Reduced-Complexity Decision-Directed Pilot-Aided Tracking of Doubly-Selective Channels

Arun P. Kannu and Philip Schniter
Dept. ECE, The Ohio State University
Columbus, OH 43210
pachaik@ee.eng.ohio-state.edu, schniter@ee.eng.ohio-state.edu

***Abstract*— We consider tracking a doubly-selective channel given a transmission stream which alternates between data and pilot blocks. Current and previous pilots, as well as previously-decoded data blocks, are used for estimation of channel parameters. The optimal Kalman and Wiener estimators for this scenario are computationally expensive. In response, we propose a novel two-stage estimation technique whose performance is close to that of the optimal estimators but whose complexity is significantly less. The first stage finds smoothed estimates of the channel during previous frames, while the second stage uses these smoothed estimates for channel prediction within the current frame.**[1]

I. Introduction

We consider identification of a doubly-selective channel when the transmitted signal is an infinite stream of frames, where each frame consists of a symbol block followed by a pilot block, as in Fig. 1. Identification is accomplished using the current pilot block and past frames, where the data in past frames is assumed to be perfectly decoded. The length-N_p pilot block is assumed to have a Kronecker delta structure with $N_p = 2N_h - 1$, where N_h is the channel delay spread.This pilot sequence has been claimed to satisfy several MSE- and capacity-based optimality criteria in the case of non-decision-directed identification of a doubly-selective channel from a single zero-padded frame [1]. The structure of our length-N_d data block is not important; it could be composed of frequency-domain symbols, as in OFDM [2], [3], or time domain symbols, as in single carrier cyclic prefix (SCCP) [4].

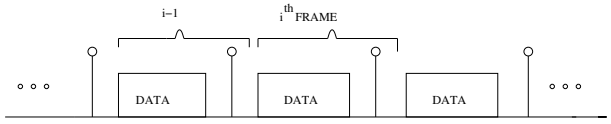

Fig. 1. Transmission pattern.

The channel estimation for OFDM systems has been considered in [5]–[10]. The works [5]–[8] considered fading scenarios in which the channel variation within a frame (or OFDM symbol) is negligible. Decision-directed adaptive algorithms for channel tracking were discussed in [5]. Pilot-based linear MMSE (LMMSE) techniques were discussed in [7] where the channel corresponding to the data portion is obtained by MMSE interpolation of the observations due to pilot portion. A decision-directed LMMSE prediction technique was proposed in [8], and modifications to make the predictor coefficients independent of the transmitted data were also given.

We consider the fast fading scenario, in which there is significant variation of the channel within each frame. For these fast fading scenarios, pilot-based least squares (LS) and LMMSE techniques were proposed in [10] and [9], respectively, and a recursive decision-directed LS technique was given in [11]. In [12], low complexity estimation techniques were presented based on a clever selection of pilot blocks. To improve on the MSE performance available from pilot-only schemes, we consider decision-directed pilot-aided MMSE estimation.

In Section III we review the "standard" methods of (linear time-varying) channel identification, which include pilot-based MMSE estimation, decision-directed pilot-aided Wiener prediction, and decision-directed pilot-aided Kalman prediction. The performance of the pilot-only technique is limited by the frame interval $N_f = N_d + N_p$; if $f_d > \frac{1}{2N_f}$, where f_d denotes the normalized Doppler frequency, then the channel is undersampled and identification breaks down. The decision-directed Kalman and Wiener predictors do not have this limitation and hence perform better for higher Doppler rates. However, once per frame they require the inversion of a matrix of size $N_f \times N_f$ or larger, so their complexity may be prohibitive for typical values of N_f. Hence in Sections IV and V we propose identification schemes with MSE performance close to that of the standard Kalman and Wiener predictors but without large matrix inverses. In Section VI we present numerical simulations that demonstrate the efficacy of our proposed techniques.

Notation: We use $(\cdot)^t$ to denote transpose, $(\cdot)^*$ conjugate, and $(\cdot)^H$ conjugate transpose. $E\{\cdot\}$ denotes expectation operator, $\delta(\cdot)$ the Kronecker delta, and $\langle\cdot\rangle_N$ the modulo-N operation.

II. System Model

We use $\mathcal{T}_{\mathsf{d}}^{(i)} := \{t_n^{(i)}\}_{n=0}^{N_d-1}$ to denote the data portion of the i^{th} transmission frame and $\mathcal{T}_{\mathsf{p}}^{(i)} := \{t_n^{(i)}\}_{n=N_d}^{N_d+N_p-1}$ to denote the corresponding pilot portion. The Kronecker pattern implies that $t_n^{(i)} = 0$ for $N_d \le n \le N_d + N_h - 2$ and $N_d + N_h \le n \le N_d + 2N_h - 2$, and that $t_{N_d+N_h-1}^{(i)} = \sqrt{2N_h - 1}$. The

[1]This work was supported by NSF CAREER CCR-0237037.

complete set of samples transmitted during the i^{th} frame is denoted by $\mathcal{T}^{(i)} := \mathcal{T}_{\mathsf{d}}^{(i)} \cup \mathcal{T}_{\mathsf{p}}^{(i)} = \{t_n^{(i)}\}_{n=0}^{N_f-1}$, and the multi-frame transmitted signal $\{t_n\}$ is defined by $t_n := t_{\langle n \rangle_{N_f}}^{(\lfloor n/N_f \rfloor)}$.

The transmitted signal passes through a noisy doubly-selective linear channel before observation at the receiver. The time-n observation can be written as

$$y_n \;=\; \sum_{d=0}^{N_h-1} h_{n,d} t_{n-d} + v_n \quad \text{for} \quad n \in \mathbb{Z}, \tag{1}$$

where $h_{n,d}$ denotes the response of the channel at time n to an impulse applied at time $n-d$, and where $\{v_n\}$ is proper complex zero-mean white Gaussian noise process with variance σ_v^2. If $y_n^{(i)} := y_{iN_f+n}$ and $h_{n,d}^{(i)} := h_{iN_f+n,d}$ and $v_n^{(i)} := v_{iN_f+n}$, then

$$y_n^{(i)} \;=\; \sum_{d=0}^{N_h-1} h_{n,d}^{(i)} t_{n-d}^{(i)} + v_n^{(i)} \quad \text{for} \quad 0 \le n \le N_f-1, \tag{2}$$

adopting the convention that $t_n^{(i)} = 0$ for $n < 0$, which is vindicated due to the sufficient number of zeros in the pilot block of the previous frame.

The following notation will be useful in the sequel. $\mathcal{Y}_{\mathsf{d}}^{(i)} := \{y_n^{(i)}\}_{n=0}^{N_d+N_h-2}$ and $\mathcal{Y}_{\mathsf{p}}^{(i)} := \{y_n^{(i)}\}_{n=N_d+N_h-1}^{N_f-1}$ will denote the data and pilot portions of the received samples in the i^{th} frame, respectively, and $\mathcal{Y}^{(i)} := \mathcal{Y}_{\mathsf{d}}^{(i)} \cup \mathcal{Y}_{\mathsf{p}}^{(i)}$. Arranging the elements of $\mathcal{Y}_{\mathsf{d}}^{(i)}, \mathcal{Y}_{\mathsf{p}}^{(i)}$, and $\mathcal{Y}^{(i)}$ in increasing order yields the vectors $\boldsymbol{y}_{\mathsf{d}}^{(i)}$, $\boldsymbol{y}_{\mathsf{p}}^{(i)}$, and $\boldsymbol{y}^{(i)}$, respectively. Similarly, $\mathcal{H}_{\mathsf{d}}^{(i)} := \{h_{n,d}^{(i)} \forall d\}_{n=0}^{N_d+N_h-2}$, $\mathcal{H}_{\mathsf{p}}^{(i)} := \{h_{n,d}^{(i)} \forall d\}_{n=N_d+N_h-1}^{N_f-1}$, and $\mathcal{H}^{(i)} := \mathcal{H}_{\mathsf{d}}^{(i)} \cup \mathcal{H}_{\mathsf{p}}^{(i)}$. The vectors $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ and $\boldsymbol{h}_{\mathsf{p}}^{(i)}$ are defined element-wise as $[\boldsymbol{h}_{\mathsf{d}}^{(i)}]_l = h_{\lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}^{(i)}$ and $[\boldsymbol{h}_{\mathsf{p}}^{(i)}]_l = h_{N_d+N_h-1+\lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}^{(i)}$, and $\boldsymbol{h}^{(i)} = [\boldsymbol{h}_{\mathsf{d}}^{(i)t}, \boldsymbol{h}_{\mathsf{p}}^{(i)t}]^t$.

Taking the point of view that $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ is useful for detection of the unknown data in $\mathcal{T}^{(i)}$, our goal is estimation of $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ using current and past observations $\{\mathcal{Y}^{(i-k)}\}_{k\ge 0}$, current pilots $\mathcal{T}_{\mathsf{p}}^{(i)}$, and past transmission frames $\{\mathcal{T}^{(i-k)}\}_{k\ge 1}$. The pilots are known *a priori* by the receiver, and, under the assumption of perfect decoding, past transmission frames are known as well. It is for reasons of simplicity that we assume perfect decoding; error propagation will be addressed in future work.

We assume the channel to be Rayleigh-fading wide-sense stationary uncorrelated scattering (WSSUS). Thus $h_{n,d}$ are zero-mean circular Gaussian with correlation $r_{hh}(l,m) = E\{h_{n,d} h_{n-l,d-m}^*\} = \sigma_d^2 J_0(2\pi f_d l)\delta(m)$, where $J_0(\cdot)$ is the 0^{th} order Bessel function of first kind, f_d is the normalized Doppler frequency, and σ_d^2 is the variance of d^{th} tap.

III. Review of Standard Estimators

Here we review the "standard" methods of channel identification and tracking. Though these techniques are well known, their presentation yields a notational framework that will be useful in the sequel.

A. Pilot-based Channel Estimator

The pilot-based channel estimator (PCE) uses only the received samples from the pilot block in the i^{th} frame and M previous frames, i.e., $\{\mathcal{Y}_{\mathsf{p}}^{(i-k)}\}_{k=0}^{k=M}$. Hence we form the observation vector $\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} = [\boldsymbol{y}_{\mathsf{p}}^{(i)t}, ..., \boldsymbol{y}_{\mathsf{p}}^{(i-M)t}]^t$. Since $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ and $\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)}$ are jointly Gaussian, the MMSE estimator is linear and given by [13]

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{pilot}} \;=\; \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{p}},\boldsymbol{h}_{\mathsf{d}}}^H (\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{p}},\underline{\boldsymbol{y}}_{\mathsf{p}}})^{-1} \underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)}, \tag{3}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{p}},\boldsymbol{h}_{\mathsf{d}}} = E\{\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$ and $\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{p}},\underline{\boldsymbol{y}}_{\mathsf{p}}} = E\{\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} \underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)H}\}$.

B. Wiener Predictor

The decision-directed pilot-aided Wiener predictor (WP) uses the pilot block in the i^{th} frame $\{\mathcal{Y}_{\mathsf{p}}^{(i)}\}$ as well as all the received samples in M previous frames $\{\mathcal{Y}^{(i-k)}\}_{k=1}^{M}$. Here we form the observation vector $\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} = [\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \boldsymbol{y}^{(i-1)t}, ..., \boldsymbol{y}^{(i-M)t}]^t$ and the channel vector $\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} = [\boldsymbol{h}_{\mathsf{p}}^{(i)t}, \boldsymbol{h}^{(i-1)t}, ..., \boldsymbol{h}^{(i-M)t}]^t$. Since the channel is linear, we can relate the observation to the channel coefficients via

$$\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} \;=\; \boldsymbol{T}_{\mathsf{w}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{w}}^{(i)} \tag{4}$$

where $\boldsymbol{T}_{\mathsf{w}}^{(i)}$ is a matrix whose elements are constructed from the set $\{\mathcal{T}^{(i-k)}\}_{k=1}^{M}$, and where $\underline{\boldsymbol{v}}_{\mathsf{w}}^{(i)}$ contains white Gaussian noise samples. Assuming that all the previous data blocks have been correctly decoded, the matrix $\boldsymbol{T}_{\mathsf{w}}^{(i)}$ is known and the quantities $\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)}$ and $\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)}$ are jointly Gaussian. Hence the MMSE predictor is given by

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{wiener}} \;=\; \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{w}},\boldsymbol{h}_{\mathsf{d}}}^{(i)H} (\boldsymbol{T}_{\mathsf{w}}^{(i)} \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\underline{\boldsymbol{h}}_{\mathsf{w}}} \boldsymbol{T}_{\mathsf{w}}^{(i)H} + \sigma_v^2 \boldsymbol{I})^{-1} \underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} \tag{5}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{w}},\boldsymbol{h}_{\mathsf{d}}}^{(i)} = E\{\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$ and $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\underline{\boldsymbol{h}}_{\mathsf{w}}} = E\{\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)H}\}$.

In (5), the matrix to be inverted is of size $(MN_f + N_h) \times (MN_f + N_h)$. The inverse must be computed for each frame index i since it depends on transmitted data. Thus increasing M leads to higher performance but increased complexity.

C. Kalman Predictor

In this section, we formulate channel estimation as a Kalman prediction (KP) problem [14].

1) Problem Setup: We assign $\boldsymbol{h}^{(i-1)}$ as the current *state* of the channel, $\boldsymbol{h}^{(i)}$ as the next state, and $\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} = [\boldsymbol{y}_{\mathsf{d}}^{(i-1)t}, \boldsymbol{y}_{\mathsf{p}}^{(i)t}]^t$ as the current observation.

2) Dynamical Equation: The state dynamics can be written as

$$\boldsymbol{h}^{(i)} \;=\; \boldsymbol{A}_{\mathsf{k}} \boldsymbol{h}^{(i-1)} + \boldsymbol{D}_{\mathsf{k}} \boldsymbol{w}_{\mathsf{k}}^{(i-1)} \tag{6}$$

where $\boldsymbol{w}_{\mathsf{k}}^{(i-1)}$ is a white Gaussian vector, i.e., $E\{\boldsymbol{w}_{\mathsf{k}}^{(i-1)} \boldsymbol{w}_{\mathsf{k}}^{(i-1-p)H}\} = \sigma_{w_{\mathsf{k}}}^2 \boldsymbol{I}\delta(p)$. The matrices $\boldsymbol{A}_{\mathsf{k}}$ and $\boldsymbol{D}_{\mathsf{k}}$, and the state noise variance $\sigma_{w_{\mathsf{k}}}^2$, are obtained by auto regressive (AR) modeling of the Doppler channel. The WSSUS assumption implies that $\boldsymbol{A}_{\mathsf{k}}$, $\boldsymbol{D}_{\mathsf{k}}$ and $\sigma_{w_{\mathsf{k}}}^2$ are constant from frame to frame.

3) Observation Equation: The first part of the observation vector, $\boldsymbol{y}_{\mathsf{d}}^{(i-1)}$, can be written as

$$\boldsymbol{y}_{\mathsf{d}}^{(i-1)} \;=\; \boldsymbol{T}_{\mathsf{k}}^{(i-1)}\boldsymbol{h}^{(i-1)} + \boldsymbol{v}_{\mathsf{d}}^{(i-1)}, \tag{7}$$

where $\boldsymbol{T}_{\mathsf{k}}^{(i-1)}$ is a matrix whose elements are constructed from $\mathcal{T}^{(i-1)}$ and where $\boldsymbol{v}_{\mathsf{d}}^{(i-1)}$ is a white Gaussian vector. The second part of the observation vector, $\boldsymbol{y}_{\mathsf{p}}^{(i)}$, can be written as

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} \;=\; \boldsymbol{G}\boldsymbol{h}^{(i)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}, \tag{8}$$

for known constant matrix $\boldsymbol{G}$ (composed of samples from pilot blocks), and where $\boldsymbol{v}_{\mathsf{p}}^{(i)}$ is a white Gaussian vector. Using (6), we rewrite (8) as

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} \;=\; \boldsymbol{G}\boldsymbol{A}_{\mathsf{k}}\boldsymbol{h}^{(i-1)} + \boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)} + \boldsymbol{v}_{\mathsf{p}}^{(i)} \tag{9}$$

Using (8) and (9), we have

$$\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} \;=\; \underbrace{\begin{bmatrix}\boldsymbol{T}_{\mathsf{k}}^{(i-1)}\\ \boldsymbol{G}\boldsymbol{A}_{\mathsf{k}}\end{bmatrix}}_{\boldsymbol{C}_{\mathsf{k}}^{(i-1)}}\boldsymbol{h}^{(i-1)} + \underbrace{\begin{bmatrix}\boldsymbol{v}_{\mathsf{d}}^{(i-1)}\\ \boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}\end{bmatrix}}_{\underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)}} \tag{10}$$

$$\;=\; \boldsymbol{C}_{\mathsf{k}}^{(i-1)}\boldsymbol{h}^{(i-1)} + \underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)} \tag{11}$$

We define $\boldsymbol{S}_{\mathsf{k}} = E\{\boldsymbol{w}_{\mathsf{k}}^{(i-1)}\underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)H}\}$ and $\boldsymbol{R}_{\mathsf{k}} = E\{\underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)}\underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)H}\}$ for use in the sequel.

4) Predictor Equation: From (6) and (11), the MMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)}, ..., \underline{\boldsymbol{y}}_{\mathsf{k}}^{(0)}\}$, denoted by $\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{kalman}}$, is given recursively as

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{kalman}} \;=\; \boldsymbol{A}_{\mathsf{k}}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{kalman}} + \boldsymbol{L}_{\mathsf{k}}^{(i-1)}[\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} - \boldsymbol{C}_{\mathsf{k}}^{(i-1)}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{kalman}}] \tag{12}$$

where the predictor gain $\boldsymbol{L}_{\mathsf{k}}^{(i-1)}$ is given by

$$\boldsymbol{L}_{\mathsf{k}}^{(i-1)} \;=\; (\boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{C}_{\mathsf{k}}^{(i-1)H} + \boldsymbol{D}_{\mathsf{k}}\boldsymbol{S}_{\mathsf{k}}) \times (\boldsymbol{C}_{\mathsf{k}}^{(i-1)}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{C}_{\mathsf{k}}^{(i-1)H} + \boldsymbol{R}_{\mathsf{k}})^{-1} \tag{13}$$

and $\boldsymbol{P}_{\mathsf{k}}^{(i-1)}$ is given recursively as

$$\boldsymbol{P}_{\mathsf{k}}^{(i-1)} \;=\; \sigma_{w_{\mathsf{k}}}^2\boldsymbol{D}_{\mathsf{k}}\boldsymbol{D}_{\mathsf{k}}^H + \boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\mathsf{k}}^{(i-2)}\boldsymbol{A}_{\mathsf{k}}^H - \boldsymbol{L}_{\mathsf{k}}^{(i-2)}(\boldsymbol{C}_{\mathsf{k}}^{(i-2)}\boldsymbol{P}_{\mathsf{k}}^{(i-2)}\boldsymbol{A}_{\mathsf{k}}^H + \boldsymbol{S}_{\mathsf{k}}^H\boldsymbol{D}_{\mathsf{k}}^H) \tag{14}$$

with initializations $\boldsymbol{P}_{\mathsf{k}}^{(0)} = E\{\boldsymbol{h}^{(0)}\boldsymbol{h}^{(0)H}\}$ and $\hat{\boldsymbol{h}}^{(0)}\big|_{\mathsf{kalman}} = \boldsymbol{0}$.

Note that the Kalman predictor uses *all* previous observations in its prediction of $\boldsymbol{h}^{(i)}$; this is the advantage of the KP over the WP. However, the performance of the KP depends on how well the model (6) describes the true evolution of the state process. Note also, from (13), that the KP requires a matrix inversion of size $N_f \times N_f$ once per frame.

IV. Low Complexity Predictor

In this section, we propose a novel computationally-efficient decision-directed channel tracker that does not require large matrix inversions.

We break the prediction of $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ into two stages. In the first stage, we find "smoothed" channel estimates during the $(i-1)^{th}$ frame. In the second stage, we use the received pilot block in the i^{th} frame as well as smoothed channel estimates from M previous frames to predict $\boldsymbol{h}_{\mathsf{d}}^{(i)}$. With some approximations, the predictor can be made time-invariant.

A. Kalman Smoothing Stage

The smoothed channel estimates of the $(i-1)^{th}$ frame are obtained using Kalman filtering. We design the Kalman smoother so that it takes L adjacent received samples and finds the smoothed estimates of the channel during that time interval. To find the smoothed estimates of the channel during the entire frame, this procedure must be repeated $K = \frac{N_f}{L}$ times. Since we use only L measurements at a time, inversion will be performed on at most an $L \times L$ matrix. Details are provided below.

Let $\mathcal{H}^{(i-1,k)}$ and $\mathcal{Y}^{(i-1,k)}$ denote the k^{th} subset of $\mathcal{H}^{(i-1)}$ and $\mathcal{Y}^{(i-1)}$, respectively, with $\mathcal{H}^{(i-1,k)} = \{h_{n,d}^{(i-1)}\ \forall d\}_{n=kL}^{(k+1)L-1}$ and $\mathcal{Y}^{(i-1,k)} = \{y_n^{(i-1)}\}_{n=kL}^{(k+1)L-1}$ for $k \in \{0, ..., K-1\}$. The vectors $\boldsymbol{h}^{(i-1,k)}$ and $\boldsymbol{y}^{(i-1,k)}$ are defined element-wise as $[\boldsymbol{h}^{(i-1,k)}]_l = h_{kL+\lfloor l/N_h\rfloor,\langle l\rangle_{N_h}}^{(i-1)}$ and $[\boldsymbol{y}^{(i-1,k)}]_l = y_{kL+l}^{(i-1)}$. Let $\boldsymbol{h}^{(i-1,k)}$ be the current channel state of the smoother, $\boldsymbol{h}^{(i-1,k+1)}$ be the next state, and $\boldsymbol{y}^{(i-1,k)}$ be the current observation vector. We write the dynamical equation as

$$\boldsymbol{h}^{(i-1,k+1)} \;=\; \boldsymbol{A}_{\mathsf{l}}\boldsymbol{h}^{(i-1,k)} + \boldsymbol{D}_{\mathsf{l}}\boldsymbol{w}_{\mathsf{l}}^{(i-1,k)} \tag{15}$$

where $\boldsymbol{w}_{\mathsf{l}}^{(i-1,k)}$ is a white Gaussian process with $E\{\boldsymbol{w}_{\mathsf{l}}^{(i-1,k)}\boldsymbol{w}_{\mathsf{l}}^{(i-1-p,k-q)}\} = \sigma_{w_{\mathsf{l}}}^2\boldsymbol{I}\delta(p)\delta(q)$. The matrices $\boldsymbol{A}_{\mathsf{l}}$ and $\boldsymbol{D}_{\mathsf{l}}$ and $\sigma_{w_{\mathsf{l}}}^2$ are obtained by AR modeling of the channel. The current observation $\boldsymbol{y}^{(i-1,k)}$can be written as

$$\boldsymbol{y}^{(i-1,k)} \;=\; \boldsymbol{T}_{\mathsf{l}}^{(i-1,k)}\boldsymbol{h}^{(i-1,k)} + \boldsymbol{v}^{(i-1,k)} \tag{16}$$

where $\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)}$ is a matrix whose elements are constructed from a subset of $\mathcal{T}^{(i-1)}$. For the system described by (15) and (16), $\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}$ denote the *smoothed* MMSE estimate of $\boldsymbol{h}^{(i-1,k)}$ using observations $\{\boldsymbol{y}^{(i-1,k)}, ..., \boldsymbol{y}^{(i-1,0)}, \boldsymbol{y}^{(i-2)}, ..., \boldsymbol{y}^{(0)}\}$ and let $\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}$ denote the *predicted* MMSE estimate of $\boldsymbol{h}^{(i-1,k)}$ using the observations $\{\boldsymbol{y}^{(i-1,k-1)}, ..., \boldsymbol{y}^{(i-1,0)}, \boldsymbol{y}^{(i-2)}, ..., \boldsymbol{y}^{(0)}\}$. The smoothed estimate is given by [14]

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} = \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} + \boldsymbol{M}_{\mathsf{l}}^{(i-1,k)}[\boldsymbol{y}^{(i-1,k)} - \boldsymbol{T}_{\mathsf{l}}^{(i-1,k)}\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}] \tag{17}$$

with smoother gain

$$\boldsymbol{M}_{\mathsf{l}}^{(i-1,k)} \;=\; \boldsymbol{P}_{\mathsf{l}}^{(i-1,k)}\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)H} \times (\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)}\boldsymbol{P}_{\mathsf{l}}^{(i-1,k)}\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)H} + \sigma_v^2\boldsymbol{I})^{-1}. \tag{18}$$

The predicted estimate is given recursively as

$$\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} \;=\; \boldsymbol{A}_{\mathsf{l}}\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k-1)} + \boldsymbol{A}_{\mathsf{l}}\boldsymbol{M}_{\mathsf{l}}^{(i-1,k-1)} \times [\boldsymbol{y}^{(i-1,k-1)} - \boldsymbol{T}_{\mathsf{l}}^{(i-1,k-1)}\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k-1)}] \tag{19}$$

where $\boldsymbol{P}_{\mathsf{l}}^{(i-1,k)}$ is given recursively as

$$\boldsymbol{P}_{\mathsf{l}}^{(i-1,k)} \;=\; \sigma_{w_{\mathsf{l}}}^2\boldsymbol{D}_{\mathsf{l}}\boldsymbol{D}_{\mathsf{l}}^H + \boldsymbol{A}_{\mathsf{l}}\boldsymbol{P}_{\mathsf{l}}^{(i-1,k-1)}\boldsymbol{A}_{\mathsf{l}}^H - \boldsymbol{A}_{\mathsf{l}}\boldsymbol{M}_{\mathsf{l}}^{(i-1,k-1)}\boldsymbol{T}_{\mathsf{l}}^{(i-1,k-1)}\boldsymbol{P}_{\mathsf{l}}^{(i-1,k-1)}\boldsymbol{A}_{\mathsf{l}}^H \tag{20}$$

with initializations $\boldsymbol{P}_{\mathsf{l}}^{(i-1,0)} = \boldsymbol{P}_{\mathsf{l}}^{(i-2,K)}$, $\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,0)} = \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-2,K)}$, $\boldsymbol{P}_{\mathsf{l}}^{(0,0)} = E\{\boldsymbol{h}^{(0,0)}\boldsymbol{h}^{(0,0)H}\}$, and $\hat{\boldsymbol{h}}_{\mathsf{l}}^{(0,0)} = \boldsymbol{0}$. We form the smoothed estimate of the channel tap vector for the $(i-1)^{th}$ frame as

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} \;=\; [\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,0)t}, ..., \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,K-1)t}]^t. \tag{21}$$

This smoothed channel estimate vector is related to the true channel vector as

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} \;=\; \boldsymbol{h}^{(i-1)} + \boldsymbol{e}_{\mathsf{l}}^{(i-1)} \tag{22}$$

where $\boldsymbol{e}_{\mathsf{l}}^{(i-1)}$ denotes the smoothing error.

B. Wiener Prediction Stage

From (8) and (22), the prediction stage observation vector $\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}$ is formed as

$$\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} \;=\; [\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)t}, ..., \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-M)t}]^t \tag{23}$$

$$= \underbrace{\begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{B}} \underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)}, \tag{24}$$

where $\underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)} = [\boldsymbol{h}^{(i)t}, ..., \boldsymbol{h}^{(i-M)t}]^t$ and $\underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)} = [\boldsymbol{v}_{\mathsf{p}}^{(i)t}, \boldsymbol{e}_{\mathsf{l}}^{(i-1)t}, ..., \boldsymbol{e}_{\mathsf{l}}^{(i-M)t}]^t$.

Experimentally, we find that the smoothing error $\boldsymbol{e}_{\mathsf{l}}^{(i)}$ is dominated by the measurement noise. So, to reduce predictor complexity, we make the following approximations.

$$E\{\underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)} \underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)H}\} \;\approx\; \sigma_v^2 \boldsymbol{I}, \tag{25}$$

$$E\{\underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\} \;\approx\; \boldsymbol{0}, \tag{26}$$

$$E\{\underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)} \underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)H}\} \;\approx\; \boldsymbol{0}. \tag{27}$$

With (25)-(27), the predictor equation is

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{lcp}} \;=\; \boldsymbol{R}^H_{\underline{\boldsymbol{y}}_{\mathsf{l}}, \boldsymbol{h}_{\mathsf{d}}} (\boldsymbol{B} \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{l}}, \underline{\boldsymbol{h}}_{\mathsf{l}}} \boldsymbol{B}^H + \sigma_v^2 \boldsymbol{I})^{-1} \underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}, \tag{28}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{l}}, \boldsymbol{h}_{\mathsf{d}}} = E\{\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$ and $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{l}}, \underline{\boldsymbol{h}}_{\mathsf{l}}} = E\{\underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)H}\}$.

The predictor coefficients in (28) are time invariant; matrix inversion is not required at the frame rate. Thus LCP requires at most an $L \times L$ matrix inversion in the smoothing stage (18). The choice of L is a tradeoff between performance and complexity.

V. LCP Modifications

Here we discuss modifications to the LCP prediction stage motivated by further reductions in complexity and memory requirements.

A. LCP with Downsampling

In "LCP with downsampling" (LCPD), we use a downsampled set of smoothed estimates for prediction. For integer downsampling ratio r, the downsampled smoothed estimates of the $(i-1)^{th}$ frame are defined element-wise as

$$[\tilde{\boldsymbol{h}}_{\mathsf{lr}}^{(i-1)}]_l = [\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)}]_{\lfloor \frac{l}{N_h} \rfloor r N_h + \langle l \rangle_{N_h}} \tag{29}$$

The prediction is identical to LCP but with $\tilde{\boldsymbol{h}}_{\mathsf{lr}}^{(i-1)}$ replacing $\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)}$. The downsampling helps in two ways.

1) The observation vector for LCPD prediction is approximately r times smaller than that of LCP, reducing complexity and memory requirements.
2) With downsampling, the time difference between smoothed estimates increases and hence approximation (25) becomes more accurate.

B. LCP with Kalman Prediction

The prediction stage of LCP was based on Wiener prediction using smoothed estimates of M previous frames. Instead, LCP could be modified to use Kalman prediction (LCKP), thus incorporating *all* previously smoothed estimates. Using $\boldsymbol{h}^{(i-1)}$ and $\boldsymbol{h}^{(i)}$ to denote the current and next channel states, respectively, the dynamical equation is described by (6). The current observation vector can be written as

$$\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} \;=\; [\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)t} \boldsymbol{y}_{\mathsf{p}}^{(i)t}]^t \tag{30}$$

$$\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)} \;=\; \underbrace{\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{B}_{\mathsf{lk}}} \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} \;=\; \boldsymbol{B}_{\mathsf{lk}} \boldsymbol{h}^{(i-1)} + \boldsymbol{B}_{\mathsf{lk}} \boldsymbol{e}_{\mathsf{l}}^{(i-1)}, \tag{31}$$

where $\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)}$ is the smoothed channel estimate during the data portion of the $(i-1)^{th}$ frame. Using (6), (8), and (31), $\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)}$ can be rewritten as

$$\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} \;=\; \underbrace{\begin{bmatrix} \boldsymbol{B}_{\mathsf{lk}} \\ \boldsymbol{G} \boldsymbol{A}_{\mathsf{k}} \end{bmatrix}}_{\boldsymbol{C}_{\mathsf{lk}}} \boldsymbol{h}^{(i-1)} + \underbrace{\begin{bmatrix} \boldsymbol{B}_{\mathsf{lk}} \boldsymbol{e}_{\mathsf{l}}^{(i-1)} \\ \boldsymbol{v}_{\mathsf{p}}^{(i)} + \boldsymbol{G} \boldsymbol{D}_{\mathsf{k}} \boldsymbol{w}_{\mathsf{k}}^{(i-1)} \end{bmatrix}}_{\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)}} \tag{32}$$

$$= \boldsymbol{C}_{\mathsf{lk}} \boldsymbol{h}^{(i-1)} + \underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)}. \tag{33}$$

Incorporating approximation (25), we have

$$\boldsymbol{R}_{\mathsf{lk}} \;=\; E\{\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)} \underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)H}\} \tag{34}$$

$$= \begin{bmatrix} \sigma_v^2 \boldsymbol{B}_{\mathsf{lk}} \boldsymbol{B}_{\mathsf{lk}}^H & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_v^2 \boldsymbol{I} + \sigma_{w_{\mathsf{k}}}^2 \boldsymbol{G} \boldsymbol{D}_{\mathsf{k}} \boldsymbol{D}_{\mathsf{k}}^H \boldsymbol{G}^H \end{bmatrix} \tag{35}$$

$$\boldsymbol{S}_{\mathsf{lk}} \;=\; E\{\boldsymbol{w}_{\mathsf{k}}^{(i-1)} \underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)H}\} \tag{36}$$

$$= \begin{bmatrix} \boldsymbol{0} & \sigma_{w_{\mathsf{k}}}^2 \boldsymbol{D}_{\mathsf{k}}^H \boldsymbol{G}^H \end{bmatrix}. \tag{37}$$

Since constant matrices are used in the dynamical equation (6), the observation equation (33), and the correlation matrix definitions (35) and (37), the *steady state* Kalman predictor suffices. Denoting $\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{lckp}}$ as the MMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\boldsymbol{y}_{\mathsf{lk}}^{(i-1)}, ..., \boldsymbol{y}_{\mathsf{lk}}^{(0)}\}$, the recursive steady state predictor equation is given by [14]

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{lckp}} = \boldsymbol{A}_{\mathsf{k}} \hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{lckp}} + \boldsymbol{L}_\infty [\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} - \boldsymbol{C}_{\mathsf{lk}} \hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{lckp}}] \tag{38}$$

with steady state predictor gain

$$\boldsymbol{L}_\infty \;=\; (\boldsymbol{A}_{\mathsf{k}} \boldsymbol{P}_\infty \boldsymbol{C}_{\mathsf{lk}} + \boldsymbol{D}_{\mathsf{k}} \boldsymbol{S}_{\mathsf{lk}})(\boldsymbol{C}_{\mathsf{lk}} \boldsymbol{P}_\infty \boldsymbol{C}_{\mathsf{lk}}^H + \boldsymbol{R}_{\mathsf{lk}})^{-1}$$

and symmetric, positive semi-definite Riccati solution

$$\boldsymbol{P}_\infty \;=\; \sigma_{w_{\mathsf{k}}}^2 \boldsymbol{D}_{\mathsf{k}} \boldsymbol{D}_{\mathsf{k}}^H + \boldsymbol{A}_{\mathsf{k}} \boldsymbol{P}_\infty \boldsymbol{A}_{\mathsf{k}}^H - (\boldsymbol{A}_{\mathsf{k}} \boldsymbol{P}_\infty \boldsymbol{C}_{\mathsf{lk}} + \boldsymbol{D}_{\mathsf{k}} \boldsymbol{S}_{\mathsf{lk}}) \times (\boldsymbol{C}_{\mathsf{lk}} \boldsymbol{P}_\infty \boldsymbol{C}_{\mathsf{lk}}^H + \boldsymbol{R}_{\mathsf{lk}})^{-1} (\boldsymbol{C}_{\mathsf{lk}} \boldsymbol{P}_\infty \boldsymbol{A}_{\mathsf{k}}^H + \boldsymbol{S}_{\mathsf{lk}}^H \boldsymbol{D}_{\mathsf{lk}}^H).$$

Note that $\boldsymbol{L}_\infty$ and $\boldsymbol{P}_\infty$ are frame invariant. Whereas the LCP observation includes smoothed estimates from M previous frames, the LCKP observation includes only smoothed estimates from the most recently decoded frame, thus reducing memory requirements.

Technique	*Size of matrix inversion per frame*
PCE	n.a.
WP	$(MN_f + N_h) \times (MN_f + N_h)$
KP	$N_f \times N_f$
LCP	$L \times L$
LCPD	$L \times L$
LCKP	$L \times L$

TABLE I
RELATIVE ALGORITHM COMPLEXITY.

VI. SIMULATION RESULTS

We consider frame size $N_f = 80$, channel delay spread $N_h = 8$, and (uniform) channel tap variance $\sigma_d^2 = N_h^{-1}$. For each estimation technique, the required matrix inversion size is given in Table I. The typical values of M and L in our simulations are 2 and 10 respectively. Though LCP, LCPD and LCKP require $K = \frac{N_f}{L}$ matrix inversions per frame, the $\mathcal{O}(N^3)$ matrix inversion complexity rule implies that LCP, LCPD and LCKP are much more computationally efficient than WP and KP. For all simulations, MSE performance is averaged over at least 1000 channel realizations, where each realization spans at least 10 frames.

As a reference, we consider the so-called "persistent training and prediction" (PTP) scheme. For PTP, we transmit persistent pilots instead of data:

$$t_n = \begin{cases} \sqrt{N_h} & \text{if } \frac{n}{N_h} \in \mathbf{Z} \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

We predict $\boldsymbol{h}^{(i)}$ using all past observations and the pilot observation of the i^{th} frame with an IIR Wiener filter.

In Fig. 2, we plot MSE versus SNR at $f_d = 0.01$. In this case, $f_d > \frac{1}{2N_f}$ and hence the performance of PCE is poor. In Fig. 3, we plot MSE versus normalized Doppler f_d at SNR = 15dB. As expected, MSE performance decreases as f_d increases. Note that LCP performance is not far from that of WP and KP.

In Fig. 4, we show the effect of M on LCP and WP. As M increases, both WP and LCP incorporate more observations, hence MSE performance improves. In Fig. 5, we show the effect of L on LCP performance. Increasing L improves performance of the Kalman smoothing stage, at the cost of increased complexity.

In Fig. 6, we plot the MSE performance of LCPD for different downsampling ratios r. Based on the discussion in Section V-A, it is not surprising that the $r = 2$ performance is slightly better than the $r = 1$ performance at high SNR. Further increase in r, however, degrades MSE performance because too much information is lost through downsampling.

In Fig. 7, we compare LCKP to LCP and KP. As expected, LCKP performs better than LCP, though the difference is significant only at high SNR. Also as expected, LCKP performs worse than KP, though the performance gap would decrease with an increase in L.

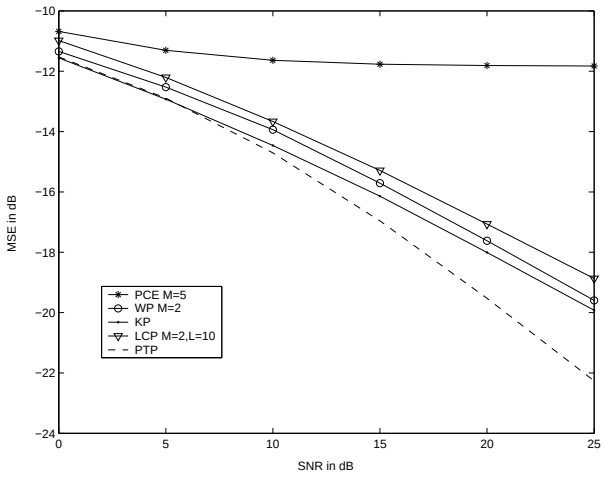

Fig. 2. Comparison of estimation techniques for $f_d = 0.01$.

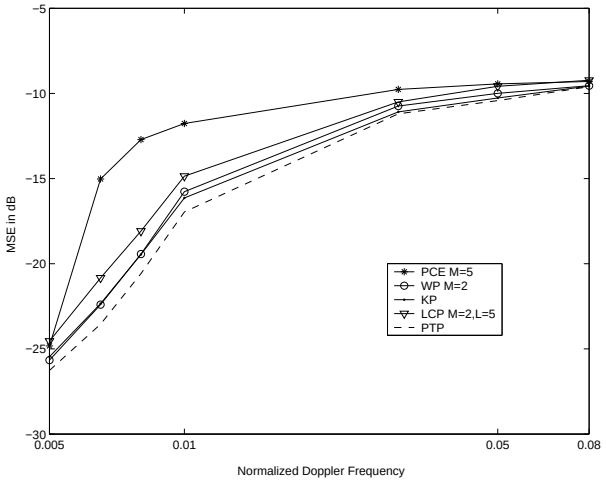

Fig. 3. Effect of f_d on MSE performance at $SNR = 15dB$.

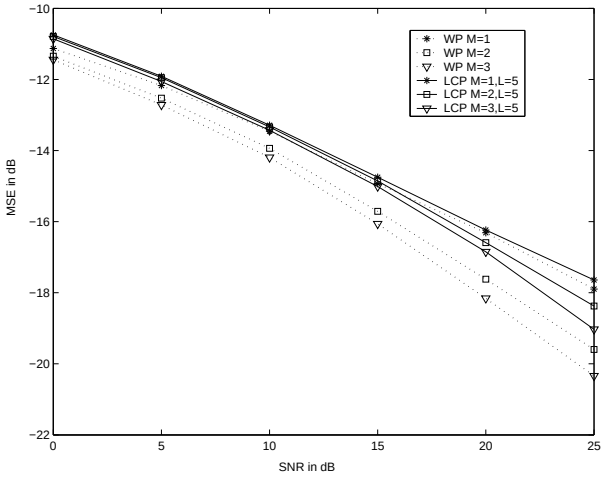

Fig. 4. Effect of M on WP and LCP for $f_d = 0.01$.

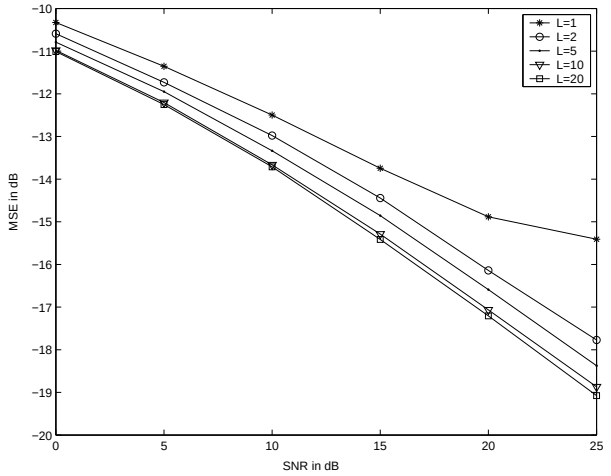

Fig. 5. Effect of L on LCP with $M = 2$ for $f_d = 0.01$.

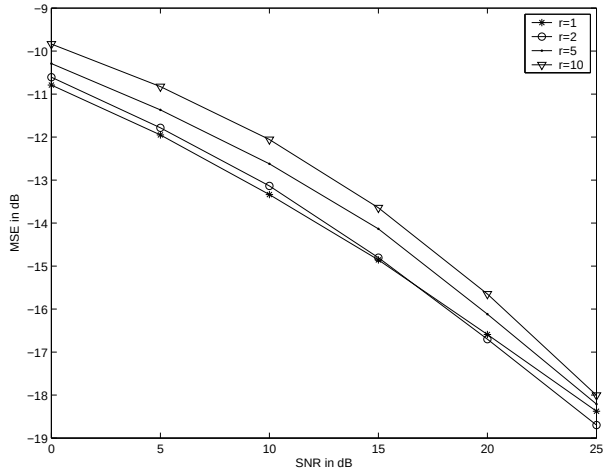

Fig. 6. Performance of LCPD with $M = 2$ and $L = 5$ for $f_d = 0.01$.

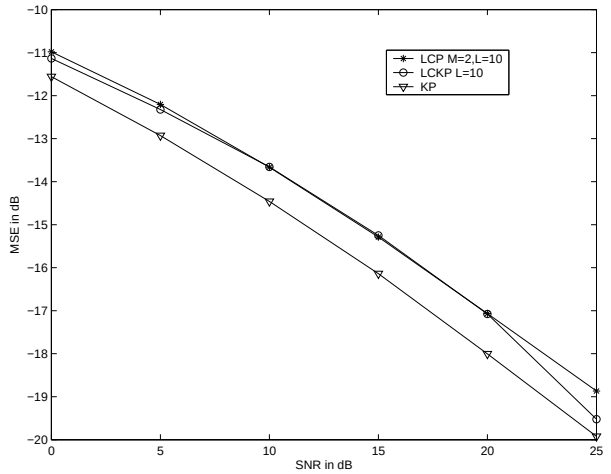

Fig. 7. Comparison of LCP, LCKP and KP for $f_d = 0.01$.

VII. Conclusion

We proposed a low-complexity two-stage doubly-selective-channel tracking scheme for transmissions composed of an infinite sequence of frames, where each frame has a data block followed by a Kronecker-delta pilot block. The first-stage computes smoothed estimates of previous channel coefficients, and the second stage uses these smoothed estimates, along with the current pilot observation, to predict the channel in the current frame. Approximations are made to reduce the complexity of the second stage predictor. Simulation results show that the proposed low complexity scheme has MSE performance comparable to the Wiener and Kalman predictors but without requiring large matrix inverses. Modifications of the low complexity scheme that reduce memory requirements were also introduced.

References

[1] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1351–1366, May 2003.

[2] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. 19, pp. 628–634, Oct. 1971.

[3] L. J. Cimini, Jr., "Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, pp. 665–765, July 1985.

[4] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, pp. 58–66, Apr. 2002.

[5] D. Schafhuber, G. Matz, and F. Hlawatsch, "Adaptive prediction of time-varying channels for coded OFDM systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, pp. 2459–2552, May 2002.

[6] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Baörjesson, "On channel estimation in OFDM systems," in *Proc. IEEE Veh. Tech. Conf.*, vol. 2, pp. 815–819, July 1995.

[7] Y. Li, L. J. Cimini, Jr., and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902–915, July 1998.

[8] D. Schafhuber, G. Matz, and F. Hlawatsch, "Predictive equalization of time-varying channels for coded OFDM/BFDM systems," in *Proc. IEEE Global Telecommunications Conf.*, vol. 2, pp. 721–725, Nov. 2000.

[9] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1375–1387, Aug. 2001.

[10] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Estimation of fast fading channels in OFDM," in *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 1, pp. 465–470, Mar. 2002.

[11] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, pp. 862–873, June 1999.

[12] P. Schniter, "Low-complexity estimation of doubly-selective channels," in *Proc. IEEE Workshop Signal Processing Advances in Wireless Commun.*, 2003. (to appear).

[13] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 2nd ed., 1994.

[14] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.