^j &= F(e_j) x_{k|k-1}^{\varphi_{k+1}^j} + K_k^j (y_k - H'(\varphi_{k+1}^j) x_{k|k-1}^{\varphi_{k+1}^j}) \\ \Sigma_{k+1|k}^j &= F(e_j) \left( \Sigma_{k|k-1}^{\varphi_{k+1}^j} - \Sigma_{k|k-1}^{\varphi_{k+1}^j} H'(\varphi_{k+1}^j) \left[ H(\varphi_{k+1}^j)' \Sigma_{k|k-1}^{\varphi_{k+1}^j} \right. \right. \\ &\quad \cdot \left. \left. H(\varphi_{k+1}^j) + R \right]^{-1} H'(\varphi_{k+1}^j) \Sigma_{k|k-1}^{\varphi_{k+1}^j} \right) F'(e_j) \\ &\quad + G(\varphi_{k+1}^j) Q G'(\varphi_{k+1}^j) \\ K_k^j &= F(e_j) \Sigma_{k|k-1}^{\varphi_{k+1}^j} H'(\varphi_{k+1}^j) \left[ H(\varphi_{k+1}^j)' \Sigma_{k|k-1}^{\varphi_{k+1}^j} \right. \\ &\quad \cdot \left. \left. H(\varphi_{k+1}^j) + R \right]^{-1} \right) F'(e_j). \end{aligned} \quad (3.3)$$

Actually,  $x_{k+1|k}^j$  and  $\Sigma_{k+1|k}^j$  are computed during Step 1, we present Step 3 for clarity.

Step 4. Set  $k$  to  $k+1$  and repeat beginning at Step 1 until  $k = \kappa - 1$ . If  $k = \kappa - 1$ , then  $i^* = \arg \min_i r_{\kappa}^i$ .

1) *Backtracking*: Start with  $X_{\kappa}^* = e_{i^*}$ . For  $k = \kappa - 1, \dots, 1$ : If  $X_{k+1}^* = e_i, i \in [1, \dots, N]$  then  $X_k^* = e_{\varphi_{k+1}^i}$ . Thus  $X_k^*, k = 1, \dots, \kappa$  is the estimated sequence of active sources.

Again the KF estimate at time  $\kappa$  from the optimal path  $X_k^*$  yields filtered estimates of  $T$ . Also a fixed-interval Kalman smoother on path  $X_k^*$  yields smoothed estimates of  $x_1$  and hence  $T, \phi$ .

2) *Computation Complexity*: For a look-ahead of  $\Delta$ , we need to run Kalman filters on  $N^{\Delta+1}$  sequences at each pulse instant. Now the computational requirement for each sequence is  $\Delta O(N^2)$  (see Section II-E). So the total complexity at each pulse instant is  $\Delta O(N^{\Delta+3})$ .

*Choice of  $\Delta$* : We are unable to give a comprehensive design rule for selecting  $\Delta$ . However, in simulations studies we show that choosing  $3 \leq \Delta \leq 5$  usually yields satisfactory estimates. It may be possible to use a similar analysis to that in [5] to compute the probability of losing a source in terms of  $\Delta$ .

#### IV. PROBABILISTIC TEACHER

Our second suboptimal scheme assumes that the sources are active at independent random time instants. Of course, if the sources are known to be periodic, then the probability of the source being active is proportional to its pulse frequency.

The probabilistic teacher (PT) makes a random decision as to which source is active at each time instant as follows: Let  $\hat{X}_k$  be the estimated active source at time  $k$  using the PT. Also let  $\hat{\mathcal{X}}_k$  denote the sequence of past decisions  $\hat{X}_1, \dots, \hat{X}_k$  from the PT based on the observations  $Y_k$ . Then with  $y_{k+1}$ , the PT decides which source is active at time  $k+1$  by selecting one source at random according to the *a posteriori* probability density of the sources at time  $k+1$ .

Even though the decision as to which source is active is made at random, on the average the density function used in the estimation is the same as the correct density for learning the values of the parameters [6]. This means that the PT scheme is statistically the same as learning without a teacher. Convergence of the PT scheme is proved in [6] under mild conditions. The main condition for convergence is that the possibility of reaching the correct value starting from the initial prior distribution of parameters is not ruled out.

We now present the PT approach to deinterleaving in detail: Let  $\pi$  denote the *a priori* probability vector of the sources:  $\pi_i = P(X_k = e_i)$ . Let  $\hat{\pi}$  denote our estimate of  $\pi$ .

Assumption: The sources are independent, i.e.,  $P(X_k = e_i | \mathcal{X}_{k-1}) = P(X_k = e_i) = \pi_i$ .

Then, given a sequence of observations  $Y_{\kappa}$ , the PT performs deinterleaving as follows:

*Initialize*: Set  $\Sigma_{1|0}, x_{1|0}$ .

*Iterations*: Given  $\Sigma_{k|k-1}, x_{k|k-1}$ .

Step 1. *Supervised learning*:

1) Compute the *a posteriori* probability density  $f(X_{k+1} = e_j | y_{k+1}, Y_k, \hat{\mathcal{X}}_k), j = 1, \dots, N$ . We do this as follows: from Bayes rule

$$\begin{aligned} f(X_{k+1} = e_j | y_{k+1}, Y_k, \hat{\mathcal{X}}_k) \\ = \frac{f(y_{k+1} | X_{k+1} = e_j, Y_k, \hat{\mathcal{X}}_k) f(X_{k+1} = e_j | Y_k, \hat{\mathcal{X}}_k)}{\sum_{i=1}^N f(y_{k+1} | X_{k+1} = e_i, Y_k, \hat{\mathcal{X}}_k) f(X_{k+1} = e_i | Y_k, \hat{\mathcal{X}}_k)} \end{aligned} \quad (4.1)$$

Let us now evaluate the terms on the RHS of (4.1). Because the sources are assumed independent,  $f(X_{k+1} = e_j | Y_k, \hat{\mathcal{X}}_k) = f(X_{k+1} = e_j) = \hat{\pi}_j$ . The density  $f(y_{k+1} | X_{k+1} = e_j, Y_k, \hat{\mathcal{X}}_k)$  is very difficult to calculate due to the time-varying nature of the DLM. So we approximate this density by a Gaussian density that has the same first two moments [10] to obtain

$$\begin{aligned} f(y_{k+1} | X_{k+1} = e_j, Y_k, \hat{\mathcal{X}}_k) \sim N[H(e_j)x_{k+1|k}^j, \Omega_{k+1}^j], \\ \Omega_{k+1}^j = H(e_j)' \sum_{k+1|k}^j H(e_j) + R. \end{aligned} \quad (4.2)$$

Here  $x_{k+1|k}^j$  and  $\Sigma_{k+1|k}^j, j = 1, \dots, N$  are computed using the KF equations as

$$\begin{aligned} x_{k+1|k}^j &= F(e_j)x_{k|k-1} + K_k^j(y_k - H'(\hat{X}_k)x_{k|k-1}) \\ \Sigma_{k+1|k}^j &= F(e_j) \left( \Sigma_{k|k-1} - \Sigma_{k|k-1} H'(\hat{X}_k) \right. \\ &\quad \cdot \left[ H'(\hat{X}_k) \Sigma_{k|k-1} H'(\hat{X}_k) + R \right]^{-1} \\ &\quad \cdot H'(\hat{X}_k) \Sigma_{k|k-1} \Big) F'(e_j) \\ &\quad + G(\hat{X}_k) Q G'(\hat{X}_k) \\ K_k^j &= F(e_j) \Sigma_{k|k-1} H'(\hat{X}_k) \left[ H'(\hat{X}_k) \Sigma_{k|k-1} \right. \\ &\quad \cdot H'(\hat{X}_k) + R \Big]^{-1}. \end{aligned} \quad (4.3)$$

2) Set  $\hat{X}_{k+1} = e_j$  for a randomly chosen  $\hat{j}, \hat{j} \in [1, \dots, N]$ , according to the *a posteriori* density  $f(\hat{X}_{k+1} = e_j | y_{k+1}, Y_k, \hat{\mathcal{X}}_k)$ .

Step 2. *Kalman filter update*: For  $\hat{X}_{k+1} = e_j$ , the PT state and covariance updates are

$$x_{k+1|k} = x_{k+1|k}^{\hat{j}}, \quad \Sigma_{k+1|k} = \Sigma_{k+1|k}^{\hat{j}} \quad (4.4)$$

where  $x_{k+1|k}^{\hat{j}}$  and  $\Sigma_{k+1|k}^{\hat{j}}$  are computed in (4.3).

Step 3. Set  $k$  to  $k+1$  and repeat beginning at Step 1, until  $k = \kappa - 1$ .

1) *Computational Complexity*: Because  $\Sigma_{k+1|k}^j, j = 1, \dots, N$  is required to be computed, the computational complexity is  $O(N^3)$  at each instant  $k$ . Thus, compared to the FDP with look-ahead, the PT scheme is particularly attractive when there are 2 large number of sources.

2) *Known a priori Probabilities*: If the relative frequencies of the periodic sources are accurately known (with the actual frequencies and periods not known) then the *a priori* probabilities  $\hat{\pi}$  can be accurately computed from these relative

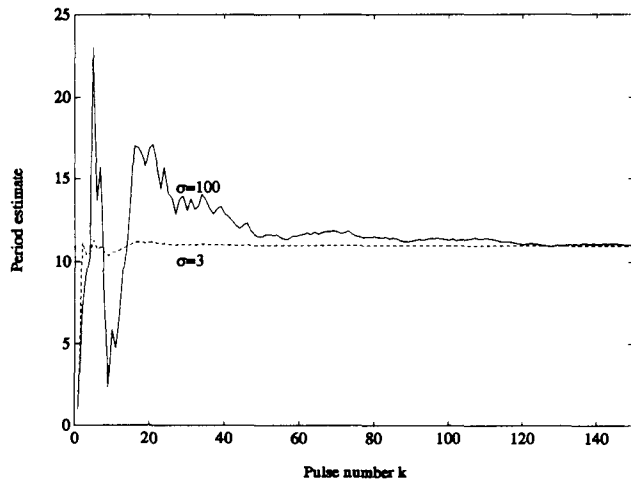


Fig. 3. Period estimate of single source in medium and large jitter noise. Parameters:  $N = 1$ ,  $T = 11$ ,  $\phi = 50$ . Noise variances:  $\sigma_w^2 = 9$  and  $1 \times 10^4$ . Initial estimate:  $x'_{1|0} = (1.1)$ ,  $\Sigma_{1|0} = 10^3 \times I_2$ .

frequencies. Simulations show that in such a case, instead of using the *a posteriori* probabilities in Step 1, one can use the *a priori* probability estimate  $\hat{\pi}$  to get excellent estimates of  $T$ . Of course,  $x'_{k+1|k}$  and  $\Sigma'_{k+1|k}$ ,  $j = 1, \dots, N$  do not have to be computed, only  $x'_{k+1|k}$  and  $\Sigma'_{k+1|k}$  are required in Step 2. So the computational requirement is  $O(N^2)$  at each pulse instant.

Thus, if the statistics of the sample path (the source probabilities) are known, the periods can be estimated despite the fact that the actual source sequence is unknown.

## V. SIMULATION EXAMPLES

We first consider the simplest case of a single periodic pulse train in jitter noise. Then the performance of the two suboptimal schemes proposed in Sections III and IV, i.e., FDP with look-ahead and PT are studied.

### A. Estimation of a Single Source

We illustrate the optimal performance of the scheme in Section II-F for estimating the period of a single periodic pulse train in jitter noise.

To a computer-generated periodic pulse train  $T = 11$ ,  $\phi = 50$  was added jitter noise with variances  $\sigma_w^2 = 9$  and  $10^4$ , respectively. The estimation scheme was initialized with  $x'_{1|0} = (1.1)$ ,  $\Sigma_{1|0} = 10^3 \times I_2$ . Fig. 3 shows the evolution of the KF period estimate given by  $x_{k+1|k}(1)$ . Notice that when  $\sigma_w = 3$ , the period estimates converge to the true value in less than 30 pulse instants. When the jitter noise is extremely large ( $\sigma_w = 100$ ), the estimates are close to the true values after 80 pulse instants. After 100 pulse instants, the period estimate is 11.04.

*Discussion:* Besides being of independent interest, the single source example illustrates that for general interleaved pulse trains, if the source sequence is known, then the pulse trains can be estimated optimally using a KF.

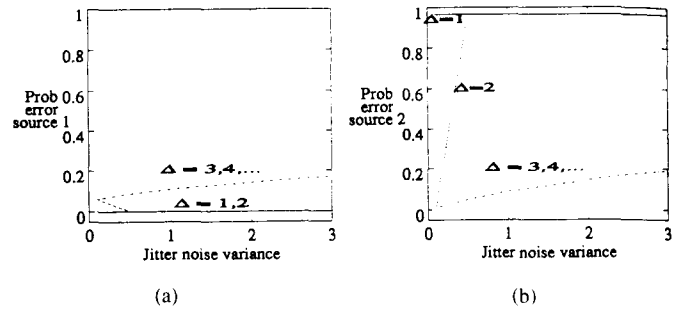


Fig. 4. Source error probabilities.

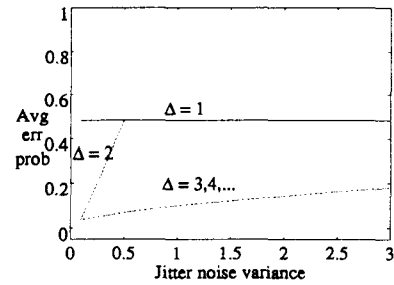


Fig. 5. Average error probabilities.

### B. FDP with Look-Ahead

1) *Error Probabilities:* Here, the same interleaved pulse train as in the example of Section II is used:  $T' = [11, 17]$ ,  $\phi' = [3, 4]$ . Initial estimates were chosen as  $x'_{1|0} = (10, 22, 1, 2)$ ,  $\Sigma_{1|0} = I_4$ , (i.e., the  $4 \times 4$  identity matrix). For various jitter-noise variances  $\sigma_w^2 \in [0.1, 3]$ , we ran the FDP with look-ahead  $\Delta = 1, \dots, 5$ .

Let us define the error probability of a source  $i$  as

$$P_e^i = \frac{\text{number of undetected pulses from source } i}{\text{number of pulses from source } i}. \quad (5.1)$$

Fig. 4 shows how  $P_e^1$  and  $P_e^2$  depend on  $\sigma_w^2$  and  $\Delta$ . For  $\Delta = 1$ , i.e., FDP with no look-ahead, Source 2 is never tracked. All pulses are detected as due to Source 1. Therefore, for  $\Delta = 1$ ,  $P_e^1 \approx 0$  and  $P_e^2 \approx 1$ .

For  $\Delta = 2$  the algorithm tracks both sources for low jitter noise,  $\sigma_w^2 < 0.5$ . For larger  $\sigma_w^2$ , Source 2 is not tracked and all pulses are detected as due to Source 1, i.e.,  $P_e^1 \approx 0$  and  $P_e^2 \approx 1$ .

For  $\Delta = 3, 4, 5, \dots$ , both sources are tracked with no noticeable improvement in performance for  $\Delta > 3$ . As expected  $P_e^1, P_e^2$  increase with  $\sigma_w^2$ .

Fig. 5 summarizes the information in Fig. 4 by showing the average error probabilities,  $\frac{1}{2}(P_e^1 + P_e^2)$ .

2) *Estimation of Two Sources:* An interleaved pulse train with jitter noise was generated with parameters  $N = 2$ ,  $\sigma_w^2 = 9.0$ ,  $T' = (11, 80)$ ,  $\phi' = (3, 4)$ . Initial estimates were chosen as  $x'_{1|0} = (60, 96, 2, 1)$ ,  $\Sigma_{1|0} = I_4$ .  $\Delta$  was taken as 3. Fig. 6 shows the evolution of the KF estimates of  $T^1$  and  $T^2$ , which are given by  $x_{k+1|k}(1)$  and  $x_{k+1|k}(2)$ , respectively.

Notice from Fig. 6 that despite  $\sigma_w^2$  being relatively large, accurate estimates of  $T^1$  and  $T^2$  are obtained after about 300 pulse instants.

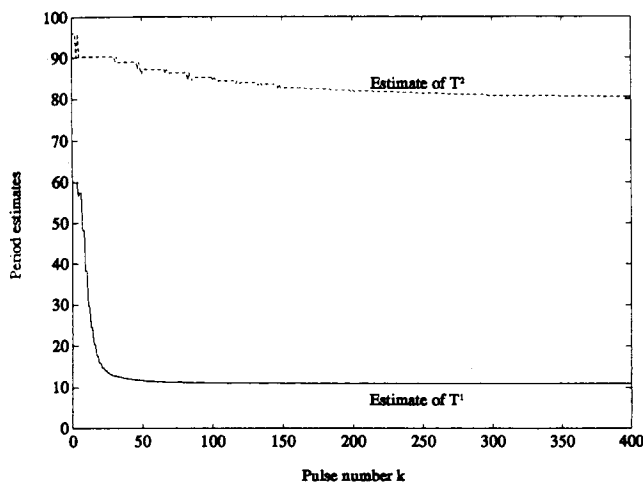


Fig. 6. Evolution of period estimates using FDP with look-ahead. Parameters:  $N = 2, \sigma_w^2 = 9.0, T' = (11, 80), \sigma' = (3, 4)$ . Initial estimate:  $x'_{1/0} = (60, 96, 2, 1)$ .

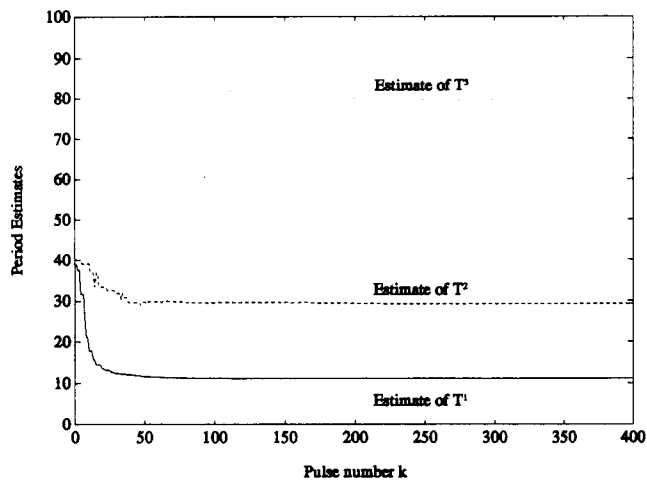
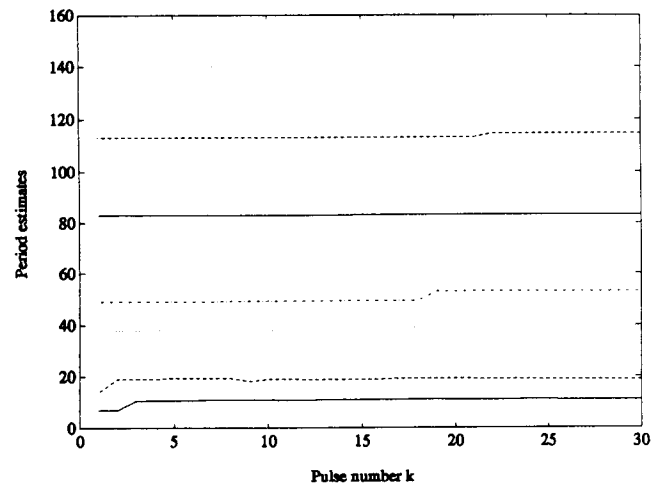


Fig. 7. Evolution of period estimates using FDP with look-ahead. Parameters:  $N = 3, \sigma_w^2 = 9.0, T' = (11, 30, 80), \phi' = (3, 4, 6)$ . Initial estimate:  $x'_{1/0} = (39, 40, 90, 3, 4, 2)$ .

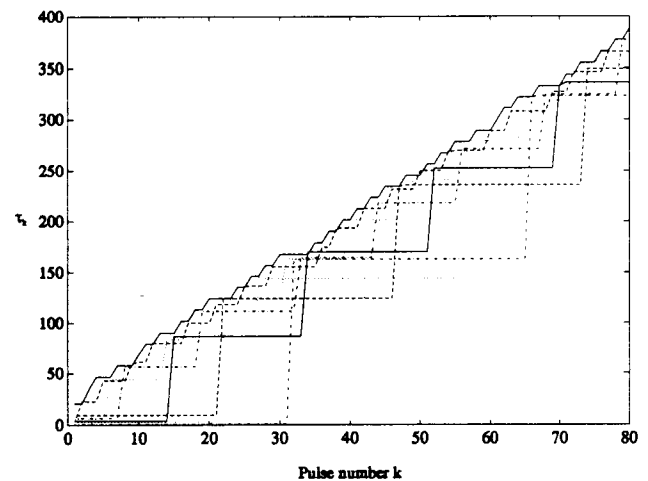
3) *Estimation of Three Sources:* An interleaved pulse train with jitter noise was generated with parameters  $N = 3, \sigma_w^2 = 9.0, T' = (11, 30, 80), \phi' = (3, 4, 6)$ . Initial estimates were chosen as  $x'_{1/0} = (39, 40, 90, 3, 4, 2)$ ,  $\Sigma_{1/0} = I_6$ . Fig. 7 shows the evolution of the KF estimates of  $T^1, T^2, T^3$ . Again after about 300 pulse instants the estimates approach to the true values.

4) *Estimation of Eight Sources:* An interleaved pulse train with jitter noise was generated with parameters  $N = 8, \sigma_w^2 = 1, T' = (11, 19, 39, 53, 83, 113, 139, 160), \phi' = (3, 4, 6, 7, 4, 10, 3, 1)$ . Initial estimates were chosen as  $x'_{1/0} = (7, 14, 38, 49, 83, 113, 140, 160, 3, 4, 6, 7, 4, 10, 3, 1)$ ,  $\Sigma_{1/0} = 10^2 \times I_{16}$ . Fig. 8 shows the evolution of the period estimates. Also shown are the estimates of  $\tau_k$  defined in (2.4). Notice that all eight sources are successfully tracked.

*Discussion:* 1) *Choice of look-ahead  $\Delta$ :* In general we found that for larger  $\sigma_w^2$ , to successfully track the sources, it is



(a)



(b)

Fig. 8. Estimation of eight sources using FDP with look-ahead. Parameters:  $N = 8, \sigma_w^2 = 1, T' = (11, 19, 39, 53, 83, 113, 139, 160), \sigma' = (3, 4, 6, 7, 4, 10, 3, 1)$ . Initial estimate:  $x'_{1/0} = (7, 14, 38, 49, 83, 113, 140, 160, 3, 4, 6, 7, 4, 10, 3, 1)$ ,  $\Sigma_{1/0} = 10^2 \times I_{16}$ .

necessary to choose a larger look-ahead  $\Delta$ . Also, the greater the separation of the periods, the smaller the  $\Delta$  necessary to track the sources.

2) *Effect of initial conditions:* Extensive simulation studies show that the smaller periods are less sensitive to initial conditions. For example, in the two-source-estimation example above, if the initial estimate of  $T^2$  is larger than 100 (i.e., 20 from the true value) then Source 2 is not tracked. In comparison, initial estimates of Source 1 could be taken up to 70 and the sources still tracked. We found that for  $\sigma_w^2 \leq 25$ , both sources were accurately estimated.

3) *Threshold effect:* Simulations shows that the FDP algorithm with look-ahead exhibits a threshold effect in that for  $\Delta < \Delta_m$  the estimates are poor and for  $\Delta \geq \Delta_m$  the estimates appear invariant of  $\Delta$  and so are optimal, or at least virtually optimal. Notice from Fig. 4 and Fig. 5 that for  $\Delta \geq \Delta_m = 3$ , the source-error probabilities do not improve. Also the period estimates do not get any better. Thus in Fig. 4 and Fig. 5, the estimates for  $\Delta \geq 3$  are "optimal."



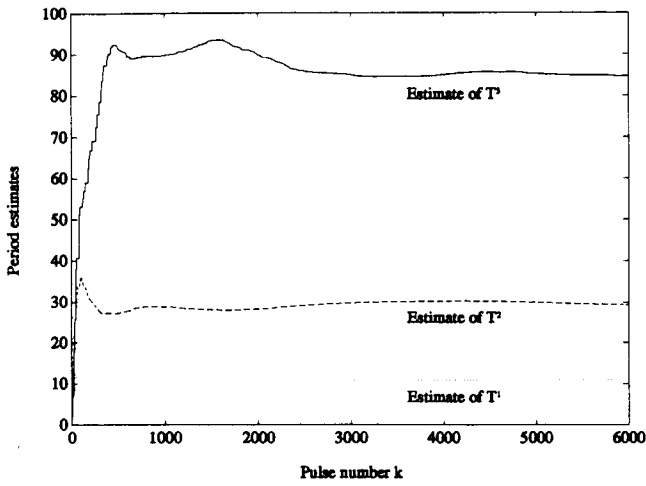


Fig. 9. Evolution of period estimates using PT with known *a priori* probabilities. Parameters:  $N = 3$ ,  $\sigma_w^2 = 36.0$ ,  $T' = (11, 30, 80)$ ,  $\sigma' = (3, 4, 6)$ . Initial estimate:  $x'_{1/0} = (1, 1, 1, 1, 1, 1)$ .

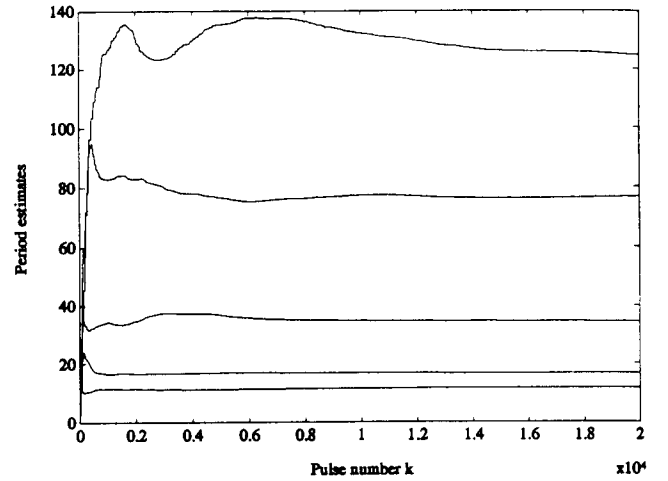


Fig. 10. Evolution of period estimates using PT with known *a priori* probabilities. Parameters:  $N = 5$ ,  $\sigma_w^2 = 64.0$ ,  $T' = (11, 19, 39, 83, 120)$ ,  $\sigma' = (3, 4, 6, 7, 4)$ . Initial estimate:  $x'_{1/0} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ .

### C. Probabilistic Teacher

The performance of the PT algorithms proposed in Section IV are now illustrated.

1) *Known a posteriori Probabilities*: Here we consider the case where the probabilities of the sources, i.e.,  $\pi_i$ ,  $i = 1, \dots, N$  are known accurately. We show that in such a case using the *a priori* probabilities instead of *a posteriori* probabilities still results in satisfactory period estimates of the sources.

2) *Estimation of Three Sources*: The same 3-source interleaved pulse train as in the three source example in Section V-A was used. We added large jitter noise to this chain,  $\sigma_w^2 = 36$ .

Assume that *a priori* probabilities  $\pi_i$ , i.e.,  $T_i/\sum_i T_i$  are known. The PT algorithm was used with these known *a priori* probabilities instead of computing *a posteriori* probabilities. Initial estimates were taken as:  $x'_{1/0} = (1, 1, 1, 1, 1, 1)$ ,  $\Sigma_{1/0} = I_6$ . Fig. 9 shows the evolution of the period estimates. Notice that despite the fact that the jitter noise is large and the initial conditions are far away from the true values, after 3000 pulse instants the estimates are close to the true values.

3) *Estimation of Five Sources*: An interleaved pulse train with jitter noise was generated with parameters  $N = 5$ ,  $\sigma_w^2 = 64.0$ ,  $T' = (11, 19, 39, 83, 120)$ ,  $\phi' = (3, 4, 6, 7, 4)$ . Initial estimates were:  $x'_{1/0} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ ,  $\Sigma_{1/0} = I_{10}$ . Fig. 10 shows the evolution of the period estimates.

*Discussion*: In general, the estimate of the larger periods converged slowly. This is because there are fewer data points from the sources with larger periods than that of the frequent sources. We found the PT algorithm using known *a priori* probabilities to be extremely robust to initial conditions, jitter noise, and missing pulses.

4) *A posteriori Probabilities*: The same 3-source interleaved pulse train as in the three-source example in Section V-A was used. Jitter noise with  $\sigma_w^2 = 1$  was added to this pulse train.

We used the PT algorithm with *a posteriori* probabilities on these data. We incorrectly assumed that  $N = 4$ . The

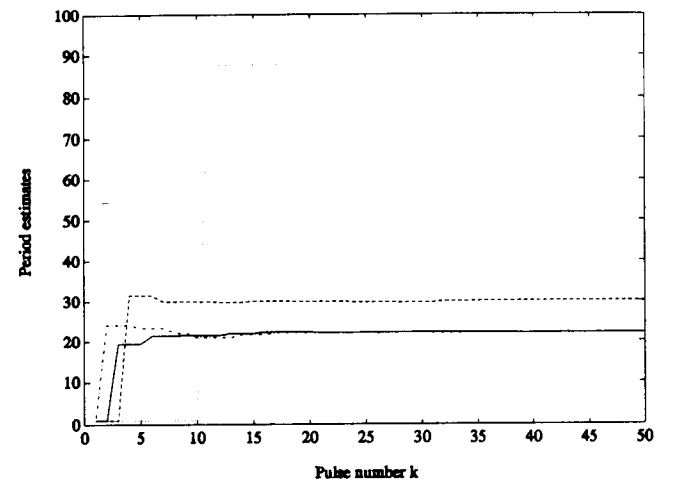


Fig. 11. Evolution of period estimates using PT with *a posteriori* probabilities. Parameters:  $N = 3$ ,  $\sigma_w^2 = 1.0$ ,  $T' = (11, 30, 80)$ ,  $\sigma' = (3, 4, 6)$ . Initial estimate:  $N = 4$ ,  $x'_{1/0} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ ,  $\Sigma_{1/0} = 10^6 \times I_8$ ,  $\pi_i = 1/4$ .

initial parameters were  $x_{1/0} = (1, 1, 1, 1, 1, 1, 1, 1)$ ,  $\Sigma_{1/0} = 10^6 \times I_8$ ,  $\pi_i = 1/4$  (so no *a priori* information was used).

Fig. 11 shows the evolution of the period estimates. Notice that the estimates of two of the sources converge to 22, i.e., a multiple of the period of the first source. The estimates of the other two sources converge to the true values.

We found in general though that the PT scheme with *a posteriori* probabilities is not robust to initial conditions. For initial conditions far away from the true parameter values, the Gaussian approximation for the *a posteriori* density is not adequate and the scheme may not necessarily recover.

### VI. ROBUSTNESS ISSUES

In Section II-A we assumed that all transmitted pulses are detected at the receiver without error. We now consider the case where if  $l$  pulses arrive simultaneously at the receiver, they are detected as a single pulse  $y_k$  (in-

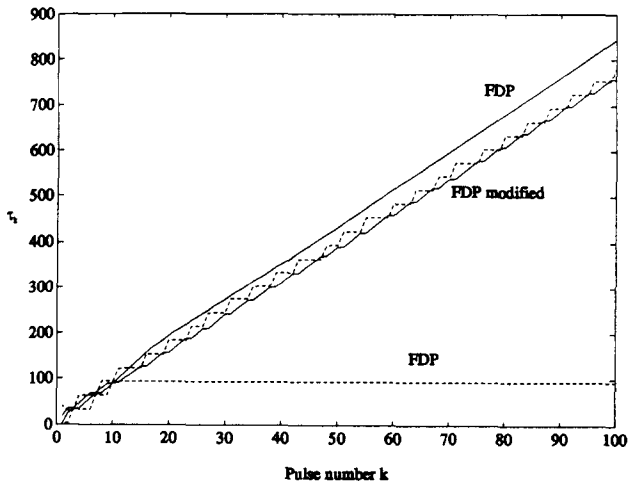


Fig. 12. State estimates using FDP in presence of missing pulses. Parameters:  $N = 2, \sigma_w^2 = 1.0, T' = (11, 30), \phi' = (3, 4)$ . Initial estimate:  $x'_{1/0} = (4, 27, 3, 4), \Sigma_{1/0} = 10^6 \times I_4$ .

stead of being detected as  $l$  pulses as in Section II-A). We give suitable modifications to the FDP algorithm with look-ahead to cope with this example of missing pulses. We also show that similar modifications can be used to deal with the case of supernumery (extra) pulses. Finally, simulation examples are presented to illustrate the performance of the FDP with look-ahead and PT schemes in the presence of missing pulses.

#### A. FDP with Look-Ahead for Missing Pulses

Recall in the FDP with look-ahead algorithm given in Section III,  $J_{k+1,k+\Delta}^p$  is computed in Step 1 from the observation sequence  $y_k, \dots, y_{k+\Delta}$ , see (3.1). To cope with the possibility of missing pulses we essentially compare the prediction errors assuming there was no missing pulse with that assuming there is a missing pulse. Consequently Step 1 and Step 4 are modified as follows:

**Modified Step 1:** For  $p = 1$  to  $N^{\Delta+1}$ , evaluate prediction error costs

$$J_{k+1,k+\Delta}^p \quad \text{based on observation sequence} \quad \{y_k, y_{k+1}, y_{k+2}, \dots, y_{k+\Delta}\} \quad (6.1)$$

$$\bar{J}_{k+1,k+\Delta}^p \quad \text{based on observation sequence} \quad \{y_k, y_k, y_{k+1}, \dots, y_{k+\Delta-1}\} \quad (6.2)$$

for each  $\Delta+1$  length sequence  $s_k^p, \dots, s_{k+\Delta}^p$ . Notice that  $J_{k+1,k+\Delta}^p$  are computed based on the assumption that there are no missing pulses at time  $k$ . Also  $\bar{J}_{k+1,k+\Delta}^p$  is computed on the assumption that there is a missing pulse at time  $k$ .

Partition the costs  $J_{k+1,k+\Delta}^p$  into  $N^2$  subsets according to if  $s_k^p = i, s_{k+1}^p = j, i = 1, \dots, N, j = 1, \dots, N$ . Let  $c_{ij}^{(1)}$  be the minimum cost in each partition.

Similarly partition the  $\bar{J}_{k+1,k+\Delta}^p$  costs into the same  $N^2$  subsets and let  $c_{ij}^{(2)}$  denote the minimum cost in each partition. Based on the prediction errors,  $J$  (no missing pulse) and  $\bar{J}$  (one

missing pulse) we make a decision whether there is a missing pulse at time  $k$  as follows:

Compare the minimum elements in  $c_{ij}^{(1)}$  and  $c_{ij}^{(2)}$ .

If  $\min_{i,j} c_{ij}^{(1)} < \min_{i,j} c_{ij}^{(2)}$ , decide that there is no missing pulse. Set  $c_{ij} = c_{ij}^{(1)}, \forall i, j$ .

Otherwise, it is decided that there is a missing pulse at time  $k$ . Set  $c_{ij} = c_{ij}^{(2)}, \forall i, j$ .

**Modified Step 4:** If it was decided in Step 1 that there was no missing pulse at time  $k$ , then set  $k$  to  $k+1$  and repeat beginning at Step 1. Otherwise go to Step 1 without incrementing  $k$ .

**1) Supernumery Pulses:** To cope with the possibility of supernumery (extra) pulses, the above procedure is slightly modified:

**Step 1:** Replace (6.2) with  $\bar{J}_{k+1,k+\Delta}^p$  based on observation sequence  $\{y_{k+1}, y_{k+2}, \dots, y_{k+\Delta+1}\}$ . This is computed based on the assumption that there is an extra pulse at time  $k$ .

**Step 4:** If it was decided in Step 1 that there was no extra pulse at time  $k$  then set  $k$  to  $k+1$  and repeat beginning at Step 1. Otherwise, go to Step 1 and set  $k$  to  $k+2$ .

**2) Simulation Example:** Here we compare the performance of the standard FDP look-ahead scheme with that proposed above in the presence of missing pulses. We generated a noisy interleaved pulse train with parameters  $N = 2, \sigma_w^2 = 1.0, T' = (11, 30), \phi' = (3, 4)$ . Also, if more than one pulse arrives simultaneously at the receiver, it is assumed that only that pulse is received at the particular pulse instant and that the other pulses are missed.

We ran both the standard FDP with look-ahead and the modified scheme proposed above on these data. We choose  $\Delta = 3$ . Initial estimates used were  $x'_{1/0} = (4, 27, 3, 4), \Sigma_{1/0} = 10^6 \times I_4$ . Fig. 12 shows the evaluation of  $\tau_k$ , where  $\tau_k$  is defined in (2.4). Notice that the standard scheme loses track of the second source after 12 pulses. The modified scheme satisfactorily tracks both sources. Also, the period estimates converged to 10.01 and 29.98, which are close to the true values.

#### B. Probabilistic Teacher for Missing Pulses

The PT scheme proposed in Section IV using known *a priori* probabilities of the sources is extremely robust to missing pulses. Recall that PT is based on the assumption that the source are random and so there is no necessity of the source being periodic. Consequently, missing pulses do not affect the performance of the PT scheme.

A noisy interleaved pulse train was generated with parameters:  $N = 3, \sigma_w^2 = 36.0, T' = (11, 30, 80), \phi' = (3, 4, 6)$ . Also, if more than one pulse arrives simultaneously at the receiver, it is assumed that only that pulse is received at the particular pulse instant. The PT algorithm with known *a priori* probabilities was run on the data with initial estimates:  $x'_{1/0} = (1, 1, 1, 1, 1), \Sigma_{1/0} = 10^6 \times I_6$ . Fig. 13 shows the evolution of the period estimates. Notice that the PT scheme is extremely robust to missing pulses.

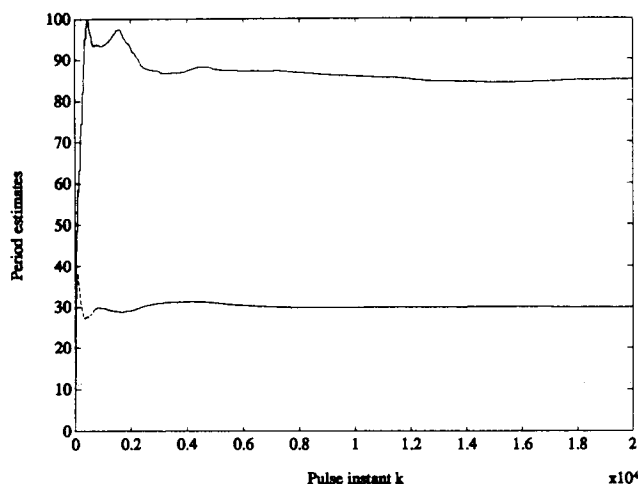


Fig. 13. Period estimates using PT with known *a priori* probabilities in presence of missing pulses. Parameters:  $N = 3$ ,  $\sigma_w^2 = 36.0$ ,  $T' = (11, 30, 80)$ ,  $\sigma' = (3, 4, 6)$ . Initial estimate:  $x'_{1/0} = (1, 1, 1, 1, 1, 1)$ ,  $\Sigma_{1/0} = 10^6 \times I_6$ .

### C. Other Robustness Issues

1) Filtered pulses: In some practical examples, the received signal from each source can be modeled as the response of a linear system to a periodic pulse train. If this linear system for each source is stable and minimum phase, then preprocessing by an inverse filter could be used to reconstruct the pulse train. Otherwise, approximate reconstruction and pulse-detection techniques could be used.

2) Non-Gaussian noise: The KF is the minimum-variance filter and so will still yield useful results when the jitter noise is non-Gaussian.

3) Change in the number of sources: A source dropout is said to occur if a source stops generating pulses. In the FDP algorithm with look-ahead, source dropout can be detected since this leads to ramping prediction errors. If the prediction errors exceed a particular threshold, then detection theory can be used to obtain the probability of a detected dropout. Also, below the threshold, detection theory gives us a probability of missed dropout. The algorithm could be reset once a dropout is detected. That is, the associated state covariance could be increased.

If a new source suddenly were to become active, again this could be detected using prediction errors and detection theory and the algorithm reset to include an additional source.

### D. Other Approaches

We have addressed two suboptimal techniques for estimating DLM's. Other techniques such as probabilistic editor, quasi-Bayes schemes ([3], pp. 214–215) are also used for estimating DLM's and may yield useful results in deinterleaving. If the source sequence is modeled as a  $N$ -state Markov chain, the observations can be formulated as a dynamic linear system driven by the Markov chain. It may be possible to use results from the theory of hidden-Markov-model estimation to achieve deinterleaving. This is one current area of research.

## VII. CONCLUSION

We have formulated the pulse train deinterleaving problem as a discrete-time stochastic dynamic linear model (DLM). The DLM allows the derivation of an optimal solution via standard techniques, although this is computationally expensive. We then proposed two suboptimal solutions that yield useful results. Simulations show that these suboptimal schemes are virtually optimal when the initial estimates of the periods and phases are reasonably close to the true values.

## REFERENCES

- [1] H. K. Mardia, "New techniques for the de-interleaving of repetitive sequences," in *IEE Proc.*, vol. 136, pt. F, no. 4, Aug. 1989.
- [2] E. Weyer, I. M. Y. Mareels, "The folded data method," Dept. of Syst. Eng., Australian National University, Report No. 2, 1991.
- [3] D. M. Titterton, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
- [4] P. J. Harrison and C. F. Stevens, "Bayesian forecasting (with discussion)," *J. Royal Statist. Soc., B*, vol. 38, pp. 205–247.
- [5] F. Gustafsson, "Estimation of discrete parameters in linear systems," *Linköping Studies in Science and Technology*. Dissertation No. 271, Dept. Elec. Eng., Linköping Univ., Sweden.
- [6] A. K. Agrawala, "Learning with a probabilistic teacher," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 4, pp. 373–379, July 1970.
- [7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [8] J. Sternby, "Consistency in least squares estimation: A Bayesian approach," *Automatica*, vol. 20, no. 4, pp. 471–475, 1984.
- [9] H. Chen, L. Guo, "Convergence rate of least squares identification and adaptive control for stochastic systems," *Int. J. Contr.*, vol. 44, pp. 1459–1476, 1986.
- [10] F. S. Chang and M. D. Srinath, "Tracking over fading channels," *IEEE Trans. Commun.*, Apr. 1975.



**John B. Moore** (F'79) was born in China in 1941. He received the bachelors and masters degrees in electrical engineering in 1963 and 1964, respectively, and the doctorate in electrical engineering from the University of Santa Clara, Santa Clara, CA in 1967.

He was appointed Senior Lecturer in the Electrical Engineering Department, University of Newcastle, Newcastle, U.K., in 1967, promoted to Associate Professor in 1968, and to full Professor (personal chair) in 1973. He was Department Head for the period 1975–1979. In 1982, he was appointed Professorial Fellow in the Department of Systems Engineering, Research School of Physical Sciences, Australian National University and promoted to Professor in 1990. He has been Head of the department since 1992. He has held visiting academic appointments at the University of Santa Clara (1968); the University of Maryland (1970); Colorado State University and Imperial College (1974); the University of California-Davis (1977); the University of Washington, Seattle (1981); Cambridge University and the National University of Singapore (1985); University of California, Berkeley (1987, 1989, and 1991); University of Alberta, Edmonton (1992, 1993); and the University of Regensburg (1993). He has spent periods in industry as a design engineer and as a consultant and currently has research grants from industry and government laboratories. He is a Team Leader in the Cooperative Research Centre for Robust and Adaptive Systems in the department. His current research is in control and communication systems and signal processing. He is coauthor with Brian Anderson of three books: *Linear Optimal Control* (Prentice-Hall, 1971), *Optimal Filtering* (Prentice-Hall, 1979), and *Optimal Control—Linear Quadratic Methods* (Prentice-Hall, 1989). He is coauthor of a book with Uwe Helmke: *Optimization and Dynamical Systems* (Springer-Verlag, 1993).

Dr. Moore is a Fellow of the Australian Academy of Technological Sciences.



**Vikram Krishnamurthy** (M'91) was born in India in 1966. He received the bachelors degree in electrical engineering with first class honours from the University of Auckland, Auckland, New Zealand, in 1987, and the doctoral degree from the Department of Systems Engineering, Australian National University, Canberra, in 1991.

During 1992–1993 he was a Senior Research Engineer at the Center for Robust and Adaptive Systems, Australian National University. In 1993 he was awarded the prestigious Queen Elizabeth II Fellowship, which he currently holds at the Research School of Information Sciences and Engineering, Australian National University. He is also a Senior Lecturer at the Department of Electrical Engineering, University of Melbourne, Melbourne, Australia. His research interests are in estimation theory, time-series analysis, and "blind-equalization" algorithms for communication channels.