Low Complexity Techniques for Tracking Doubly-Selective Channels

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Science in the

Graduate School of The Ohio State University

By

Arun Pachai Kannu, B.E.

* * * * *

The Ohio State University

2004

Master's Examination Committee:

Prof. Philip Schniter, Adviser

Prof. Hesham El-Gamal

Approved by

Adviser

Department of Electrical
and Computer Engineering

ABSTRACT

We consider channel identification techniques for a transmission scheme which consists of a sequence of frames with each frame consisting of pilots and data. We consider two systems, the first one having pilots in the time multiplexed fashion, and the second one being a multi-carrier system with frequency multiplexed pilots. We consider scenarios in which the channel is both frequency and time selective.

For the time multiplexed pilots system, channel estimation performance of the pilot-only technique is limited by pilot spacing. The standard optimal decision-directed techniques 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.

The frequency multiplexed pilots system suffers from the inter-carrier interference (ICI) in doubly-selective channels. We present channel estimation techniques which take the ICI effects into account. We also present a novel pilot tone selection criterion and the theoretical MSE calculations show its efficacy for a wide range of Doppler frequencies. The performance limitation due to ICI can be reduced by employing decision-directed estimators. We develop a low-complexity estimation technique and simulation results show good performance of the proposed techniques.

Low Complexity Techniques for Tracking Doubly-Selective Channels

By

Arun Pachai Kannu, M.S.

The Ohio State University, 2004

Prof. Philip Schniter, Adviser

We consider channel identification techniques for a transmission scheme which consists of a sequence of frames with each frame consisting of pilots and data. We consider two systems, the first one having pilots in the time multiplexed fashion, and the second one being a multi-carrier system with frequency multiplexed pilots. We consider scenarios in which the channel is both frequency and time selective.

For the time multiplexed pilots system, channel estimation performance of the pilot-only technique is limited by pilot spacing. The standard optimal decision-directed techniques 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.

The frequency multiplexed pilots system suffers from the inter-carrier interference (ICI) in doubly-selective channels. We present channel estimation techniques which take the ICI effects into account. We also present a novel pilot tone selection criterion and the theoretical MSE calculations show its efficacy for a wide range of Doppler frequencies. The performance limitation due to ICI can be reduced by employing decision-directed estimators. We develop a low-complexity estimation technique and simulation results show good performance of the proposed techniques.

dedicated to my family ...

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Prof. Phil Schniter, for his valuable advice, guidance and help throughout this work. I thank Prof. Hesham El-Gamal for agreeing to be in the thesis committee. Learning Comm theory was a great experience with his insightful comments. I thank Prof. Serrani for his excellent teaching of Estimation theory, which helped me a lot with my research.

I thank my parents and sister, for their love and support. I also thank my uncles, aunts and cousins; they gave me the motivation and inspiration to do well in studies. Time spent together with them has always been so much fun.

I thank Praveen, Sib, Aditi, Arul for the fun times during lunch breaks and the tough times while solving homeworks! I thank Ravi for his help with linux and latex. I thank members of the IPS lab for their help and friendly conversations.

This material is based upon work supported by the National Science Foundation under Grant No. 0237037.

VITA

September 10, 1979	Born - Virudhunagar, India
2001	B.E. Electronics and Communications Anna University, India
2001 - 2002	Research Associate AU-KBC Research Center Anna University, India
2002 - 2003	University Fellow The Ohio State University
Jan - Mar 2004	Graduate Teaching Associate The Ohio State University
Sep - Dec 2003, Apr 2004 - present	Graduate Research Associate The Ohio State University

PUBLICATIONS

Research Publications

A. P. Kannu and P. Schniter, "Reduced-Complexity Pilot-aided Decision-directed Tracking of Doubly-Selective Channels," *Proc. Conf. Information Sciences and Systems*, March 2004.

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in:

Detection and Estimation theory	Prof. Philip Schniter
Communication and Information theory	Prof. Hesham El-Gamal

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Wireless Communications

The rapidly evolving global information structure includes broadband wireless communication as a key component. Future wireless links are expected to provide high data rate transmission of multimedia services in high mobility situations.

The reliable communication over wireless channels has been a challenging problem because of the *multipath fading* phenomenon [1]. The transmitted signal arrives at the receiver via multiple propagation paths at different delays. These signal components may add constructively or destructively at the receiver resulting in wide variations in the signal strength and this phenomenon is generally referred as the multipath fading. When the bandwidth of the transmitted signal is large, the fading results in inter-symbol interference (ISI) and the channel becomes frequency-selective.

When there is mobility between the transmitter and receiver, the channel fading characteristics change with time and the channel becomes time-selective. The channels which are both time and frequency selective are referred as doubly-selective channels. A simple way of characterizing these channels is by modeling them as a linear time varying (LTV) filter. The received sequence y_n (after match filtering and

sampling) is related to the transmit sequence t_n (in the complex baseband) by

$$y_n = \sum_{d=0}^{N_h-1} h_{n,d} t_{n-d} + v_n \tag{1.1}$$

where $h_{n,d}$ denotes the fading coefficients at time instant n and v_n is additive white Gaussian noise (AWGN). The fading coefficients $h_{n,d}$ are samples of a random process which can be characterized statistically.

1.2 Channel Estimation Problem

The quality of the channel estimation has a major impact on the overall system performance and hence reliable estimation of doubly-selective channels is well motivated. One of the common ways of acquiring channel state information (CSI) is by transmitting training symbols which are known *a priori* at the receiver. For time-invariant channels, a training sequence is usually sent at the beginning of each transmission burst which may not work well for time-selective channels. This motivates periodic transmission of the training sequence during the transmission.

We consider channel estimation for two different transmission systems which transmit multiple frames. The first system transmits pilots in time multiplexed fashion, where each frame has a data block followed by a pilot block as shown in Fig. 1.1. Now onwards, we refer this system as time-multiplexed pilots (TMP) system.

The second system is a multi-carrier system with frequency multiplexed pilots (hereafter referred as FMP system). In this system, pilot symbols are assigned frequencies different from that of data symbols. We focus on the orthogonal frequency division multiplexing (OFDM) systems.

Channel estimation for OFDM systems was previously considered in, e.g., [2–7]. The works [2–5] considered fading scenarios in which the channel variation within a

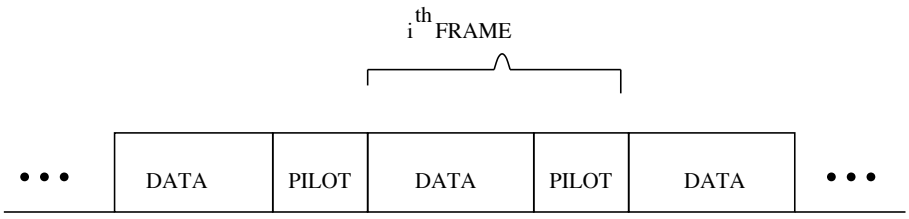

Figure 1.1: Time Multiplexed Pilots System.

frame is negligible. Decision-directed adaptive algorithms for channel tracking were discussed in [2]. Pilot-based linear MMSE (LMMSE) techniques were discussed in [4] 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 [5], including modifications to make the predictor coefficients invariant to the transmitted data.

We focus on the fast fading scenario, in which there is significant variation of the channel within each frame. OFDM systems suffer from inter-carrier interference (ICI) in fast fading channels [8]. For these fast fading scenarios, pilot-based LMMSE and least squares (LS) techniques were proposed in [6] and [7], respectively, and a recursive decision-directed LS technique was given in [9]. In [10], 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.

1.3 Organization and Contribution

In Chapter 2, we give the details of the doubly-selective channels and their statistical characterization. We also present the standard estimation techniques that can be applied for channel tracking in the TMP system. In Chapter 3 and Chapter 4, we present the novel computationally less expensive channel tracking techniques for the TMP and FMP systems in the fast fading doubly-selective channels.

1.3.1 Time Multiplexed Pilots System

A low complexity two-stage decision/pilot aided channel tracking technique (LCP) is presented in Chapter 3. The modifications for LCP are presented which helps in reducing the computational and memory requirements further. We find a benchmark on the prediction performance by transmitting the 'asymptotically optimal' sequence for prediction and using infinite past observations. The simulation results are presented which show that the LCP performance is close to that of standard optimal estimators.

1.3.2 Frequency Multiplexed Pilots System

In Chapter 4, we present channel estimation techniques for OFDM systems which take the ICI effects into account. We present a novel pilot tone selection criterion for OFDM systems and the theoretical MSE calculations show its efficacy for a wide range of Doppler frequencies. We also present a low complexity decision-directed estimation technique and the numerical results are presented which show the good performance of the proposed techniques.

CHAPTER 2

BACKGROUND

2.1 Wireless Channels

Because of the multipath fading, the received signal is modeled as a filtered version of the transmitted signal, with the filter taps being random fading coefficients. In doubly-selective channels, these filter taps are continuously changing with time. There are different representations for the doubly-selective channels and the ones used in the thesis are discussed below.

2.1.1 Time-Lag Representation

Time-lag model is basically the impulse response of the LTV channel, denoted by $h_{n,d},\ n \in \mathbb{Z},\ d \in \{0, .., N_h - 1\}$, where N_h is the number of taps of the channel. Here $h_{n,d}$ denotes the response of the channel at time index n to a Kronecker delta applied at time index $n - d$. One particular realization of the channel is shown in Fig. 2.1.

2.1.2 Doppler-Lag Representation

The time-lag coefficients $h_{n,d}$ within time window $n \in \{0, ..., N - 1\}$ are equivalently represented by Doppler-lag coefficients $\breve{h}_{k,d}$ by taking the DFT across the time

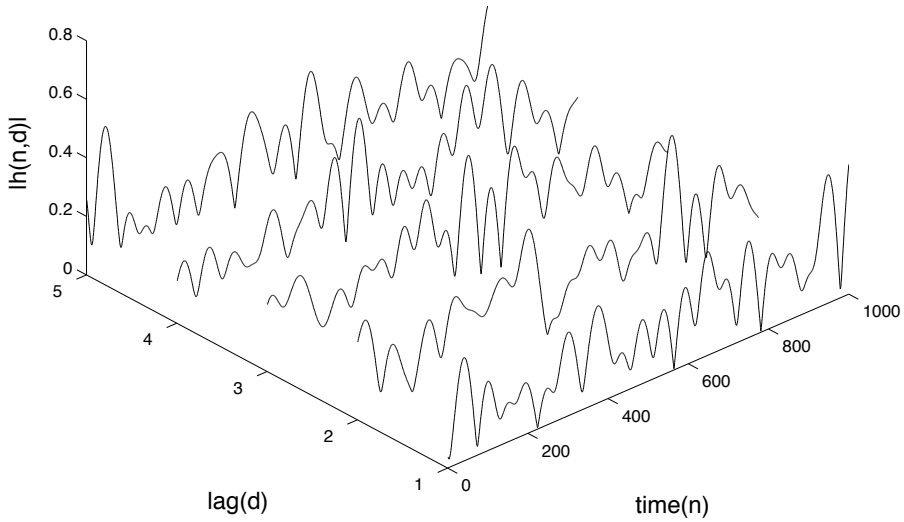

Figure 2.1: A realization of the channel in time-lag domain

dimension. We have,

$$\breve{h}_{k,d} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} h_{n,d} e^{-j\frac{2\pi}{N}nk},\ k \in \{0, ..., N-1\} \tag{2.1}$$

2.2 Statistics of Wireless Channels

Since the fading coefficients are random, they are characterized statistically. One of the common ways of characterizing the random sequence is by its auto-correlation function. We assume the channel fading statistics to be wide sense stationary uncorrelated scattering (WSSUS). Hence, we have

$$E\{h_{n,d} h^*_{n-k,d-m}\} = r_{hh}(k,d)\delta(m), \tag{2.2}$$

where $\delta(\cdot)$ denotes the Kronecker-Delta function. For the Rayleigh fading channel [11], the auto-correlation function is given by

$$r_{hh}(k,d) \quad = \quad \sigma_d^2 J_0(2\pi f_d k) \tag{2.3}$$

where σ_d^2 denotes the variance of the d^{th} path, $J_0(\cdot)$ is the 0^{th} order Bessel function of the first kind and f_d denotes the normalized Doppler frequency. The normalized Doppler frequency characterizes the rate of variation of the channel in time, higher the value of f_d, more rapidly the channel varies. The auto-correlation function (2.3) for different Doppler frequencies is shown in Fig. 2.2. Note that, for higher f_d, the correlation 'decays' fast which means that the rate of channel variation is high. The

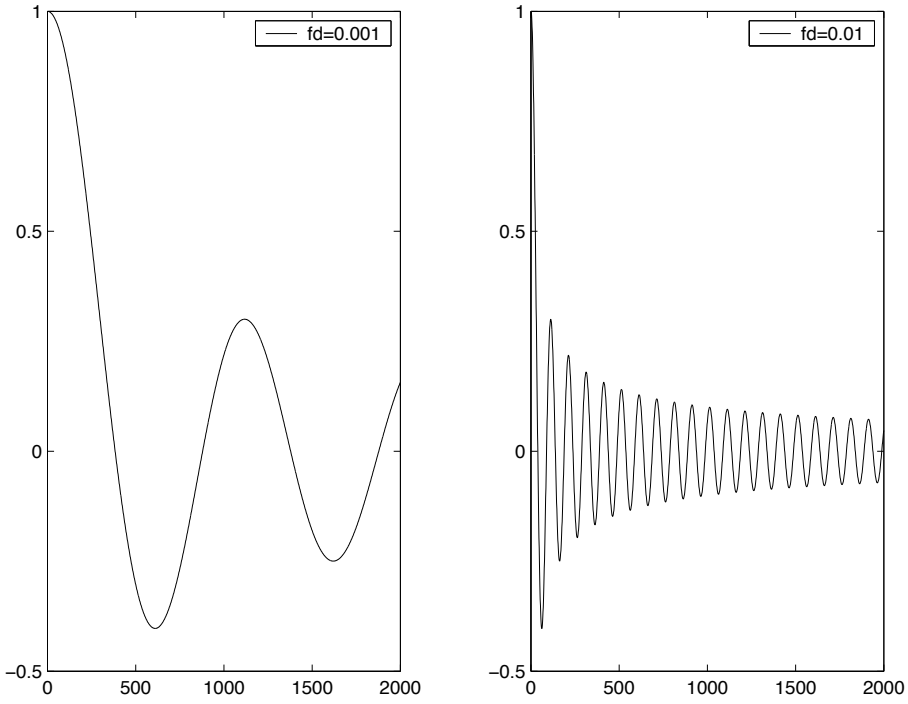

Figure 2.2: $r_{hh}(k,d)$ as a function of k with $\sigma_d^2 = 1$

corresponding power spectral density of the process is given by

$$S_{hh}(f,d) = \begin{cases} \frac{\sigma_d^2}{\pi f_d \sqrt{1-(f/f_d)^2}}, & |f| < f_d \\ 0, & |f| \geq f_d \end{cases} \tag{2.4}$$

So, the Rayleigh fading channel is a band limited random process with the maximum frequency being f_d. In this thesis we consider Rayleigh fading scenario with $\sigma_d^2 = N_h^{-1}$, though the techniques presented are applicable to other WSSUS fading scenarios too.

2.3 System Model for TMP

We consider identification of a doubly-selective channel when the transmitted signal consists of a sequence of frames, and when each frame consists of a data block followed by a pilot block (see Fig. 2.3). 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, where $N_p = 2N_h - 1$ and 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 [12]. We derive some additional optimality results in the Appendix A. The structure of the length-N_d data block is not important; it could be composed of frequency-domain symbols, as in OFDM [13], or time domain symbols, as in single carrier cyclic prefix (SCCP) [14]. $N_f := N_d + N_p$ denotes the frame interval.

We use $\mathcal{T}_{\mathsf{d}}^{(i)} := \left\{t_n^{(i)}\right\}_{n=0}^{N_d-1}$ to denote the data portion of the i^{th} transmission frame and $\mathcal{T}_{\mathsf{p}}^{(i)} := \left\{t_n^{(i)}\right\}_{n=N_d}^{N_d+N_p-1}$ to denote the corresponding pilot portion. The Kronecker pattern implies that $t_{N_d+N_h-1}^{(i)} = \sqrt{2N_h - 1}$ and that all the other elements in $\mathcal{T}_{\mathsf{p}}^{(i)}$ are zero. The complete set of samples transmitted during the i^{th} frame is denoted

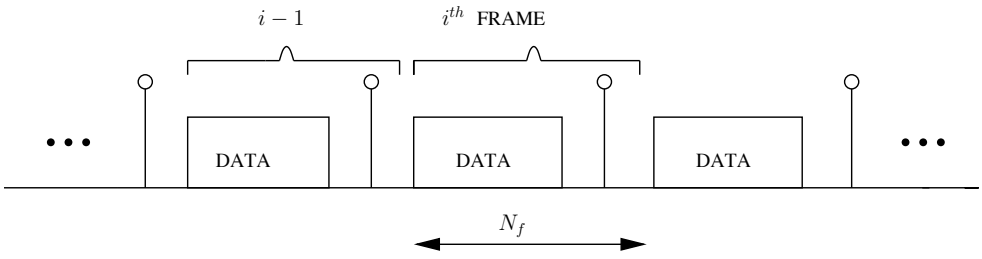

Figure 2.3: Transmission pattern of TMP.

by $\mathcal{T}^{(i)} := \mathcal{T}_{\mathsf{d}}^{(i)} \cup \mathcal{T}_{\mathsf{p}}^{(i)} = \left\{t_n^{(i)}\right\}_{n=0}^{N_f-1}$, 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

$$y_n \;=\; \sum_{d=0}^{N_h-1} h_{n,d}\, t_{n-d} + v_n \quad \text{for} \quad n \in \mathbb{Z}, \tag{2.5}$$

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

$$y_n^{(i)} \;=\; \sum_{d=0}^{N_h-1} h_{n,d}^{(i)} t_{n-d}^{(i)} + v_n^{(i)} \quad \text{for} \quad 0 \le n \le N_f - 1, \tag{2.6}$$

with $t_n^{(i)} = 0$, $-(N_h-1) \le n < 0$.

The following notation will be useful in the sequel. $\mathcal{Y}_{\mathsf{d}}^{(i)} := \left\{y_n^{(i)}\right\}_{n=0}^{N_d+N_h-2}$ and $\mathcal{Y}_{\mathsf{p}}^{(i)} := \left\{y_n^{(i)}\right\}_{n=N_d+N_h-1}^{N_f-1}$ 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, $\mathcal{H}_{\mathsf{d}}^{(i)} := \left\{h_{n,d}^{(i)} \forall d\right\}_{n=0}^{N_d+N_h-2}$, $\mathcal{H}_{\mathsf{p}}^{(i)} := \left\{h_{n,d}^{(i)} \forall d\right\}_{n=N_d+N_h-1}^{N_f-1}$, 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 $\left[\boldsymbol{h}_{\mathsf{d}}^{(i)}\right]_l = h_{\lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}^{(i)}$ and $\left[\boldsymbol{h}_{\mathsf{p}}^{(i)}\right]_l = h_{N_d+N_h-1+\lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}^{(i)}$, and $\boldsymbol{h}^{(i)} = \left[\boldsymbol{h}_{\mathsf{d}}^{(i)t}, \boldsymbol{h}_{\mathsf{p}}^{(i)t}\right]^t$.

Based on the convolution (2.6), $\boldsymbol{y}_{\mathsf{d}}^{(i)}$ can be written

$$\boldsymbol{y}_{\mathsf{d}}^{(i)} \;=\; \boldsymbol{T}_{\mathsf{d}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)} + \boldsymbol{v}_{\mathsf{d}}^{(i)} \tag{2.7}$$

where $\boldsymbol{v}_{\mathsf{d}}^{(i)} = \left[v_0^{(i)}, v_1^{(i)}, \ldots, v_{N_d+N_h-2}^{(i)}\right]^t$ and where $\boldsymbol{T}_{\mathsf{d}}^{(i)}$ is given by

$$\boldsymbol{T}_{\mathsf{d}}^{(i)} \;=\; \begin{bmatrix} t_0^{(i)} & \cdots & t_{-(N_h-1)}^{(i)} & & & & & & \\ & & & t_1^{(i)} & \cdots & t_{-(N_h-2)}^{(i)} & & & \\ & & & & & & \ddots & & \\ & & & & & & & t_{N_d+N_h-2}^{(i)} & \cdots & t_{N_d-1}^{(i)} \end{bmatrix}$$

Let $\boldsymbol{b}_p$ denote the p^{th} column of $\boldsymbol{I}_{N_h}$ and let $\boldsymbol{B}_p$ be the diagonal matrix formed from $\boldsymbol{b}_p$. With $\boldsymbol{G}_{\mathsf{p}} := \sqrt{2N_h - 1}\,[\boldsymbol{B}_0, \ldots, \boldsymbol{B}_{N_h-1}]$, we have

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} \;=\; \boldsymbol{G}_{\mathsf{p}} \boldsymbol{h}_{\mathsf{p}}^{(i)} + \boldsymbol{v}_{\mathsf{p}}^{(i)} \tag{2.8}$$

where $\boldsymbol{v}_{\mathsf{p}}^{(i)} = \left[v_{N_d+N_h-1}^{(i)}, \ldots, v_{N_f-1}^{(i)}\right]^t$. Recalling $\boldsymbol{y}^{(i)} = \left[\boldsymbol{y}_{\mathsf{d}}^{(i)t}, \boldsymbol{y}_{\mathsf{p}}^{(i)t}\right]^t$, (2.7) and (2.8) imply

$$\boldsymbol{y}^{(i)} \;=\; \underbrace{\begin{bmatrix} \boldsymbol{T}_{\mathsf{d}}^{(i)} & \\ & \boldsymbol{G}_{\mathsf{p}} \end{bmatrix}}_{\boldsymbol{T}^{(i)}} \underbrace{\begin{bmatrix} \boldsymbol{h}_{\mathsf{d}}^{(i)} \\ \boldsymbol{h}_{\mathsf{p}}^{(i)} \end{bmatrix}}_{\boldsymbol{h}^{(i)}} + \underbrace{\begin{bmatrix} \boldsymbol{v}_{\mathsf{d}}^{(i)} \\ \boldsymbol{v}_{\mathsf{p}}^{(i)} \end{bmatrix}}_{\boldsymbol{v}^{(i)}} \tag{2.9}$$

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 $\left\{\mathcal{Y}^{(i-k)}\right\}_{k\geq 0}$, current pilots $\mathcal{T}_{\mathsf{p}}^{(i)}$, and past transmission frames $\left\{\mathcal{T}^{(i-k)}\right\}_{k\geq 1}$. The pilots are known *a priori* by the receiver, and, under the assumption of perfect decoding, past transmission frames are known as well.

2.4 Review of Standard Estimators

In this section, we review the "standard" methods of channel identification and tracking. These techniques are MMSE estimators and uses the second-order channel fading statistics for estimation [15]. We also present their draw backs.

2.4.1 Pilot-aided Wiener Estimator

The pilot-aided Wiener estimator (PW) uses only the received samples from the pilot block in the i^{th} frame and M previous frames, i.e., $\left\{\mathcal{Y}_{\mathsf{p}}^{(i-k)}\right\}_{k=0}^{k=M}$. Hence we form the observation vector $\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} = \left[\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \ldots, \boldsymbol{y}_{\mathsf{p}}^{(i-M)t}\right]^t$. Using (2.8), we have

$$\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} = \underbrace{\begin{bmatrix} \boldsymbol{G}_{\mathsf{p}} & & \\ & \ddots & \\ & & \boldsymbol{G}_{\mathsf{p}} \end{bmatrix}}_{\boldsymbol{G}_M} \underbrace{\begin{bmatrix} \boldsymbol{h}_{\mathsf{p}}^{(i)} \\ \vdots \\ \boldsymbol{h}_{\mathsf{p}}^{(i-M)} \end{bmatrix}}_{\underline{\boldsymbol{h}}_{\mathsf{p}}^{(i)}} + \begin{bmatrix} \boldsymbol{v}_{\mathsf{p}}^{(i)} \\ \vdots \\ \boldsymbol{v}_{\mathsf{p}}^{(i-M)} \end{bmatrix}. \tag{2.10}$$

Since $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ and $\underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)}$ are jointly Gaussian, the MMSE estimator is linear and given by [15]

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{pilot}} = \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{p}},\boldsymbol{h}_{\mathsf{d}}}^H \boldsymbol{G}_M^H \left(\boldsymbol{G}_M \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{p}},\underline{\boldsymbol{h}}_{\mathsf{p}}} \boldsymbol{G}_M^H + \sigma_v^2 \boldsymbol{I}\right)^{-1} \underline{\boldsymbol{y}}_{\mathsf{p}}^{(i)} \tag{2.11}$$

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

The estimator coefficients in (2.11) are frame invariant. Hence PW doesn't require matrix inversion at frame rate. The pilot observations give the samples of the channel process with the interval of N_f. If $f_d > \frac{1}{2N_f}$ then the channel is under-sampled and identification breaks down. This limits the MSE performance of PW at high Doppler frequencies.

2.4.2 Pilot-aided Decision-directed Wiener Estimator

The pilot-aided decision-directed Wiener estimator (PDW) uses the pilot block in the i^{th} frame $\mathcal{Y}_{\mathsf{p}}^{(i)}$ as well all received samples in M previous frames $\left\{\mathcal{Y}^{(i-k)}\right\}_{k=1}^{M}$. Here we form the observation vector $\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} = \left[\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \boldsymbol{y}^{(i-1)t}, \ldots, \boldsymbol{y}^{(i-M)t}\right]^t$ and the channel vector $\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} = \left[\boldsymbol{h}_{\mathsf{p}}^{(i)t}, \boldsymbol{h}^{(i-1)t}, \ldots, \boldsymbol{h}^{(i-M)t}\right]^t$. Using (2.8) and (2.9), we have

$$\underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} \;=\; \boldsymbol{T}_{\mathsf{w}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{w}}^{(i)} \tag{2.12}$$

where $\underline{\boldsymbol{v}}_{\mathsf{w}}^{(i)} = \left[\boldsymbol{v}_{\mathsf{p}}^{(i)t}, \boldsymbol{v}^{(i-1)t}, \ldots, \boldsymbol{v}^{(i-M)t}\right]^t$ and
$\boldsymbol{T}_{\mathsf{w}}^{(i)} = \text{blkdiag}\left(\boldsymbol{G}_{\mathsf{p}}, \boldsymbol{T}^{(i-1)}, \boldsymbol{T}^{(i-2)}, \ldots, \boldsymbol{T}^{(i-M)}\right)$. 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 estimator is linear and given by

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\Big|_{\mathsf{wiener}} \;=\; \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\boldsymbol{h}_{\mathsf{d}}}^{H} \boldsymbol{T}_{\mathsf{w}}^{(i)H} \left(\boldsymbol{T}_{\mathsf{w}}^{(i)} \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\underline{\boldsymbol{h}}_{\mathsf{w}}} \boldsymbol{T}_{\mathsf{w}}^{(i)H} + \sigma_v^2 \boldsymbol{I}\right)^{-1} \underline{\boldsymbol{y}}_{\mathsf{w}}^{(i)} \tag{2.13}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\boldsymbol{h}_{\mathsf{d}}} = E\left\{\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\right\}$ and $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{w}},\underline{\boldsymbol{h}}_{\mathsf{w}}} = E\left\{\underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{w}}^{(i)H}\right\}$. In the estimator equation (2.13), the inverted matrix is of size $(MN_f + N_h) \times (MN_f + N_h)$ and the inverse must be computed for each frame index i. Though PDW doesn't suffer from the sampling limitation, it has huge computational complexity. Thus increasing M leads to higher performance but increased complexity.

2.4.3 Pilot-aided Decision-directed Kalman Estimator

In this section, we find the pilot-aided decision-directed Kalman estimator (PDK) by formulating our channel estimation as a Kalman prediction problem [16]. We assign $\boldsymbol{h}^{(i-1)}$ as the current *state* of the channel, $\boldsymbol{h}^{(i)}$ as the next state, and $\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} = \left[\boldsymbol{y}_{\mathsf{d}}^{(i-1)t}, \boldsymbol{y}_{\mathsf{p}}^{(i)t}\right]^t$ as the current observation. The state dynamics can be written as

$$\boldsymbol{h}^{(i)} \;=\; \boldsymbol{A}_{\mathsf{k}} \boldsymbol{h}^{(i-1)} + \boldsymbol{D}_{\mathsf{k}} \boldsymbol{w}_{\mathsf{k}}^{(i-1)} \tag{2.14}$$

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. Using (2.7),

$$\boldsymbol{y}_{\mathsf{d}}^{(i-1)} = \boldsymbol{T}_{\mathsf{d}}^{(i-1)}\boldsymbol{h}^{(i-1)} + \boldsymbol{v}_{\mathsf{d}}^{(i-1)}, \tag{2.15}$$

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} = \boldsymbol{G}\boldsymbol{h}^{(i)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}, \tag{2.16}$$

where $\boldsymbol{G} = [\boldsymbol{0}_{N_h\times(N_d+N_h-1)N_h}, \boldsymbol{G}_{\mathsf{p}}]$. Using (2.14), we can rewrite (2.16) as

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} = \boldsymbol{G}\boldsymbol{A}_{\mathsf{k}}\boldsymbol{h}^{(i-1)} + \boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}. \tag{2.17}$$

With (2.15) and (2.17) we have

$$\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} = \underbrace{\begin{bmatrix}\boldsymbol{T}_{\mathsf{d}}^{(i-1)}\\ \boldsymbol{G}\boldsymbol{A}_{\mathsf{k}}\end{bmatrix}}_{\boldsymbol{C}_{\mathsf{k}}^{(i-1)}}\boldsymbol{h}^{(i-1)} + \underbrace{\begin{bmatrix}\boldsymbol{v}_{\mathsf{d}}^{(i-1)}\\ \boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}\end{bmatrix}}_{\underline{\boldsymbol{v}}_{\mathsf{k}}^{(i-1)}} \tag{2.18}$$

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. From (2.14) and (2.18), the MMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)}, \ldots, \underline{\boldsymbol{y}}_{\mathsf{k}}^{(0)}\}$, denoted by $\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{kalman}}$, is given recursively as

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{kalman}} = \boldsymbol{A}_{\mathsf{k}}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{kalman}} + \boldsymbol{L}_{\mathsf{k}}^{(i-1)}\left(\underline{\boldsymbol{y}}_{\mathsf{k}}^{(i-1)} - \boldsymbol{C}_{\mathsf{k}}^{(i-1)}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{kalman}}\right)$$

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

$$\boldsymbol{L}_{\mathsf{k}}^{(i-1)} = \left(\boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{C}_{\mathsf{k}}^{(i-1)H} + \boldsymbol{D}_{\mathsf{k}}\boldsymbol{S}_{\mathsf{k}}\right)\left(\boldsymbol{C}_{\mathsf{k}}^{(i-1)}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{C}_{\mathsf{k}}^{(i-1)H} + \boldsymbol{R}_{\mathsf{k}}\right)^{-1} \tag{2.19}$$

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

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

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

Note that the Kalman estimator uses *all* previous observations in its prediction of $\boldsymbol{h}^{(i)}$; this is the advantage of the PDK over the PDW. However, the performance of the PDK depends on how well the model (2.14) describes the true evolution of the state process. Increasing the AR model order helps in the making the 'match' of the true and model auto-correlation function 'longer' (See Fig. 2.4). But even with higher model order, note that there is deviation from the true auto-correlation for large values of the 'time difference'. Note also, from (2.19), that the PDK requires a matrix inversion of size $N_f \times N_f$ once per frame.

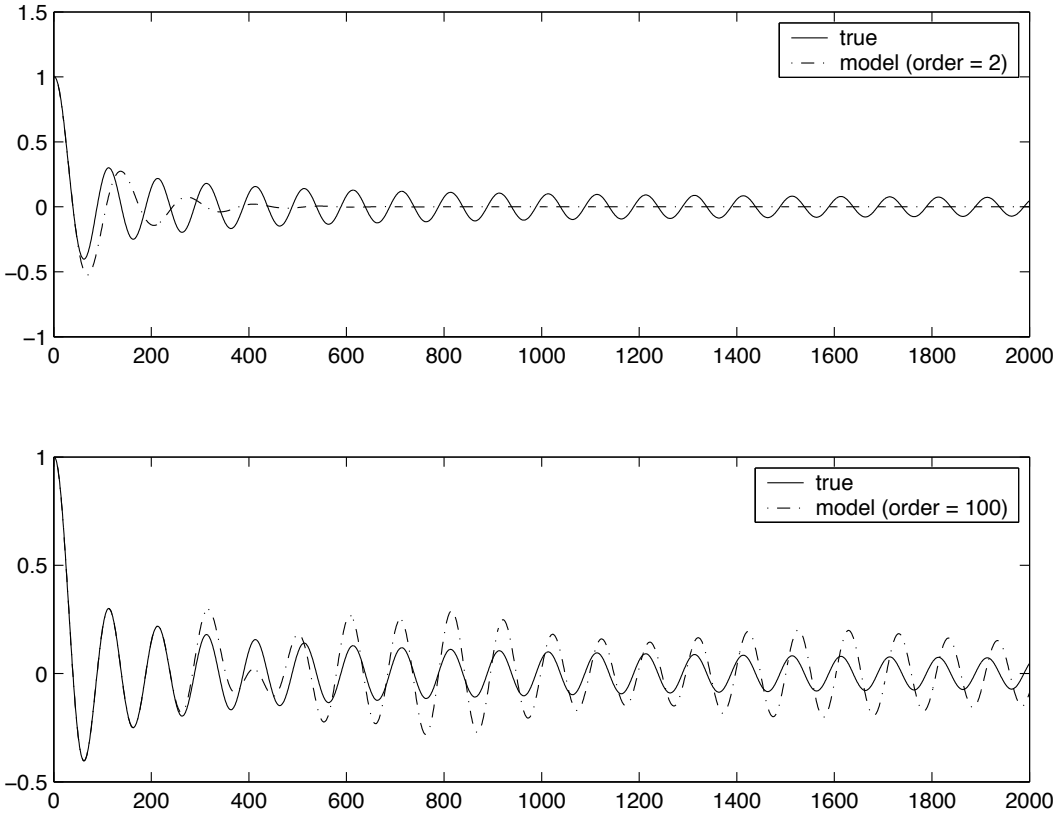

Figure 2.4: Auto-correlation mismatch for $f_d = 0.01$.

CHAPTER 3

LOW COMPLEXITY ESTIMATION FOR TIME MULTIPLEXED PILOTS SYSTEM

3.1 Low Complexity Predictor

In this section, we describe a novel computationally-efficient decision-directed channel tracker (LCP) that does not require large matrix inversions [17]. 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, leading to significant computational savings.

3.1.1 Kalman Smoothing Stage

The smoothed channel estimates of the $(i-1)^{th}$ frame are obtained using Kalman filtering. We divide each frame into $K = \frac{N_f}{L}$ sub-frames of size L and design the Kalman smoother such that it finds the smoothed channel estimates of each particular sub-frame using the received samples during that sub-frame. Since we process only L measurements at a time, this requires only K inversions of size $L \times L$ per frame. Details are provided below.

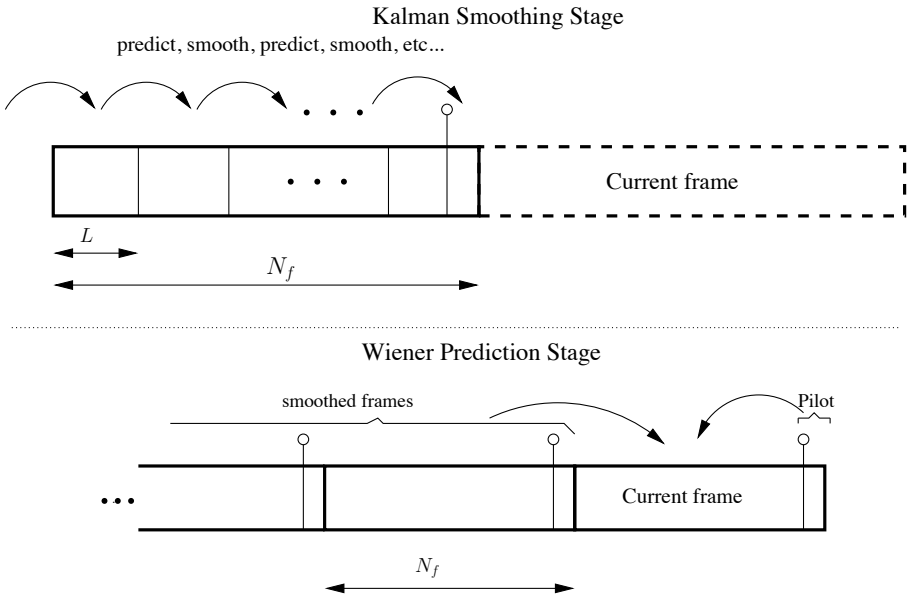

Figure 3.1: Two stages of LCP.

Let $\mathcal{H}^{(i-1,k)}$ and $\mathcal{Y}^{(i-1,k)}$ denote the k^{th} sub-frame's channel coefficients and observations, respectively, with $\mathcal{H}^{(i-1,k)} = \left\{h_{n,d}^{(i-1)} \forall d\right\}_{n=kL}^{(k+1)L-1}$ and $\mathcal{Y}^{(i-1,k)} = \left\{y_n^{(i-1)}\right\}_{n=kL}^{(k+1)L-1}$ for $k \in \{0, \ldots, K-1\}$. The vectors $\boldsymbol{h}^{(i-1,k)}$ and $\boldsymbol{y}^{(i-1,k)}$ are defined element-wise as $\left[\boldsymbol{h}^{(i-1,k)}\right]_l = h_{kL+\lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}^{(i-1)}$ and $\left[\boldsymbol{y}^{(i-1,k)}\right]_l = y_{kL+l}^{(i-1)}$. The smoother's dynamical equation relating the current channel state, $\boldsymbol{h}^{(i-1,k)}$, the next state, $\boldsymbol{h}^{(i-1,k+1)}$ is

$$\boldsymbol{h}^{(i-1,k+1)} \quad = \quad \boldsymbol{A}_{\mathsf{I}} \boldsymbol{h}^{(i-1,k)} + \boldsymbol{D}_{\mathsf{I}} \boldsymbol{w}_{\mathsf{I}}^{(i-1,k)} \tag{3.1}$$

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

$$\boldsymbol{y}^{(i-1,k)} \quad = \quad \boldsymbol{T}_{\mathsf{I}}^{(i-1,k)} \boldsymbol{h}^{(i-1,k)} + \boldsymbol{v}^{(i-1,k)} \tag{3.2}$$

where $\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)}$ is an $L \times LN_h$ matrix with the structure

$$\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)} = \begin{bmatrix} t_{kL}^{(i-1)} & \cdots & t_{kL-(N_h-1)}^{(i-1)} & & & & & \\ & & & t_{kL+1}^{(i-1)} & \cdots & t_{kL+1-(N_h-1)}^{(i-1)} & & \\ & & & & & & \ddots & \\ & & & & & & t_{(k+1)L-1}^{(i-1)} & \cdots & t_{(k+1)L-N_h}^{(i-1)} \end{bmatrix} \quad (3.3)$$

and $\boldsymbol{v}^{(i-1,k)} = \left[v_{kL}^{(i-1)}, \ldots, v_{(k+1)L-1}^{(i-1)}\right]^t$.

For the system described by (3.1) and (3.2), let $\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}$ denote the *smoothed* MMSE estimate of $\boldsymbol{h}^{(i-1,k)}$ using observations $\left\{\boldsymbol{y}^{(i-1,k)}, \ldots, \boldsymbol{y}^{(i-1,0)}, \boldsymbol{y}^{(i-2)}, \ldots, \boldsymbol{y}^{(0)}\right\}$, and let $\hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}$ denote the *predicted* MMSE estimate of $\boldsymbol{h}^{(i-1,k)}$ using the observations $\{\boldsymbol{y}^{(i-1,k-1)}, \ldots, \boldsymbol{y}^{(i-1,0)}, \boldsymbol{y}^{(i-2)}, \ldots, \boldsymbol{y}^{(0)}\}$. The smoothed estimate and smoother gain are [16]

$$\begin{aligned} \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} &= \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} + \boldsymbol{M}_{\mathsf{l}}^{(i-1,k)} \left(\boldsymbol{y}^{(i-1,k)} - \boldsymbol{T}_{\mathsf{l}}^{(i-1,k)} \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)}\right) \\ \boldsymbol{M}_{\mathsf{l}}^{(i-1,k)} &= \boldsymbol{P}_{\mathsf{l}}^{(i-1,k)} \boldsymbol{T}_{\mathsf{l}}^{(i-1,k)H} \left(\boldsymbol{T}_{\mathsf{l}}^{(i-1,k)} \boldsymbol{P}_{\mathsf{l}}^{(i-1,k)} \boldsymbol{T}_{\mathsf{l}}^{(i-1,k)H} + \sigma_v^2 \boldsymbol{I}\right)^{-1} \end{aligned} \quad (3.4)$$

The predicted estimate and $\boldsymbol{P}_{\mathsf{l}}^{(i-1,k)}$ are given recursively as

$$\begin{aligned} \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k)} &= \boldsymbol{A}_{\mathsf{l}} \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k-1)} + \boldsymbol{A}_{\mathsf{l}} \boldsymbol{M}_{\mathsf{l}}^{(i-1,k-1)} \left(\boldsymbol{y}^{(i-1,k-1)} - \boldsymbol{T}_{\mathsf{l}}^{(i-1,k-1)} \hat{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,k-1)}\right) \\ \boldsymbol{P}_{\mathsf{l}}^{(i-1,k)} &= \sigma_{w_{\mathsf{l}}}^2 \boldsymbol{D}_{\mathsf{l}} \boldsymbol{D}_{\mathsf{l}}^H + \boldsymbol{A}_{\mathsf{l}} \boldsymbol{P}_{\mathsf{l}}^{(i-1,k-1)} \boldsymbol{A}_{\mathsf{l}}^H - \boldsymbol{A}_{\mathsf{l}} \boldsymbol{M}_{\mathsf{l}}^{(i-1,k-1)} \boldsymbol{T}_{\mathsf{l}}^{(i-1,k-1)} \boldsymbol{P}_{\mathsf{l}}^{(i-1,k-1)} \boldsymbol{A}_{\mathsf{l}}^H \end{aligned}$$

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 by collecting the smoothed estimates of the channel in all K subframes as $\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} = \left[\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,0)t}, \ldots, \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,K-1)t}\right]^t$. This smoothed channel estimate vector is related to the true channel vector $\boldsymbol{h}^{(i-1)}$ and the smoothing error $\boldsymbol{e}_{\mathsf{l}}^{(i-1)}$ via

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} = \boldsymbol{h}^{(i-1)} + \boldsymbol{e}_{\mathsf{l}}^{(i-1)}. \quad (3.5)$$

3.1.2 Wiener Prediction Stage

We form the prediction stage observation vector as $\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} = \left[\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)t}, \ldots, \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-M)t}\right]^t$. From (2.16) and (3.5), we have

$$\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} = \underbrace{\begin{bmatrix} \boldsymbol{G} & & & \\ & \boldsymbol{I}_{N_f N_h} & & \\ & & \ddots & \\ & & & \boldsymbol{I}_{N_f N_h} \end{bmatrix}}_{\boldsymbol{B}} \underbrace{\begin{bmatrix} \boldsymbol{h}^{(i)} \\ \boldsymbol{h}^{(i-1)} \\ \vdots \\ \boldsymbol{h}^{(i-M)} \end{bmatrix}}_{\underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)}} + \underbrace{\begin{bmatrix} \boldsymbol{v}_{\mathsf{p}}^{(i)} \\ \boldsymbol{e}_{\mathsf{l}}^{(i-1)} \\ \vdots \\ \boldsymbol{e}_{\mathsf{l}}^{(i-M)} \end{bmatrix}}_{\underline{\boldsymbol{v}}_{\mathsf{l}}^{(i)}}$$

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.

1. The smoothing errors are white with variance σ_v^2,

$$E\{\boldsymbol{e}_{\mathsf{l}}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \sigma_v^2 \boldsymbol{I}\delta(p-q) \tag{3.6}$$

$$E\{\boldsymbol{v}_{\mathsf{p}}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \boldsymbol{0}\,\forall\, p \neq q. \tag{3.7}$$

2. The smoothing errors are uncorrelated with the channel,

$$E\{\boldsymbol{h}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \boldsymbol{0}\,\forall\, p,\ q. \tag{3.8}$$

With (3.6)-(3.8), the predictor equation is

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{lcp}} = \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{l}},\boldsymbol{h}_{\mathsf{d}}}^H \boldsymbol{B}^H \left(\boldsymbol{B}\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{l}},\underline{\boldsymbol{h}}_{\mathsf{l}}}\boldsymbol{B}^H + \sigma_v^2\boldsymbol{I}\right)^{-1} \underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}, \tag{3.9}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{l}},\boldsymbol{h}_{\mathsf{d}}} = E\{\underline{\boldsymbol{h}}_{\mathsf{l}}^{(i)}\boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$ 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 (3.9) 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 (3.4). The choice of L is a tradeoff between performance and complexity.

3.2 LCP Modifications

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

3.2.1 LCP with Doppler-Lag Coefficients

The Doppler-lag channel coefficients are obtained by taking the DFT of the time-lag coefficients across the time dimension. With $\boldsymbol{F}$ denoting the $N_f \times N_f$ unitary DFT matrix, the i^{th}-frame Doppler-lag coefficients $\boldsymbol{h}_{\mathsf{dl}}^{(i)}$ are obtained as

$$\boldsymbol{h}_{\mathsf{dl}}^{(i)} \;=\; (\boldsymbol{F} \otimes \boldsymbol{I}_{N_h})\, \boldsymbol{h}^{(i)} \tag{3.10}$$

where $\otimes$ denotes Kronecker product. Out of these, only a few coefficients have significant energy [10]. In LCP with Doppler-lag coefficients (LCPD), we use only these significant Doppler coefficients of the past smoothed frames. Let $\boldsymbol{S}$ be the selection matrix,

$$\boldsymbol{S} \;=\; \begin{bmatrix} \boldsymbol{I}_{r+1} & \boldsymbol{0}_{(r+1)\times(N_f-2r-1)} & \boldsymbol{0}_{(r+1)\times r} \\ \boldsymbol{0}_{r\times(r+1)} & \boldsymbol{0}_{r\times(N_f-2r-1)} & \boldsymbol{I}_r \end{bmatrix} \tag{3.11}$$

with $2r+1$ being the number of significant Doppler-coefficients of each tap. The significant Doppler coefficients in the i^{th} frame $\boldsymbol{h}_{\mathsf{ds}}^{(i)}$ are

$$\begin{aligned} \boldsymbol{h}_{\mathsf{ds}}^{(i)} &= (\boldsymbol{S} \otimes \boldsymbol{I}_{N_h})\boldsymbol{h}_{\mathsf{dl}}^{(i)} \\ &= (\boldsymbol{S} \otimes \boldsymbol{I}_{N_h})(\boldsymbol{F} \otimes \boldsymbol{I}_{N_h})\boldsymbol{h}^{(i)} \\ &= \underbrace{(\boldsymbol{S}\boldsymbol{F} \otimes \boldsymbol{I}_{N_h})}_{\boldsymbol{J}}\boldsymbol{h}^{(i)}. \end{aligned} \tag{3.12}$$

The significant smoothed Doppler coefficients of the $(i-1)^{th}$ frame are

$$\begin{aligned}\tilde{\boldsymbol{h}}_{\mathsf{ds}}^{(i-1)} &= \boldsymbol{J}\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} \\ &= \boldsymbol{h}_{\mathsf{ds}}^{(i-1)} + \underbrace{\boldsymbol{J}\boldsymbol{e}_{\mathsf{l}}^{(i-1)}}_{\boldsymbol{e}_{\mathsf{ds}}^{(i-1)}}. \end{aligned} \tag{3.13}$$

With (3.6), we have [18]

$$\begin{aligned} E\{\boldsymbol{e}_{\mathsf{ds}}^{(i-1)}\boldsymbol{e}_{\mathsf{ds}}^{(i-1)H}\} &= E\{\boldsymbol{J}\boldsymbol{e}_{\mathsf{l}}^{(i)}\boldsymbol{e}_{\mathsf{l}}^{(i)H}\boldsymbol{J}^H\} \\ &= (\boldsymbol{S}\boldsymbol{F}\otimes\boldsymbol{I}_{N_h})(\sigma_v^2\boldsymbol{I})(\boldsymbol{S}\boldsymbol{F}\otimes\boldsymbol{I}_{N_h})^H \\ &= \sigma_v^2(\boldsymbol{S}\boldsymbol{F}\otimes\boldsymbol{I}_{N_h})(\{\boldsymbol{S}\boldsymbol{F}\}^H\otimes\boldsymbol{I}_{N_h}) \\ &= \sigma_v^2(\boldsymbol{S}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{S}^H)\otimes\boldsymbol{I}_{N_h} \\ &= \sigma_v^2\boldsymbol{I}_{(2r+1)N_h}. \end{aligned} \tag{3.14}$$

We form the LCPD observation vector as $\underline{\boldsymbol{y}}_{\mathsf{ds}}^{(i)} = \left[\boldsymbol{y}_{\mathsf{p}}^{(i)t}, \tilde{\boldsymbol{h}}_{\mathsf{ds}}^{(i-1)t}, \ldots, \tilde{\boldsymbol{h}}_{\mathsf{ds}}^{(i-M)t}\right]^t$, yielding

$$\underline{\boldsymbol{y}}_{\mathsf{ds}}^{(i)} = \underbrace{\begin{bmatrix}\boldsymbol{G} & & \\ & \boldsymbol{J} & \\ & & \ddots \\ & & & \boldsymbol{J}\end{bmatrix}}_{\boldsymbol{A}_{\mathsf{ds}}} \underbrace{\begin{bmatrix}\boldsymbol{h}^{(i)} \\ \boldsymbol{h}^{(i-1)} \\ \vdots \\ \boldsymbol{h}^{(i-M)}\end{bmatrix}}_{\underline{\boldsymbol{h}}_{\mathsf{ds}}^{(i)}} + \underbrace{\begin{bmatrix}\boldsymbol{v}_{\mathsf{p}}^{(i)} \\ \boldsymbol{e}_{\mathsf{ds}}^{(i-1)} \\ \vdots \\ \boldsymbol{e}_{\mathsf{ds}}^{(i-M)}\end{bmatrix}}_{\underline{\boldsymbol{e}}_{\mathsf{ds}}^{(i)}}. \tag{3.15}$$

Using (3.6)-(3.8) and (3.14), the MMSE estimate of $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ using $\underline{\boldsymbol{y}}_{\mathsf{ds}}^{(i)}$ is given by

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\big|_{\mathsf{lcpd}} = \boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{ds}},\boldsymbol{h}_{\mathsf{d}}}^H\boldsymbol{A}_{\mathsf{ds}}^H\left(\boldsymbol{A}_{\mathsf{ds}}\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{ds}}\underline{\boldsymbol{h}}_{\mathsf{ds}}}\boldsymbol{A}_{\mathsf{ds}}^H + \sigma_v^2\boldsymbol{I}\right)^{-1}\underline{\boldsymbol{y}}_{\mathsf{ds}}^{(i)}$$

where $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{ds}},\boldsymbol{h}_{\mathsf{d}}} = E\{\underline{\boldsymbol{h}}_{\mathsf{ds}}^{(i)}\boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$ and $\boldsymbol{R}_{\underline{\boldsymbol{h}}_{\mathsf{ds}},\underline{\boldsymbol{h}}_{\mathsf{ds}}} = E\{\underline{\boldsymbol{h}}_{\mathsf{ds}}^{(i)}\underline{\boldsymbol{h}}_{\mathsf{ds}}^{(i)H}\}$. Since, typically, $(2r+1) \ll N_f$, LCPD has a reduced computational and memory load relative to LCP.

3.2.2 LCP with Kalman Prediction

The prediction stage of LCP was based on Wiener prediction using smoothed estimates of M previous frames. Here we present LCP with Kalman prediction (LCKP).

As stated before, the Kalman predictor incorporates *all* previously smoothed estimates, giving an advantage over the Wiener predictor *if* the dynamical equation—in this case (2.14)—models the true evolution of the state process. For prediction, the current observation can be written as

$$\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} = \left[\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)t}, \boldsymbol{y}_{\mathsf{p}}^{(i)t}\right]^t \tag{3.16}$$

where $\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)}$ is the smoothed channel estimate during the data portion of the $(i-1)^{th}$ frame:

$$\begin{aligned}\tilde{\boldsymbol{h}}_{\mathsf{ld}}^{(i-1)} &= \underbrace{\left[\boldsymbol{I}_{(N_d+N_h-1)N_h} \quad \boldsymbol{0}_{(N_d+N_h-1)N_h \times N_h^2}\right]}_{\boldsymbol{B}_{\mathsf{lk}}} \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} \\ &= \boldsymbol{B}_{\mathsf{lk}}\boldsymbol{h}^{(i-1)} + \boldsymbol{B}_{\mathsf{lk}}\boldsymbol{e}_{\mathsf{l}}^{(i-1)}. \end{aligned} \tag{3.17}$$

Using (2.14), (2.16), and (3.17), $\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)}$ can be rewritten as

$$\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} = \underbrace{\begin{bmatrix}\boldsymbol{B}_{\mathsf{lk}} \\ \boldsymbol{G}\boldsymbol{A}_{\mathsf{k}}\end{bmatrix}}_{\boldsymbol{C}_{\mathsf{lk}}} \boldsymbol{h}^{(i-1)} + \underbrace{\begin{bmatrix}\boldsymbol{B}_{\mathsf{lk}}\boldsymbol{e}_{\mathsf{l}}^{(i-1)} \\ \boldsymbol{v}_{\mathsf{p}}^{(i)} + \boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)}\end{bmatrix}}_{\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)}} \tag{3.18}$$

Defining $\boldsymbol{R}_{\mathsf{lk}} = E\{\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)}\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)H}\}$ and $\boldsymbol{S}_{\mathsf{lk}} = E\{\boldsymbol{w}_{\mathsf{k}}^{(i-1)}\underline{\boldsymbol{v}}_{\mathsf{lk}}^{(i-1)H}\}$, approximations (3.6)-(3.8) imply

$$\boldsymbol{R}_{\mathsf{lk}} = \begin{bmatrix}\sigma_v^2\boldsymbol{B}_{\mathsf{lk}}\boldsymbol{B}_{\mathsf{lk}}^H & \\ & \sigma_v^2\boldsymbol{I} + \sigma_{w_{\mathsf{k}}}^2\boldsymbol{G}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{D}_{\mathsf{k}}^H\boldsymbol{G}^H\end{bmatrix} \tag{3.19}$$

$$\boldsymbol{S}_{\mathsf{lk}} = \left[\boldsymbol{0}_{N_fN_h \times (N_d+N_h-1)N_h} \quad \sigma_{w_{\mathsf{k}}}^2\boldsymbol{D}_{\mathsf{k}}^H\boldsymbol{G}^H\right]. \tag{3.20}$$

Since constant matrices are used in the dynamical equation (2.14), the observation equation (3.18), and the correlation matrix definitions (3.19) and (3.20), the *steady state* Kalman predictor suffices [16]. Denoting $\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{lckp}}$ as the MMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\boldsymbol{y}_{\mathsf{lk}}^{(i-1)}, \ldots, \boldsymbol{y}_{\mathsf{lk}}^{(0)}\}$, the recursive steady state predictor

equations are

$$\begin{aligned}\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{lckp}} &= \boldsymbol{A}_{\mathsf{k}}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{lckp}} + \boldsymbol{L}_{\infty}\left(\underline{\boldsymbol{y}}_{\mathsf{lk}}^{(i-1)} - \boldsymbol{C}_{\mathsf{lk}}\hat{\boldsymbol{h}}^{(i-1)}\big|_{\mathsf{lckp}}\right)\\ \boldsymbol{L}_{\infty} &= \left(\boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\infty}\boldsymbol{C}_{\mathsf{lk}} + \boldsymbol{D}_{\mathsf{k}}\boldsymbol{S}_{\mathsf{lk}}\right)\left(\boldsymbol{C}_{\mathsf{lk}}\boldsymbol{P}_{\infty}\boldsymbol{C}_{\mathsf{lk}}^{H} + \boldsymbol{R}_{\mathsf{lk}}\right)^{-1}\end{aligned}$$

with symmetric, positive semi-definite Riccati solution

$$\begin{aligned}\boldsymbol{P}_{\infty} &= \sigma_{w_{\mathsf{k}}}^{2}\boldsymbol{D}_{\mathsf{k}}\boldsymbol{D}_{\mathsf{k}}^{H} + \boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\infty}\boldsymbol{A}_{\mathsf{k}}^{H} - \left(\boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\infty}\boldsymbol{C}_{\mathsf{lk}} + \boldsymbol{D}_{\mathsf{k}}\boldsymbol{S}_{\mathsf{lk}}\right)\\ &\quad\times\left(\boldsymbol{C}_{\mathsf{lk}}\boldsymbol{P}_{\infty}\boldsymbol{C}_{\mathsf{lk}}^{H} + \boldsymbol{R}_{\mathsf{lk}}\right)^{-1}\left(\boldsymbol{C}_{\mathsf{lk}}\boldsymbol{P}_{\infty}\boldsymbol{A}_{\mathsf{k}}^{H} + \boldsymbol{S}_{\mathsf{lk}}^{H}\boldsymbol{D}_{\mathsf{lk}}^{H}\right).\end{aligned}$$

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.

3.3 Persistent Training and Prediction

In this section, we consider a reference scheme "persistent training and prediction" (PTP), in which we consider channel estimation with a persistent training sequence. The training sequence is chosen to minimize the approximate expression for MSE, and we use infinite past observations for prediction. The performance of PTP will give the benchmark of the performance of our estimators.

3.3.1 Training Sequence Design

We search for the MSE-minimizing training sequence for doubly-selective channel parameter estimation. The estimator is assumed to know the pilots in the current frame and both data and pilots in M past frames, for the frame structure of Fig. 2.3. For simplicity, though, we assume the frame duration is an integer multiple of the

delay spread N_h, i.e., $N_f = QN_h$ for $Q \in \mathbb{Z}$, and that the channel is *block-fading*: the coefficients are fixed within N_h-sample temporal blocks but change arbitrarily between blocks. In addition, we focus on the high SNR regime, in that we find the sequence which minimizes a truncated series expansion of MSE rather than MSE itself. Finally, we restrict ourselves to N_f-periodic transmissions, since we are interested in estimation of $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ at every frame index i.

Adapting (2.9) to the block fading model, the i^{th}-frame observation becomes

$$\boldsymbol{y}^{(i)} \;=\; \boldsymbol{T}_{\mathsf{b}}^{(i)}\boldsymbol{h}_{\mathsf{b}}^{(i)} + \boldsymbol{v}^{(i)}.$$

where the block-fading coefficients $\boldsymbol{h}_{\mathsf{b}}^{(i)}$ are defined element-wise as $[\boldsymbol{h}_{\mathsf{b}}^{(i)}]_l := h^{(i)}_{N_h\lfloor l/N_h\rfloor,\langle l\rangle_{N_h}}$, and where $\boldsymbol{T}_{\mathsf{b}}^{(i)} := \text{blkdiag}\big(\boldsymbol{T}_0^{(i)},\ldots,\boldsymbol{T}_{Q-1}^{(i)}\big)$ for Toeplitz

$$\boldsymbol{T}_q^{(i)} \;:=\; \begin{bmatrix} t^{(i)}_{qN_h} & \cdots & t^{(i)}_{qN_h-N_h+1} \\ \vdots & \ddots & \vdots \\ t^{(i)}_{qN_h+N_h-1} & \cdots & t^{(i)}_{qN_h} \end{bmatrix}.$$

Note that the assumed frame structure guarantees $\boldsymbol{T}_{Q-1}^{(i)} = \sqrt{2N_h-1}\,\boldsymbol{I}_{N_h}$. Since N_f-periodicity implies that $\boldsymbol{T}_{\mathsf{b}}^{(i)}$ is invariant to i, we henceforth drop the superscript notation on transmission matrices.

The observations due to the 'pilot' in the i^{th} frame is written as

$$\boldsymbol{y}_{\mathsf{p}}^{(i)} \;=\; \boldsymbol{T}_{Q-1}\boldsymbol{h}_{\mathsf{p}}^{(i)} + \boldsymbol{v}_{\mathsf{p}}^{(i)}$$

where $\boldsymbol{h}_{\mathsf{p}}^{(i)} = [h^{(i)}_{N_f-N_h,0},\ldots h^{(i)}_{N_f-N_h,N_h-1}]^t$ and $\boldsymbol{v}_{\mathsf{p}}^{(i)} = [v^{(i)}_{N_f-N_h},\ldots,v^{(i)}_{N_f-1}]^t$.

The multi-frame observation $\underline{\boldsymbol{y}}^{(i)} := \big[\boldsymbol{y}_{\mathsf{p}}^{(i)t},\boldsymbol{y}^{(i-1)t},\ldots,\boldsymbol{y}^{(i-M)t}\big]^t$ can be written

$$\underline{\boldsymbol{y}}^{(i)} \;=\; \underline{\boldsymbol{T}}_{\mathsf{b}}\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{b}}^{(i)},$$

where $\underline{\boldsymbol{T}}_{\mathsf{b}} := \text{blkdiag}(\boldsymbol{T}_{Q-1},\ \boldsymbol{I}_M \otimes \boldsymbol{T}_{\mathsf{b}})$, $\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} := \big[\boldsymbol{h}_{\mathsf{p}}^{(i)t},\boldsymbol{h}_{\mathsf{b}}^{(i-1)t},\ldots,\boldsymbol{h}_{\mathsf{b}}^{(i-M)t}\big]^t$ and $\underline{\boldsymbol{v}}_{\mathsf{b}}^{(i)} := \big[\boldsymbol{v}_{\mathsf{p}}^{(i)t},\boldsymbol{v}^{(i-1)t},\ldots,\boldsymbol{v}^{(i-M)t}\big]^t$.

The MMSE estimate of $\boldsymbol{h}_{\mathsf{d}}^{(i)}$ using $\underline{\boldsymbol{y}}^{(i)}$ is given by

$$\hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)} \;=\; \boldsymbol{P}^H \underline{\boldsymbol{T}}_{\mathsf{b}}^H \left(\underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{R}_{\mathsf{b}} \underline{\boldsymbol{T}}_{\mathsf{b}}^H + \sigma_v^2 \boldsymbol{I}\right)^{-1} \underline{\boldsymbol{y}}^{(i)}$$

where $\boldsymbol{R}_{\mathsf{b}} := E\{\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)H}\}$ and $\boldsymbol{P} := E\{\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$. Defining $\boldsymbol{R}_{\mathsf{d}} := E\{\boldsymbol{h}_{\mathsf{d}}^{(i)} \boldsymbol{h}_{\mathsf{d}}^{(i)H}\}$, the total MSE is given by

$$\begin{aligned}
(\sigma_e^{(i)})^2 \;&:=\; \mathrm{tr}\left(E\left[\left(\boldsymbol{h}_{\mathsf{d}}^{(i)} - \hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\right)\left(\boldsymbol{h}_{\mathsf{d}}^{(i)} - \hat{\boldsymbol{h}}_{\mathsf{d}}^{(i)}\right)^H\right]\right) \\
&=\; \mathrm{tr}\left(\boldsymbol{R}_{\mathsf{d}} - \boldsymbol{P}^H \underline{\boldsymbol{T}}_{\mathsf{b}}^H \left(\underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{R}_{\mathsf{b}} \underline{\boldsymbol{T}}_{\mathsf{b}}^H + \sigma_v^2 \boldsymbol{I}\right)^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{P}\right) \\
&=\; \mathrm{tr}\left(\boldsymbol{R}_{\mathsf{d}} - \boldsymbol{P}^H \left(\boldsymbol{R}_{\mathsf{b}} + \sigma_v^2 \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} \boldsymbol{P}\right). \qquad (3.21)
\end{aligned}$$

At high SNR, the inverse in (3.21) can be well approximated using a truncated power series expansion [19]:

$$\begin{aligned}
&\left(\boldsymbol{R}_{\mathsf{b}} + \sigma_v^2 \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} \\
&\quad=\; \left(\boldsymbol{I} + \sigma_v^2 \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} \boldsymbol{R}_{\mathsf{b}}^{-1} \\
&\quad\approx\; \boldsymbol{R}_{\mathsf{b}}^{-1} - \sigma_v^2 \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{R}_{\mathsf{b}}^{-1}. \qquad (3.22)
\end{aligned}$$

Equations (3.21) and (3.22) suggest that the "data" sequence should be chosen to minimize, subject to unit power constraint,

$$\mathrm{tr}\left(\boldsymbol{P}^H \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{R}_{\mathsf{b}}^{-1} \boldsymbol{P}\right),$$

i.e., $\mathrm{tr}\left(\underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{Z}\right)$ for $\boldsymbol{Z} := \boldsymbol{R}_{\mathsf{b}}^{-1} \boldsymbol{P} \boldsymbol{P}^H \boldsymbol{R}_{\mathsf{b}}^{-1}$.

For $k \in \{0, \ldots, MQ\}$, let $\boldsymbol{Z}_k$ denote the k^{th} $N_h \times N_h$ diagonal sub-matrix of $\boldsymbol{Z}$, i.e., $\boldsymbol{Z}_k = [\boldsymbol{Z}]_{kN_h:kN_h+N_h-1,kN_h:kN_h+N_h-1}$. Because $\underline{\boldsymbol{T}}_{\mathsf{b}}$ is block diagonal with $N_h \times N_h$ blocks, we have

$$\mathrm{tr}\left(\underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{Z}\right) \geq \sum_{k=0}^{M} \lambda_{Q-1}^{-1} \,\mathrm{tr}(\boldsymbol{Z}_{kQ}) + \sum_{k=0}^{Q-2} \lambda_k^{-1} z(k) \qquad (3.23)$$

where λ_k denotes the *maximum* eigenvalue of $\boldsymbol{T}_k\boldsymbol{T}_k^H$ and $z(k) := \sum_{l=0}^{M-1} \mathrm{tr}(\boldsymbol{Z}_{k+lQ+1})$. The first term on the right of (3.23) is contributed by pilots and the second by data. Equality in (3.23) is achieved iff *all* eigenvalues of $\boldsymbol{T}_k\boldsymbol{T}_k^H$ equal λ_k. Assuming that this is the case—(we shall give an example) we minimize the second term under the unit-power constraint. Note that all eigenvalues of $\boldsymbol{T}_{Q-1}\boldsymbol{T}_{Q-1}^H$ equal $\lambda_{Q-1} = 2N_h - 1$. Using Lagrange optimization to minimize $\sum_{k=0}^{Q-2} \lambda_k^{-1} z(k)$ subject to $\sum_{k=0}^{Q-2} \lambda_k = N_f - 2N_h + 1$, we find the optimal values to be

$$\lambda_k \;=\; \frac{(N_f - 2N_h + 1)}{\sum_{l=0}^{Q-2} \sqrt{z(l)}} \sqrt{z(k)},\; k \in \{0, \ldots, Q-2\} \tag{3.24}$$

where $z(k) \geq 0\; \forall k$ since $\boldsymbol{Z}$ is positive semi-definite. Now, consider the frame sequence

$$t_n^{(i-1)} \;=\; \sqrt{\lambda_{\lfloor \frac{n}{N_h} \rfloor}}\, \delta\big(\langle n \rangle_{N_h}\big) \quad n \in \{0, \ldots, N_f - 2N_h\},$$

with $\{\lambda_k\}_{k=0}^{Q-2}$ from (3.24). This sequence satisfies the criterion that all the eigen values of $\boldsymbol{T}_k\boldsymbol{T}_k^H$ are equal to the optimal values found in (3.24) and hence achieves the minimum value of the lower bound (3.23) on the truncated MSE expansion.

We find the the performance of PTP by letting M approach ∞, using an IIR Wiener estimator and transmitting the sequence found from the previous section. This will give the benchmark for the prediction techniques considered earlier, since we are transmitting an optimal sequence in minimizing the approximate MSE and we are predicting the channel using infinite number of past observations.

3.4 Simulation Results

We consider frame size $N_f = 80$ and channel delay spread $N_h = 8$. For each estimation technique, the required matrix inversion size is given in Table 3.1. The nominal values of M and L in our simulations are 2 and 10 respectively. Though

Technique	*Size of matrix inversion per frame*
PW	n.a.
PDW	$(MN_f + N_h) \times (MN_f + N_h)$
PDK	$N_f \times N_f$
LCP,LCPD,LCKP	$L \times L$

Table 3.1: Relative Algorithm Complexity.

LCP, LCPD and LCKP require $K = \frac{N_f}{L}$ matrix inversions (of size L) per frame, the $\mathcal{O}(N^3)$ matrix inversion complexity rule implies that LCP, LCPD and LCKP are more computationally efficient than PDW and PDK. For all simulations, MSE performance is averaged over at least 1000 channel realizations, where each realization spans at least 10 frames.

In Fig. 3.2, we plot MSE versus SNR at $f_d = 0.01$, and in Fig. 3.3, we plot MSE versus f_d at SNR = 15dB. When $f_d > (2N_f)^{-1} \approx 0.006$, it is seen that the performance of PW is poor. For all f_d, LCP performance is relatively close to PDW and PDK. For SNRs below about 15dB, LCP, PDW, and PDK are relatively close to the PTP benchmark.

In Fig. 3.4, we show the effect of M, the number of previous frames, on LCP and PDW. Increasing M improves both PDW and LCP performance at the cost of increased complexity. In Fig. 3.5, we show the effect of L, the sub-frame size, on LCP performance. Increasing L improves Kalman smoothing performance at the cost of increased complexity.

In Fig. 3.6, we compare the MSE performance of LCP and LCPD for different values of L. The number of significant Doppler coefficients are chosen according to $r = \lceil N_f f_d + 4 \rceil = 5$. The results show that the MSE performance of LCPD is very

close to that of LCP, even though the number of smoothed coefficients employed in prediction was reduced by a factor of $\frac{N_f}{2r+1} \approx 7$.

In Fig. 3.7, we compare LCKP to LCP and PDK. As expected, LCKP performs better than LCP, though the difference is significant only at high SNR. Also as expected, LCKP performs worse than PDK, though the performance gap would decrease with an increase in L.

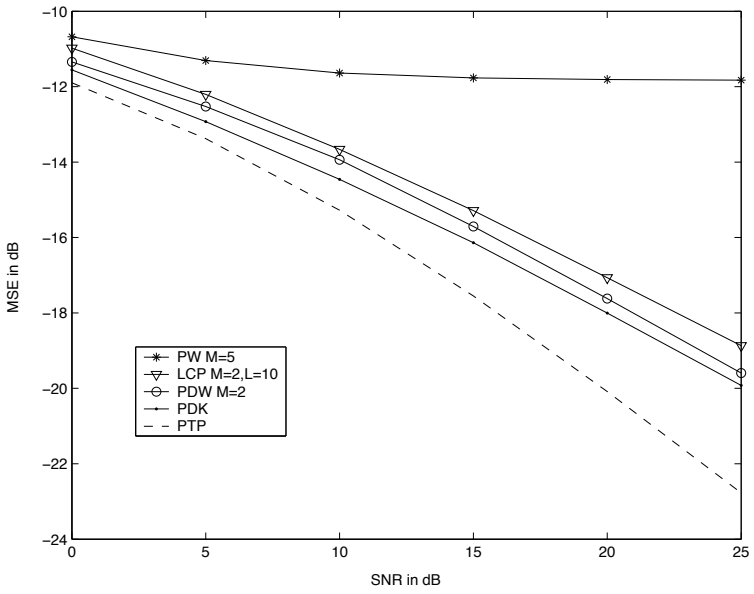

Figure 3.2: Comparison of estimation techniques for $f_d = 0.01$.

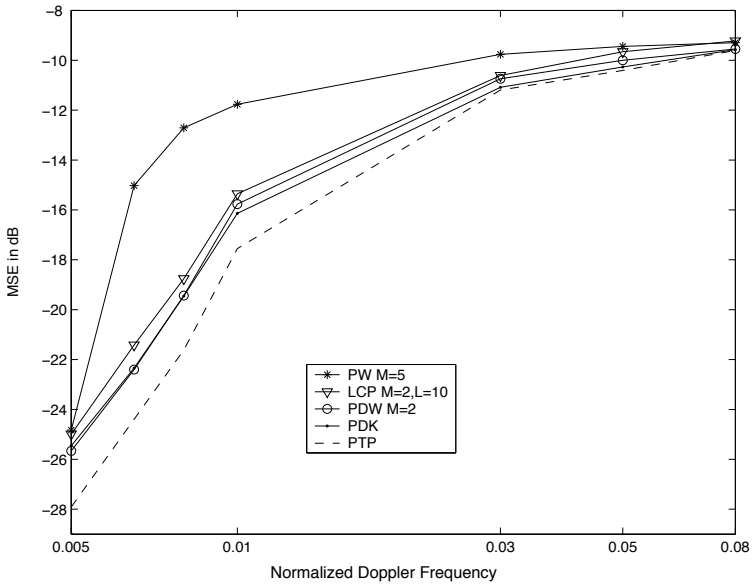

Figure 3.3: Effect of f_d on MSE performance at $SNR = 15dB$.

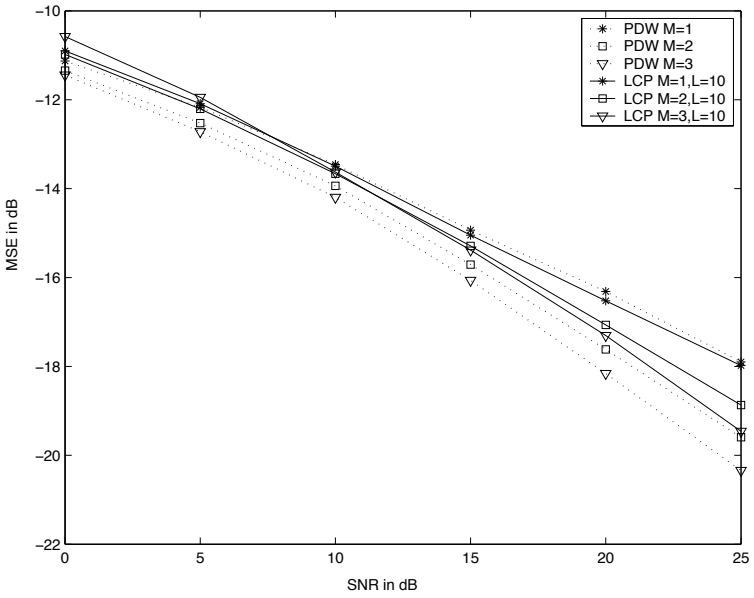

Figure 3.4: Effect of M on PDW and LCP for $f_d = 0.01$.

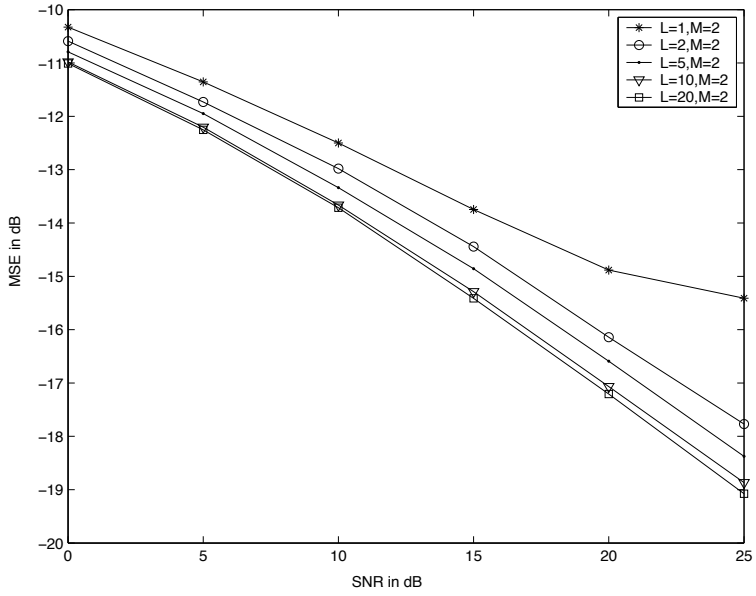

Figure 3.5: Effect of L on LCP with $M = 2$ for $f_d = 0.01$.

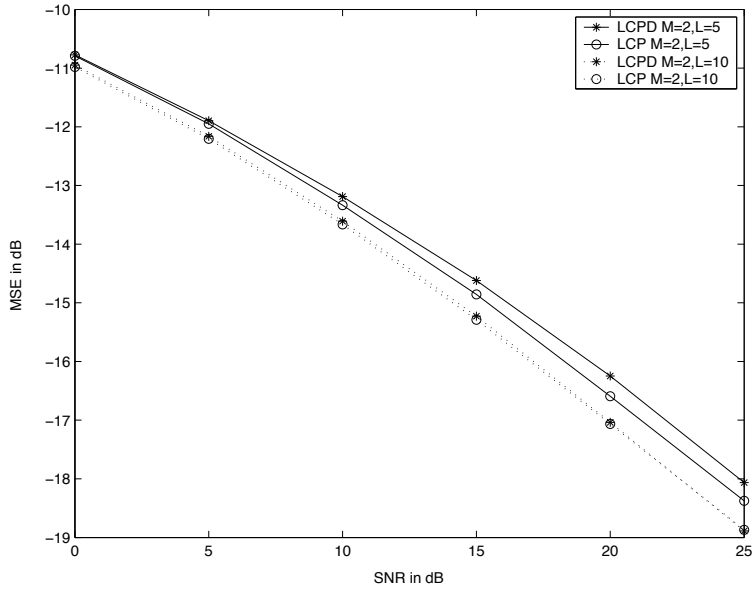

Figure 3.6: Performance of LCPD with $M = 2$ for $f_d = 0.01$.

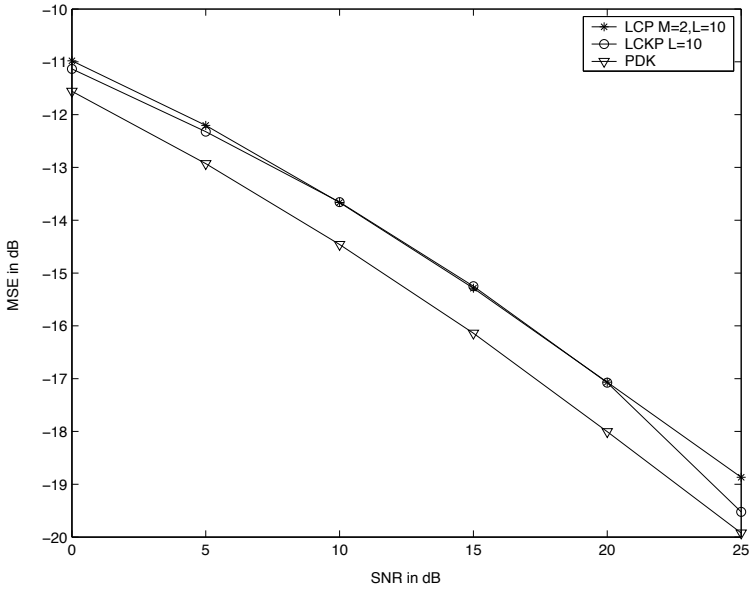

Figure 3.7: Comparison of LCP, LCKP and PDK for $f_d = 0.01$.

CHAPTER 4

CHANNEL TRACKING FOR FREQUENCY MULTIPLEXED PILOTS SYSTEM

In this section, we consider the channel tracking problem for a multi-carrier system with pilot tones. The multi-carrier systems suffer from inter-carrier interference (ICI) in doubly-selective channels and this poses challenges in estimating the channel using pilot tones.

We consider a cyclic prefix OFDM (CP-OFDM) system (Fig. 4.1), since it is one of the most common multi-carrier systems. The channel estimation techniques presented are applicable to zero-prefix OFDM (ZP-OFDM) systems as well but with slight modifications.

4.1 System Model

Since we consider a system different from that of Chapter 2 and Chapter 3, we reintroduce the notation applicable to the new system, which will be used in the rest of this chapter.

We consider a cyclic-prefix OFDM system with the transmission pattern shown in Fig. 4.1. Each frame consists of a cyclic prefix portion followed by a OFDM symbol portion. The length of OFDM symbol is N, while that of cyclic-prefix portion is N_c

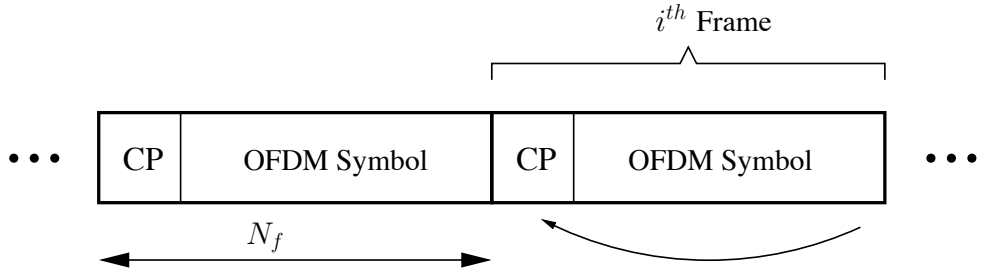

Figure 4.1: Transmission Pattern of FMP.

and N_c is chosen to be $N_h - 1$, where N_h denotes the delay spread of the channel. Let $N_f = N + N_c$ denote the frame interval. Let $\mathcal{S}^{(i)} = \{s_k^{(i)}\}_{k=0}^{k=N-1}$ denote the set of frequency domain symbols of the i^{th} OFDM symbol. N_p out of N frequency domain symbols are reserved for pilots and the rest are data symbols. The transmitted signal $t_n^{(i)}$ in the i^{th} frame is given by

$$t_{n+N_c}^{(i)} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k^{(i)} e^{j\frac{2\pi}{N}kn}, \;\; n = -N_c, ..., N-1 \tag{4.1}$$

The complete set of samples transmitted during the i^{th} frame is denoted by $\mathcal{T}^{(i)} = \{t_n^{(i)}\}_{n=0}^{N_f-1}$, 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

$$y_n = \sum_{d=0}^{N_h-1} h_{n,d} t_{n-d} + v_n \;\; \text{for} \;\; n \in \mathbb{Z}, \tag{4.2}$$

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

$$y_n^{(i)} = \sum_{d=0}^{N_h-1} h_{n,d}^{(i)} t_{n-d}^{(i)} + v_n^{(i)} \;\; \text{for} \;\; 0 \le n \le N_f - 1, \tag{4.3}$$

with $t_{-n}^{(i)} = t_{N_f-n}^{(i-1)}$ for $0 < n \leq N_h - 1$.

The following notation will be useful in the sequel. $\mathcal{Y}^{(i)} := \{y_n^{(i)}\}_{n=0}^{n=N_f-1}$ and $\mathcal{H}^{(i)} := \{h_{n,d}^{(i)} \forall d\}_{n=0}^{N_f-1}$ will denote the set of received samples and the set of channel tap coefficients of the i^{th} frame respectively.

Taking the point of view that $\mathcal{H}^{(i)}$ is useful for detection of the unknown data in $\mathcal{T}^{(i)}$, our goal is estimation of $\mathcal{H}^{(i)}$ using current and past observations $\{\mathcal{Y}^{(i-k)}\}_{k\geq 0}$, and past transmission frames $\{\mathcal{T}^{(i-k)}\}_{k\geq 1}$. The pilot symbols are known *a priori* by the receiver, and, under the assumption of perfect decoding, past transmission frames are known as well.

4.1.1 Matrix/Vector Notation

For convenience, we introduce vector/matrix notation for our system which will be used in the subsequent sections. Let $\boldsymbol{s}_{\mathsf{p}}{}^{(i)}$ denote $N \times 1$ vector of pilot symbols in the i^{th} OFDM block with zeros put in the position of data symbols. The pilot symbols are kept the same in all the OFDM blocks and we define $\boldsymbol{s}_{\mathsf{p}} := \boldsymbol{s}_{\mathsf{p}}^{(i)}$. Similarly, we form the vector $\boldsymbol{s}_{\mathsf{d}}^{(i)}$ of data symbols with zeros inserted in the pilot positions. We define $\boldsymbol{s}^{(i)} := \boldsymbol{s}_{\mathsf{p}} + \boldsymbol{s}_{\mathsf{d}}^{(i)}$. Let $\boldsymbol{F}$ denote N-point energy-preserving IDFT matrix. We define $\boldsymbol{t}_{\mathsf{p}} := \boldsymbol{F}\boldsymbol{s}_{\mathsf{p}}$ and $\boldsymbol{t}_{\mathsf{d}}^{(i)} := \boldsymbol{F}\boldsymbol{s}_{\mathsf{d}}^{(i)}$. Forming the vector $\boldsymbol{t}^{(i)} = [t_0^{(i)}, ..., t_{N_f-1}^{(i)}]^t$, we have

$$\boldsymbol{t}^{(i)} = \underbrace{\begin{bmatrix}\boldsymbol{A}_0 \\ \boldsymbol{I}_N\end{bmatrix}}_{\boldsymbol{A}} \underbrace{(\boldsymbol{t}_{\mathsf{p}} + \boldsymbol{t}_{\mathsf{d}}^{(i)})}_{\boldsymbol{t}_{\mathsf{s}}^{(i)}}, \tag{4.4}$$

where $\boldsymbol{A}_0 = \begin{bmatrix}\boldsymbol{0}_{N_c \times N - N_c} \boldsymbol{I}_{N_c}\end{bmatrix}$. $\boldsymbol{t}_{\mathsf{cp}}^{(i)}$ - the cyclic-prefix portion of i^{th} frame is given by

$$\boldsymbol{t}_{\mathsf{cp}}^{(i)} = \boldsymbol{A}_0 \boldsymbol{t}_{\mathsf{s}}^{(i)}. \tag{4.5}$$

The received sample vector $\boldsymbol{y}^{(i)} = [y_0^{(i)}, ..., y_{N_f-1}^{(i)}]^t$ can be written as

$$\boldsymbol{y}^{(i)} = \boldsymbol{H}^{(i)} \underbrace{\begin{bmatrix} \boldsymbol{t}_{\mathsf{cp}}^{(i-1)} \\ \boldsymbol{t}^{(i)} \end{bmatrix}}_{\underline{\boldsymbol{t}}^{(i)}} + \boldsymbol{v}^{(i)} \tag{4.6}$$

where

$$\boldsymbol{H}^{(i)} = \begin{bmatrix} h_{0,N_h-1}^{(i)} & h_{0,N_h-2}^{(i)} & \cdots & h_{0,0}^{(i)} & & \\ & h_{1,N_h-1}^{(i)} & h_{1,N_h-2}^{(i)} & \cdots & h_{1,0}^{(i)} & \\ & \ddots & \ddots & \cdots & \ddots & \\ & & h_{N_f-1,N_h-1}^{(i)} & h_{N_f-1,N_h-2}^{(i)} & \cdots & h_{N_f-1,0}^{(i)} \end{bmatrix} \tag{4.7}$$

and $\boldsymbol{v}^{(i)} = [v_0^{(i)}, ..., v_{N_f-1}^{(i)}]$. The convolution in (4.6) can be rearranged as

$$\boldsymbol{y}^{(i)} = \boldsymbol{T}_{\mathsf{s}}^{(i)} \boldsymbol{h}^{(i)} + \boldsymbol{v}^{(i)} \tag{4.8}$$

where

$$\boldsymbol{T}_{\mathsf{s}}^{(i)} = \begin{bmatrix} t_0^{(i)} & \cdots & t_{-N_c}^{(i)} & & & & & & \\ & & & t_1^{(i)} & \cdots & t_{-(N_c-1)}^{(i)} & & & \\ & & & & & & \ddots & & \\ & & & & & & & t_{N_f-1}^{(i)} & \cdots & t_{N_f-N_h}^{(i)} \end{bmatrix} \tag{4.9}$$

and $\boldsymbol{h}^{(i)}$ is $N_f N_h \times 1$ vector, described element-wise as $[\boldsymbol{h}^{(i)}]_l = h^{(i)}_{\lfloor \frac{l}{N_h} \rfloor, \langle l \rangle_{N_h}}$. Using (4.4) we divide $\boldsymbol{T}_{\mathsf{s}}^{(i)}$ into the pilot and data components as $\boldsymbol{T}_{\mathsf{s}}^{(i)} = \boldsymbol{T}_{\mathsf{p}} + \boldsymbol{T}_{\mathsf{d}}^{(i)}$ and similarly $\underline{\boldsymbol{t}}^{(i)} = \underline{\boldsymbol{t}}_{\mathsf{p}} + \underline{\boldsymbol{t}}_{\mathsf{d}}^{(i)}$.

4.2 Pilot-aided Wiener Estimation

Reasonably accurate estimates of time-invariant channels for OFDM systems can be obtained with N_h pilot tones. But in doubly-selective channels, the scenario is different because of the ICI introduced. We give a pilot- based Wiener estimation technique(PW) which takes complete account of these ICI effects.

We use the received samples of M previous frames and also the current frame to estimate the channel during the current frame. We form the observation vector, $\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} = [\boldsymbol{y}^{(i-M)t}, ..., \boldsymbol{y}^{(i)t}]^t$ and we have

$$\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} = \boldsymbol{H}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{m}}^{(i)} \tag{4.10}$$

where

$$\boldsymbol{H}_{\mathsf{m}}^{(i)} = \begin{bmatrix} h_{0,N_h-1}^{(i-M)} & \cdots & h_{0,0}^{(i-M)} & & & & & & \\ & \ddots & & \ddots & & & & & \\ & & h_{N_f-1,N_h-1}^{(i-M)} & \cdots & h_{N_f-1,0}^{(i-M)} & & & & \\ & & & h_{0,N_h-1}^{(i-M+1)} & \cdots & h_{0,0}^{(i-M+1)} & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & h_{N_f-1,N_h-1}^{(i)} & \cdots & h_{N_f-1,0}^{(i)} \end{bmatrix}$$

$\underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} = [\boldsymbol{t}_{\mathsf{cp}}^{(i-M-1)t}, \boldsymbol{t}^{(i-M)t}, \boldsymbol{t}^{(i-M+1)t}, ..., \boldsymbol{t}^{(i)t}]^t$ and $\underline{\boldsymbol{v}}_{\mathsf{m}}^{(i)} = [\boldsymbol{v}^{(i-M)t}, ..., \boldsymbol{v}^{(i)t}]^t$. We assume that the data vector $\boldsymbol{t}_{\mathsf{d}}^{(i)}$ is zero-mean and uncorrelated with the data vector of other OFDM symbols, i.e., $E\{\boldsymbol{t}_{\mathsf{d}}^{(i)} \boldsymbol{t}_{\mathsf{d}}^{(j)H}\} = \boldsymbol{R}_{\mathsf{d}} \delta(i-j)$. Now, we have

$$\begin{aligned} \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{m}}, \underline{\boldsymbol{y}}_{\mathsf{m}}} &= E\{\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)H}\} \\ &= E\{\boldsymbol{H}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)H} \boldsymbol{H}_{\mathsf{m}}^{(i)H}\} + \sigma_v^2 \boldsymbol{I} \\ &= E\{E\{\boldsymbol{H}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)H} \boldsymbol{H}_{\mathsf{m}}^{(i)H} | \boldsymbol{H}_{\mathsf{m}}^{(i)}\}\} + \sigma_v^2 \boldsymbol{I} \\ &= E\{\boldsymbol{H}_{\mathsf{m}}^{(i)} \underbrace{E\{\underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)H}\}}_{\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}}, \underline{\boldsymbol{t}}_{\mathsf{m}}}} \boldsymbol{H}_{\mathsf{m}}^{(i)H}\} + \sigma_v^2 \boldsymbol{I} \qquad (4.11) \\ &= \sigma_v^2 \boldsymbol{I} + \sum_{k=0}^{N_h-1} \boldsymbol{R}_{\mathsf{m}} \odot \boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}}, \underline{\boldsymbol{t}}_{\mathsf{m}}}^{(k)} \qquad (4.12) \end{aligned}$$

where $\odot$ denotes point-wise multiplication, $\boldsymbol{R}_{\mathsf{m}}$ is a $(M+1)N_f \times (M+1)N_f$ Toeplitz matrix with its first column and row being $[\boldsymbol{R}_{\mathsf{m}}]_{l,0} = [\boldsymbol{R}_{\mathsf{m}}]_{0,l} = N_h^{-1} J_0(2\pi f_d l)$ and $\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}}, \underline{\boldsymbol{t}}_{\mathsf{m}}}^{(k)} = [\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}}, \underline{\boldsymbol{t}}_{\mathsf{m}}}]_{k:(M+1)N_f+k-1, k:(M+1)N_f+k-1}$. (4.12) follows from (4.11) since different

channel taps are uncorrelated but have the same fading statistics. Now,

$$\underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} = \underbrace{\begin{bmatrix} \boldsymbol{A}_0 & & \\ & \boldsymbol{A} & \\ & & \ddots \\ & & & \boldsymbol{A} \end{bmatrix}}_{\boldsymbol{A}_{\mathsf{m}}} \Big(\underbrace{\begin{bmatrix} \boldsymbol{t}_{\mathsf{p}} \\ \boldsymbol{t}_{\mathsf{p}} \\ \vdots \\ \boldsymbol{t}_{\mathsf{p}} \end{bmatrix}}_{\underline{\boldsymbol{t}}_{\mathsf{pm}}} + \underbrace{\begin{bmatrix} \boldsymbol{t}_{\mathsf{d}}^{(i-M-1)} \\ \boldsymbol{t}_{\mathsf{d}}^{(i-M)} \\ \vdots \\ \boldsymbol{t}_{\mathsf{d}}^{(i)} \end{bmatrix}}_{\underline{\boldsymbol{t}}_{\mathsf{dm}}} \Big) \tag{4.13}$$

and hence, we have

$$\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}},\underline{\boldsymbol{t}}_{\mathsf{m}}} = E\{\underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)} \underline{\boldsymbol{t}}_{\mathsf{m}}^{(i)H}\} \tag{4.14}$$

$$= \boldsymbol{A}_{\mathsf{m}} \{\underline{\boldsymbol{t}}_{\mathsf{pm}} \underline{\boldsymbol{t}}_{\mathsf{pm}}^{H} + \text{blkdiag}(\underbrace{\boldsymbol{R}_{\mathsf{d}}, ..., \boldsymbol{R}_{\mathsf{d}}}_{M+2 \text{ times}})\} \boldsymbol{A}_{\mathsf{m}}^{H}. \tag{4.15}$$

Now, we rewrite (4.10) as

$$\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} = \text{blkdiag}(\underbrace{\boldsymbol{T}_{\mathsf{p}}, ..., \boldsymbol{T}_{\mathsf{p}}}_{M+1 \text{ times}}) \underline{\boldsymbol{h}}^{(i)} + \text{blkdiag}(\boldsymbol{T}_{\mathsf{d}}^{(i-M)}, ..., \boldsymbol{T}_{\mathsf{d}}^{(i)}) \underline{\boldsymbol{h}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{m}}^{(i)} \tag{4.16}$$

where $\underline{\boldsymbol{h}}^{(i)} = [\boldsymbol{h}^{(i-M)t}, ..., \boldsymbol{h}^{(i)t}]^t$. We have,

$$\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}} = E\{\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} \boldsymbol{h}^{(i)H}\} \tag{4.17}$$

$$= \text{blkdiag}(\underbrace{\boldsymbol{T}_{\mathsf{p}}, ..., \boldsymbol{T}_{\mathsf{p}}}_{M+1 \text{ times}}) \, E\{\underline{\boldsymbol{h}}^{(i)} \boldsymbol{h}^{(i)H}\} \tag{4.18}$$

and the LMMSE estimate of $\boldsymbol{h}^{(i)}$ using $\underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)}$ is given by

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{pilot}} = \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}}^{H} \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\underline{\boldsymbol{y}}_{\mathsf{m}}}^{-1} \underline{\boldsymbol{y}}_{\mathsf{m}}^{(i)} \tag{4.19}$$

4.2.1 Pilot Tone Selection

In case of LTI channels, the pilot tones are uniformly spaced across the band [20]. In the doubly-selective channels, OFDM systems suffer from ICI [8]. In flat fading time-selective channels, it has been shown in [7], all the pilot tones have to be grouped together to minimize the ICI effects. So, based on these results, it has been suggested

in [7], that, for doubly-selective channels, the pilot tones can be grouped as small clusters and these clusters are spaced uniformly across the band. We address the issues of cluster size N_{cs} selection in this section.

We define $\boldsymbol{H}_{\mathsf{df}}^{(i)} := \boldsymbol{F}^H \boldsymbol{H}_{\mathsf{d}}^{(i)} \boldsymbol{F}$ where $[\boldsymbol{H}_{\mathsf{d}}^{(i)}]_{n,l} = h^{(i)}_{n+N_h-1,\langle l-n\rangle_N}$, $n, l \in \{0, ..., N-1\}$. With $\boldsymbol{y}_{\mathsf{d}}^{(i)} = [y^{(i)}_{N_h-1}, ..., y^{(i)}_{N_f-1}]^t$ and $\boldsymbol{v}_{\mathsf{d}}^{(i)} = [v^{(i)}_{N_h-1}, ..., v^{(i)}_{N_f-1}]^t$, we have

$$\boldsymbol{F}^H \boldsymbol{y}_{\mathsf{d}}^{(i)} = \boldsymbol{H}_{\mathsf{df}}^{(i)} \boldsymbol{s}^{(i)} + \boldsymbol{F}^H \boldsymbol{v}_{\mathsf{d}}^{(i)} \tag{4.20}$$

The above equation relates the DFT of the received samples and the frequency domain transmitted symbols. Note that the elements of $\boldsymbol{H}_{\mathsf{df}}^{(i)}$ are obtained from a linear transformation of $\boldsymbol{h}^{(i)}$ and hence the MMSE estimate of the elements of $\boldsymbol{H}_{\mathsf{df}}^{(i)}$ is obtained by the same linear transformation of the MMSE estimate of $\boldsymbol{h}^{(i)}$. Choosing the pseudo-random pilot symbols in the frequency domain, we calculate the MSE for the estimate for the elements of pilot and data 'columns' of $\boldsymbol{H}_{\mathsf{df}}^{(i)}$. Note, k^{th} column of $\boldsymbol{H}_{\mathsf{df}}^{(i)}$ is decided as a pilot or data column depending on whether $s_k^{(i)}$ is a pilot or data symbol.

If we increase the cluster size, the MSE of the pilot columns decreases because of less interference from the data. This is illustrated in Fig. 4.2. If we increase the cluster size, the MSE of the data columns decreases up to a certain value of cluster size and increases afterwards (Fig. 4.3). This is because, as the cluster size increases, the maximum spacing between the pilot subcarrier and the data subcarrier increases, which causes the correlation to decrease and hence the MSE performance degrades. For LTI channel with N_h taps, reliable channel estimates can be obtained with N_h equi-spaced pilot tones [20]. From the plot Fig. 4.3, we note that, for doubly-selective channels, the pilot pattern with N_h equi-spaced clusters performs well. Hence the

cluster size will be

$$N_{cs} \approx \frac{N_p}{N_h}. \tag{4.21}$$

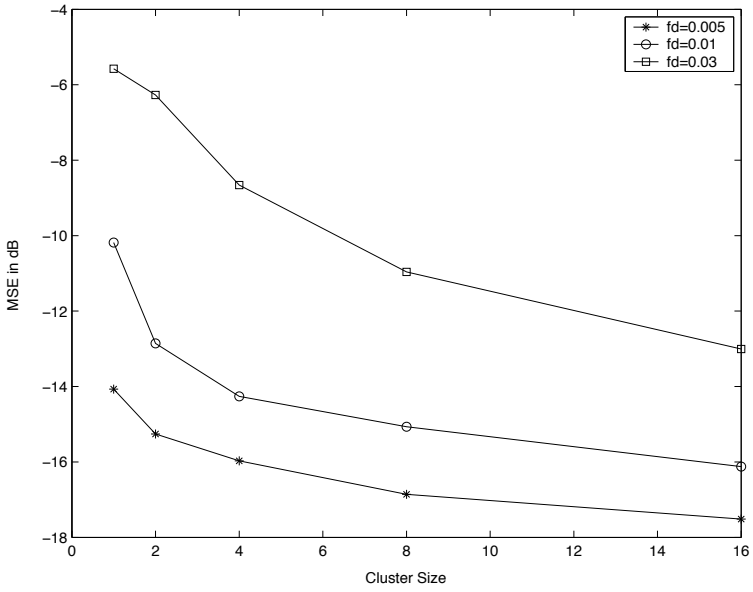

Figure 4.2: MSE of pilot 'columns', $N = 32$, $N_h = 4$, $N_p = 16$

The MSE per tap for the estimator in (4.19) is given by

$$\sigma_e^2 = \frac{1}{N_f N_h} \operatorname{tr}(\boldsymbol{R}_{\mathsf{h}} \otimes \boldsymbol{I}_{N_h} - \boldsymbol{R}^H_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}} \boldsymbol{R}^{-1}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\underline{\boldsymbol{y}}_{\mathsf{m}}} \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}}) \tag{4.22}$$

where $\boldsymbol{R}_{\mathsf{h}}$ is a $N_f \times N_f$ Toeplitz matrix with its first column and row being $[\boldsymbol{R}_{\mathsf{h}}]_{l,0} = [\boldsymbol{R}_{\mathsf{h}}]_{0,l} = N_h^{-1} J_0(2\pi f_d l)$.

For a system with $N = 64$, $N_p = 16$, and pilot symbols selected from a pseudo-random sequence, the value of σ_e^2 in dB is plotted for different channel parameters and N_{cs} values in Fig. 4.4 and Fig. 4.5.

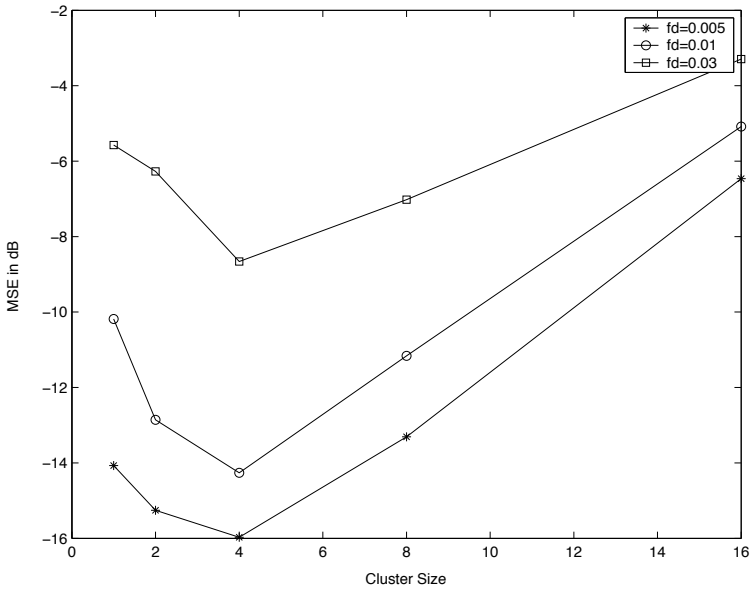

Figure 4.3: MSE of data 'columns', $N = 32$, $N_h = 4$, $N_p = 16$

In Fig. 4.4, we experiment with different values of N_h at $f_d = 0.01$. For each delay spread, the cluster size given by (4.21) is found to have minimum MSE.

In Fig. 4.5, we experiment with different Doppler frequencies. We note that for high Doppler frequencies, the cluster size given by (4.21) is better than others. But, for very low Doppler frequencies, the cluster size of 1 seems to be better than others. This is because, for very low Doppler frequencies, ICI effects are very little and hence there is no advantage in grouping the pilot tones. Even in this case, the difference in performance between cluster size of 1 and that given by (4.21) is very little.

So, the given pilot tone selection procedure provides an extension for doubly-selective channels and the theoretical MSE plots show its efficacy for wide range of Doppler frequencies.

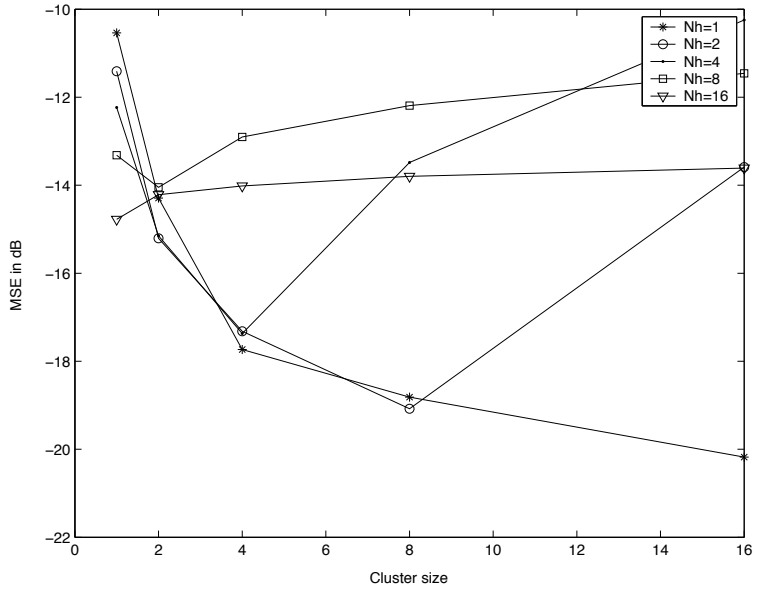

Figure 4.4: Effect of N_h at $f_d = 0.01$, $SNR = 15dB$, $M = 0$.

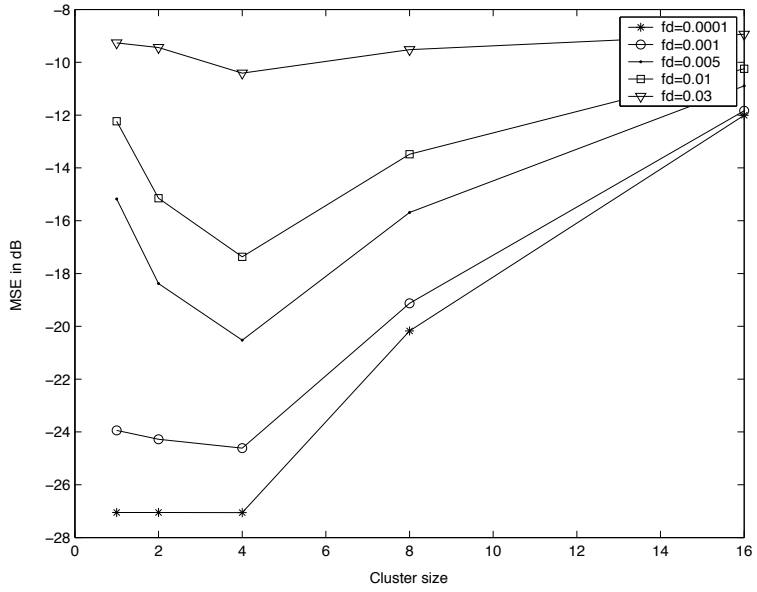

Figure 4.5: Effect of f_d at $N_h = 4$, $SNR = 15dB$, $M = 0$.

Though, PW is optimal LMMSE estimator, its performance is dominated by ICI at high Doppler rates.

4.3 Pilot-aided Decision-directed Wiener Estimation

In this pilot-aided decision-directed Wiener estimation(PDW), we use previously decoded data to improve the MSE performance. The details are given below:

Forming the same observation vector in (4.10), with the assumption that the previous frames are decoded properly, we compute the auto-correlation matrix $\boldsymbol{R}^{(i)}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\underline{\boldsymbol{y}}_{\mathsf{m}}}$ as in (4.12), with $\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{m}},\underline{\boldsymbol{t}}_{\mathsf{m}}}$ replaced by

$$\boldsymbol{R}^{(i)}_{\underline{\boldsymbol{t}}_{\mathsf{m}},\underline{\boldsymbol{t}}_{\mathsf{m}}} = \boldsymbol{A}_{\mathsf{m}}\{\underline{\boldsymbol{t}}_{\mathsf{w}}\underline{\boldsymbol{t}}^{H}_{\mathsf{w}} + \text{blkdiag}(\underbrace{\boldsymbol{0}_{N_f\times N_f},...,\boldsymbol{0}_{N_f\times N_f}}_{M+1 \text{ times}},\boldsymbol{R}_{\mathsf{d}})\}\boldsymbol{A}^{H}_{\mathsf{m}} \tag{4.23}$$

where $\underline{\boldsymbol{t}}_{\mathsf{w}} = [\boldsymbol{t}^{(i-M-1)t}_{\mathsf{s}},...,\boldsymbol{t}^{(i-1)t}_{\mathsf{s}},\boldsymbol{t}^{t}_{\mathsf{p}}]^{t}$. Similarly, we have

$$\boldsymbol{R}^{(i)}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}} = E\{\underline{\boldsymbol{y}}^{(i)}_{\mathsf{m}}\boldsymbol{h}^{(i)H}\} \tag{4.24}$$

$$= \text{blkdiag}(\boldsymbol{T}^{(i-M)}_{\mathsf{s}},...,\boldsymbol{T}^{(i-1)}_{\mathsf{s}},\boldsymbol{T}_{\mathsf{p}})E\{\underline{\boldsymbol{h}}^{(i)}\boldsymbol{h}^{(i)H}\} \tag{4.25}$$

and the LMMSE estimate of $\boldsymbol{h}^{(i)}$ using $\underline{\boldsymbol{y}}^{(i)}_{\mathsf{m}}$ is given by

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{wiener}} = \boldsymbol{R}^{(i)H}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\boldsymbol{h}}(\boldsymbol{R}^{(i)}_{\underline{\boldsymbol{y}}_{\mathsf{m}},\underline{\boldsymbol{y}}_{\mathsf{m}}})^{-1}\underline{\boldsymbol{y}}^{(i)}_{\mathsf{m}}. \tag{4.26}$$

We note that the PDW requires a matrix inversion of size $(M+1)N_f \times (M+1)N_f$ per frame (4.26). When we increase M, PDW uses more observations for estimation and hence MSE performance improves but the matrix inversion complexity also increases.

4.4 Pilot-aided Decision-directed Kalman Estimation

In this section, we find the pilot-aided decision-directed Kalman estimator (PDK) by formulating our channel estimation as a Kalman prediction problem [16]. We

assign $\boldsymbol{h}^{(i-1)}$ as the current *state* of the channel, $\boldsymbol{h}^{(i)}$ as the next state, and $\boldsymbol{y}^{(i-1)}$ as the current observation. The state dynamics can be written as

$$\boldsymbol{h}^{(i)} = \boldsymbol{A}_{\mathsf{k}}\boldsymbol{h}^{(i-1)} + \boldsymbol{D}_{\mathsf{k}}\boldsymbol{w}_{\mathsf{k}}^{(i-1)} \tag{4.27}$$

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. Let $\hat{\boldsymbol{h}}_{\mathsf{k}}^{(i)}$ denote the LMMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\boldsymbol{y}^{(i-1)}, ..., \boldsymbol{y}^{(0)}\}$ with the knowledge of previous data and $\tilde{\boldsymbol{h}}_{\mathsf{k}}^{(i)}$ denote the LMMSE estimate of $\boldsymbol{h}^{(i)}$ using the observations $\{\boldsymbol{y}^{(i)}, ..., \boldsymbol{y}^{(0)}\}$ with the knowledge of previous data and the current pilot. From the Kalman filter theory [16], we have

$$\hat{\boldsymbol{h}}_{\mathsf{k}}^{(i)} = \boldsymbol{A}_{\mathsf{k}}\hat{\boldsymbol{h}}_{\mathsf{k}}^{(i-1)} + \boldsymbol{L}_{\mathsf{k}}^{(i-1)}\{\boldsymbol{y}^{(i-1)} - \boldsymbol{T}_{\mathsf{s}}^{(i-1)}\hat{\boldsymbol{h}}_{\mathsf{k}}^{(i-1)}\} \tag{4.28}$$

with $\hat{\boldsymbol{h}}_{\mathsf{k}}^{(0)} = \boldsymbol{0}$ and the predictor gain is given by

$$\boldsymbol{L}_{\mathsf{k}}^{(i-1)} = \boldsymbol{A}_{\mathsf{k}}^{(i-1)}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{T}_{\mathsf{s}}^{(i-1)H}(\boldsymbol{T}_{\mathsf{s}}^{(i-1)}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{T}_{\mathsf{s}}^{(i-1)H} + \sigma_v^2\boldsymbol{I})^{-1} \tag{4.29}$$

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

$$\boldsymbol{P}_{\mathsf{k}}^{(i)} = \boldsymbol{A}_{\mathsf{k}}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{A}_{\mathsf{k}}^H - \boldsymbol{L}_{\mathsf{k}}^{(i-1)}\boldsymbol{T}_{\mathsf{s}}^{(i-1)}\boldsymbol{P}_{\mathsf{k}}^{(i-1)}\boldsymbol{A}_{\mathsf{k}}^H + \sigma_{w_{\mathsf{k}}}^2\boldsymbol{D}_{\mathsf{k}}\boldsymbol{D}_{\mathsf{k}}^H \tag{4.30}$$

with $\boldsymbol{P}_{\mathsf{k}}^{(0)} = \boldsymbol{R}_{\mathsf{h}} \otimes \boldsymbol{I}_{N_h}$. Now, $\tilde{\boldsymbol{h}}_{\mathsf{k}}^{(i)}$ is given as

$$\tilde{\boldsymbol{h}}_{\mathsf{k}}^{(i)} = \hat{\boldsymbol{h}}_{\mathsf{k}}^{(i)} + \boldsymbol{M}_{\mathsf{k}}^{(i)}\{\boldsymbol{y}^{(i)} - \boldsymbol{T}_{\mathsf{p}}\hat{\boldsymbol{h}}_{\mathsf{k}}^{(i)}\} \tag{4.31}$$

and

$$\boldsymbol{M}_{\mathsf{k}}^{(i)} = \boldsymbol{P}_{\mathsf{k}}^{(i)}\boldsymbol{T}_{\mathsf{p}}^H(\boldsymbol{T}_{\mathsf{p}}\boldsymbol{P}_{\mathsf{k}}^{(i)}\boldsymbol{T}_{\mathsf{p}}^H + \boldsymbol{R}_{\mathsf{yd}} + \sigma_v^2\boldsymbol{I})^{-1} \tag{4.32}$$

where

$$\boldsymbol{R}_{\mathsf{yd}} = E\{\boldsymbol{H}^{(i)}\underline{\boldsymbol{t}}_{\mathsf{d}}^{(i)}\underline{\boldsymbol{t}}_{\mathsf{d}}^{(i)H}\boldsymbol{H}^{(i)H}\} \tag{4.33}$$

$$= \sum_{k=0}^{N_h-1} \boldsymbol{R}_{\mathsf{h}} \odot \boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{d}},\underline{\boldsymbol{t}}_{\mathsf{d}}}(k:k+N_f-1, k:k+N_f-1) \tag{4.34}$$

and

$$\boldsymbol{R}_{\underline{\boldsymbol{t}}_{\mathsf{d}},\underline{\boldsymbol{t}}_{\mathsf{d}}} = E\{\underline{\boldsymbol{t}}_{\mathsf{d}}^{(i)}\underline{\boldsymbol{t}}_{\mathsf{d}}^{(i)H}\} \tag{4.35}$$

$$= \begin{bmatrix} \boldsymbol{A}_0 & \\ & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{\mathsf{d}} & \\ & \boldsymbol{R}_{\mathsf{d}} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \\ & \boldsymbol{A} \end{bmatrix}^H. \tag{4.36}$$

Note that the Kalman estimator uses *all* previous observations in its prediction of $\boldsymbol{h}^{(i)}$; this is the advantage of the PDK over the PDW. However, the performance of the PDK depends on how well the model (4.27) describes the true evolution of the state process. These mismatch related issues are already discussed in Chapter 2. Note also, from (4.29) and (4.32), that the PDK requires matrix inversions of size $N_f \times N_f$ for each frame.

4.5 Low Complexity Prediction

In this section, we derive a low-complexity predictor for CP-OFDM system similar to the one discussed in Chapter 3, which has a smoothing stage followed by a prediction stage. The details are given below:

4.5.1 Kalman Smoothing Stage

The smoothed channel estimates of the $(i-1)^{th}$ frame are obtained using Kalman filtering. We divide each frame into sub-frames of size L and design the Kalman smoother such that it takes the received samples of each sub-frame and finds the smoothed estimates of the channel of that sub-frame. 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 set of k^{th} sub-frame's channel coefficients and observations 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

$$\boldsymbol{h}^{(i-1,k+1)} \quad = \quad \boldsymbol{A}_{\mathsf{I}} \boldsymbol{h}^{(i-1,k)} + \boldsymbol{D}_{\mathsf{I}} \boldsymbol{w}_{\mathsf{I}}^{(i-1,k)} \tag{4.37}$$

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

$$\boldsymbol{y}^{(i-1,k)} \quad = \quad \boldsymbol{T}_{\mathsf{I}}^{(i-1,k)} \boldsymbol{h}^{(i-1,k)} + \boldsymbol{v}^{(i-1,k)} \tag{4.38}$$

where $\boldsymbol{T}_{\mathsf{I}}^{(i-1,k)}$ is a $L \times LN_h$ matrix with the same structure given in (3.3). For the system described by (4.37) and (4.38), $\tilde{\boldsymbol{h}}_{\mathsf{I}}^{(i-1,k)}$ denote *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 $\hat{\boldsymbol{h}}_{\mathsf{I}}^{(i-1,k)}$ denote *predicted* MMSE estimate of $\boldsymbol{h}^{(i-1,k)}$ using $\{\boldsymbol{y}^{(i-1,k-1)}, ..., \boldsymbol{y}^{(i-1,0)}, \boldsymbol{y}^{(i-2)}, ..., \boldsymbol{y}^{(0)}\}$. The smoothed estimate is given by [16]

$$\tilde{\boldsymbol{h}}_{\mathsf{I}}^{(i-1,k)} \quad = \quad \hat{\boldsymbol{h}}_{\mathsf{I}}^{(i-1,k)} + \boldsymbol{M}_{\mathsf{I}}^{(i-1,k)} \{\boldsymbol{y}^{(i-1,k)} - \boldsymbol{T}_{\mathsf{I}}^{(i-1,k)} \hat{\boldsymbol{h}}_{\mathsf{I}}^{(i-1,k)}\}$$

with smoother gain

$$\boldsymbol{M}_{\mathsf{I}}^{(i-1,k)} \quad = \quad \boldsymbol{P}_{\mathsf{I}}^{(i-1,k)} \boldsymbol{T}_{\mathsf{I}}^{(i-1,k)H} (\boldsymbol{T}_{\mathsf{I}}^{(i-1,k)} \boldsymbol{P}_{\mathsf{I}}^{(i-1,k)} \boldsymbol{T}_{\mathsf{I}}^{(i-1,k)H} + \sigma_v^2 \boldsymbol{I})^{-1} \tag{4.39}$$

The predicted estimate is given recursively as

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

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

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

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 by collecting the smoothed estimates of channel of all the K subframes as

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} = [\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,0)t}, ..., \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1,K-1)t}]^t. \tag{4.40}$$

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

$$\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)} = \boldsymbol{h}^{(i-1)} + \boldsymbol{e}_{\mathsf{l}}^{(i-1)} \tag{4.41}$$

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

4.5.2 Wiener Prediction Stage

We form the observation vector as $\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} = [\tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-M)t}, ..., \tilde{\boldsymbol{h}}_{\mathsf{l}}^{(i-1)t}, \boldsymbol{y}^{(i)t}]^t$ and

$$\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)} = \underbrace{\begin{bmatrix} \boldsymbol{I}_{MN_fN_h} & \boldsymbol{T}_{\mathsf{p}} \end{bmatrix}}_{\boldsymbol{A}_{\mathsf{l}}} \underline{\boldsymbol{h}}^{(i)} + \begin{bmatrix} \boldsymbol{0}_{MN_fN_h \times 1} \\ \boldsymbol{T}_{\mathsf{d}}^{(i)}\boldsymbol{h}^{(i)} \end{bmatrix} + \underbrace{\begin{bmatrix} \boldsymbol{e}_{\mathsf{l}}^{(i-M)} \\ \vdots \\ \boldsymbol{e}_{\mathsf{l}}^{(i-1)} \\ \boldsymbol{v}^{(i)} \end{bmatrix}}_{\underline{\boldsymbol{e}}_{\mathsf{l}}^{(i)}} \tag{4.42}$$

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.

1. The smoothing errors are uncorrelated with variance σ_v^2,

$$E\{\boldsymbol{e}_{\mathsf{l}}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \sigma_v^2\boldsymbol{I}\delta(p-q). \tag{4.43}$$

$$E\{\boldsymbol{v}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \boldsymbol{0}\,\forall\, p \neq q. \tag{4.44}$$

2. The smoothing errors are uncorrelated with the channel,

$$E\{\boldsymbol{h}^{(p)}\boldsymbol{e}_{\mathsf{l}}^{(q)H}\} = \boldsymbol{0}\,\forall\, p,\ q. \tag{4.45}$$

Now, with (4.43) - (4.45), we have

$$\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{l}},\underline{\boldsymbol{y}}_{\mathsf{l}}} = E\{\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)H}\} \tag{4.46}$$

$$= \boldsymbol{A}_{\mathsf{l}}(\boldsymbol{R}_{\mathsf{m}}\otimes\boldsymbol{I}_{N_h})\boldsymbol{A}_{\mathsf{l}}^H + \begin{bmatrix} \boldsymbol{0}_{MN_fN_h\times MN_fN_h} & \\ & \boldsymbol{R}_{\mathsf{yd}} \end{bmatrix} + \sigma_v^2\boldsymbol{I} \tag{4.47}$$

We also have

$$\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{l}},\boldsymbol{h}} = E\{\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}\boldsymbol{h}^{(i)H}\} \tag{4.48}$$

$$= \boldsymbol{A}_{\mathsf{l}}E\{\underline{\boldsymbol{h}}^{(i)}\boldsymbol{h}^{(i)H}\}. \tag{4.49}$$

The LMMSE estimate of $\boldsymbol{h}^{(i)}$ using $\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}$ is given by

$$\hat{\boldsymbol{h}}^{(i)}\big|_{\mathsf{lcp}} = \boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{l}},\boldsymbol{h}}^H\boldsymbol{R}_{\underline{\boldsymbol{y}}_{\mathsf{l}},\underline{\boldsymbol{y}}_{\mathsf{l}}}^{-1}\underline{\boldsymbol{y}}_{\mathsf{l}}^{(i)}. \tag{4.50}$$

Note that, the predictor coefficients in (4.50) are time invariant. The matrix inversion of size $L \times L$ is needed in smoothing stage of LCP (4.39) and choice of L is a tradeoff between performance and complexity.

4.6 Persistent Training and Prediction

In this section, we consider a reference scheme "persistent training and prediction" (PTP) similar to Section 3.3. We find the benchmark for the performance of our

estimation techniques by transmitting a persistent training sequence and using a estimator which incorporates infinite past observations. The training sequence is chosen to minimize the approximate expression for MSE of the MMSE estimator.

4.6.1 Training Sequence Design

We search for the MSE-minimizing training sequence for doubly-selective channel parameter estimation. The estimator is assumed to know the pilots in the current frame and both data and pilots in M past frames, for the frame structure of Fig. 4.1. For simplicity, though, we assume the frame duration is an integer multiple of the delay spread N_h, i.e., $N_f = QN_h$ for $Q \in \mathbb{Z}$, and that the channel is *block-fading*: the coefficients are fixed within N_h-sample temporal blocks but change arbitrarily between blocks. In addition, we find the sequence which minimizes a truncated series expansion of MSE rather than MSE itself.

With the block fading model, the i^{th}-frame observation becomes

$$\boldsymbol{y}^{(i)} \;=\; \boldsymbol{T}_{\mathsf{b}}^{(i)}\boldsymbol{h}_{\mathsf{b}}^{(i)} + \boldsymbol{v}^{(i)}.$$

where the block-fading coefficients $\boldsymbol{h}_{\mathsf{b}}^{(i)}$ are defined element-wise as $[\boldsymbol{h}_{\mathsf{b}}^{(i)}]_l := h^{(i)}_{N_h\lfloor l/N_h\rfloor,\langle l\rangle_{N_h}}$, and where $\boldsymbol{T}_{\mathsf{b}}^{(i)} := \text{blkdiag}\big(\boldsymbol{T}_0^{(i)},\ldots,\boldsymbol{T}_{Q-1}^{(i)}\big)$ for Toeplitz

$$\boldsymbol{T}_q^{(i)} \;:=\; \begin{bmatrix} t^{(i)}_{qN_h} & \cdots & t^{(i)}_{qN_h-N_h+1} \\ \vdots & \ddots & \vdots \\ t^{(i)}_{qN_h+N_h-1} & \cdots & t^{(i)}_{qN_h} \end{bmatrix}.$$

Since the transmitted sequence in the i^{th} frame is composed of pilots and data (4.4), we can split $\boldsymbol{T}_{\mathsf{b}}^{(i)}$ into pilot and data component as $\boldsymbol{T}_{\mathsf{b}}^{(i)} = \boldsymbol{T}_{\mathsf{bp}} + \boldsymbol{T}_{\mathsf{bd}}^{(i)}$. In the past frames, we transmit a persistent training sequence, since we are interested in minimal estimation error for each frame, we restrict our training sequence to be N_f periodic. Let $\boldsymbol{T}_{\mathsf{b}} := \boldsymbol{T}_{\mathsf{b}}^{(i-k)} \forall\, k > 0$.

The multi-frame observation $\underline{\boldsymbol{y}}^{(i)} := \left[\boldsymbol{y}^{(i)t}, \boldsymbol{y}^{(i-1)t}, \ldots, \boldsymbol{y}^{(i-M)t}\right]^t$ can be written

$$\underline{\boldsymbol{y}}^{(i)} \;=\; \underline{\boldsymbol{T}}_{\mathsf{b}} \underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} + \underline{\boldsymbol{v}}_{\mathsf{b}}^{(i)},$$

where $\underline{\boldsymbol{T}}_{\mathsf{b}} := \text{blkdiag}(\boldsymbol{T}_{\mathsf{bp}},\ \boldsymbol{I}_M \otimes \boldsymbol{T}_{\mathsf{b}})$, $\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} := \left[\boldsymbol{h}_{\mathsf{b}}^{(i)t}, \boldsymbol{h}_{\mathsf{b}}^{(i-1)t}, \ldots, \boldsymbol{h}_{\mathsf{b}}^{(i-M)t}\right]^t$ and $\underline{\boldsymbol{v}}_{\mathsf{b}}^{(i)} := \left[\boldsymbol{v}^{(i)t} + \boldsymbol{T}_{\mathsf{bd}}^{(i)} \boldsymbol{h}_{\mathsf{b}}^{(i)}, \boldsymbol{v}^{(i-1)t}, \ldots, \boldsymbol{v}^{(i-M)t}\right]^t$.

The MMSE estimate of $\boldsymbol{h}_{\mathsf{b}}^{(i)}$ using $\underline{\boldsymbol{y}}^{(i)}$ is given by

$$\hat{\boldsymbol{h}}_{\mathsf{b}}^{(i)} \;=\; \boldsymbol{P}^H \underline{\boldsymbol{T}}_{\mathsf{b}}^H \left(\underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{R}_{\mathsf{b}} \underline{\boldsymbol{T}}_{\mathsf{b}}^H + \sigma_v^2 \boldsymbol{R}_v\right)^{-1} \underline{\boldsymbol{y}}^{(i)}$$

where $\boldsymbol{R}_{\mathsf{b}} := E\{\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} \underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)H}\}$, $\boldsymbol{P} := E\{\underline{\boldsymbol{h}}_{\mathsf{b}}^{(i)} \boldsymbol{h}_{\mathsf{b}}^{(i)H}\}$ and $\boldsymbol{R}_v = \text{blkdiag}(\frac{1}{\sigma_v^2} E\{\boldsymbol{T}_{\mathsf{bd}}^{(i)} \boldsymbol{h}_{\mathsf{b}}^{(i)} \boldsymbol{h}_{\mathsf{b}}^{(i)H} \boldsymbol{T}_{\mathsf{bd}}^{(i)H}\} + \boldsymbol{I}_{N_f}, \boldsymbol{I}_{MN_f})$. Defining $\boldsymbol{R} := E\{\boldsymbol{h}_{\mathsf{b}}^{(i)} \boldsymbol{h}_{\mathsf{b}}^{(i)H}\}$, the total MSE is given by

$$\begin{aligned}
(\sigma_e^{(i)})^2 &= \text{tr}\left(E\left[\left(\boldsymbol{h}_{\mathsf{b}}^{(i)} - \hat{\boldsymbol{h}}_{\mathsf{b}}^{(i)}\right)\left(\boldsymbol{h}_{\mathsf{b}}^{(i)} - \hat{\boldsymbol{h}}_{\mathsf{b}}^{(i)}\right)^H\right]\right) \\
&= \text{tr}\left(\boldsymbol{R} - \boldsymbol{P}^H \underline{\boldsymbol{T}}_{\mathsf{b}}^H \left(\underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{R}_{\mathsf{b}} \underline{\boldsymbol{T}}_{\mathsf{b}}^H + \sigma_v^2 \boldsymbol{R}_v\right)^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}} \boldsymbol{P}\right) \\
&= \text{tr}\left(\boldsymbol{R} - \boldsymbol{P}^H \left(\boldsymbol{R}_{\mathsf{b}} + \sigma_v^2 \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} \boldsymbol{P}\right). \qquad (4.51)
\end{aligned}$$

We approximate the inverse in the above equation by a truncated power series expansion [19]:

$$\begin{aligned}
\left(\boldsymbol{R}_{\mathsf{b}} + \sigma_v^2 \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} &= \left(\boldsymbol{I} + \sigma_v^2 \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\right)^{-1} \boldsymbol{R}_{\mathsf{b}}^{-1} \\
&\approx \boldsymbol{R}_{\mathsf{b}}^{-1} - \sigma_v^2 \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{R}_{\mathsf{b}}^{-1}. \qquad (4.52)
\end{aligned}$$

Equations (4.51) and (4.52) suggest that the training sequence should be chosen to minimize, subject to unit power constraint,

$$\text{tr}\left(\boldsymbol{P}^H \boldsymbol{R}_{\mathsf{b}}^{-1} \underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{R}_{\mathsf{b}}^{-1} \boldsymbol{P}\right), \qquad (4.53)$$

i.e., $\text{tr}\left(\underline{\boldsymbol{T}}_{\mathsf{b}}^{-1} \boldsymbol{R}_v \underline{\boldsymbol{T}}_{\mathsf{b}}^{-H} \boldsymbol{Z}\right)$ for $\boldsymbol{Z} := \boldsymbol{R}_{\mathsf{b}}^{-1} \boldsymbol{P} \boldsymbol{P}^H \boldsymbol{R}_{\mathsf{b}}^{-1}$.

For $l \in \{0, ..., M\}$, let $\boldsymbol{Z}_l$ denote the l^{th} $N_f \times N_f$ sub-diagonal matrix of $\boldsymbol{Z}$, i.e., $\boldsymbol{Z}_l = [\boldsymbol{Z}]_{lN_f:lN_f+N_f-1,lN_f:lN_f+N_f-1}$. Due to the block diagonal structure of $\underline{\boldsymbol{T}}_{\mathsf{b}}$, we have

$$\begin{aligned} \mathrm{tr}\left(\underline{\boldsymbol{T}}_{\mathsf{b}}^{-1}\boldsymbol{R}_v\underline{\boldsymbol{T}}_{\mathsf{b}}^{-H}\boldsymbol{Z}\right) &= \mathrm{tr}\left(\boldsymbol{T}_{\mathsf{bp}}^{-1}([\boldsymbol{R}_v]_{0:N_f-1,0:N_f-1})\boldsymbol{T}_{\mathsf{bp}}^{-H}\boldsymbol{Z}_0\right) + \\ &\quad \sum_{l=1}^{M}\mathrm{tr}\left(\boldsymbol{T}_{\mathsf{b}}^{-1}\boldsymbol{T}_{\mathsf{b}}^{-H}\boldsymbol{Z}_l\right) \end{aligned} \tag{4.54}$$

We minimize the second term in the above equation, by 'proper' selection of $\boldsymbol{T}_{\mathsf{b}}$. Now, with the block diagonal structure of $\boldsymbol{T}_{\mathsf{b}}$ we have,

$$\sum_{l=1}^{M}\mathrm{tr}\left(\boldsymbol{T}_{\mathsf{b}}^{-1}\boldsymbol{T}_{\mathsf{b}}^{-H}\boldsymbol{Z}_l\right) \geq \sum_{k=0}^{Q-1}\frac{1}{\lambda_k}z(k) \tag{4.55}$$

where λ_k is the *maximum* eigen value of $\boldsymbol{T}_k\boldsymbol{T}_k^H$, $z(k) = \sum_{l=1}^{M}\mathrm{tr}\left(\boldsymbol{Z}_{l,k}\right)$ and $\boldsymbol{Z}_{l,k}$ is k^{th} $N_h \times N_h$ diagonal sub-matrix of $\boldsymbol{Z}_l$, i.e., $\boldsymbol{Z}_{l,k} = [\boldsymbol{Z}_l]_{kN_h:kN_h+N_h-1,kN_h:kN_h+N_h-1}$. The equality in (4.55) holds iff all the eigen values of $\boldsymbol{T}_k\boldsymbol{T}_k^H$ equal λ_k. Assuming that this is the case - (we shall give an example), we minimize RHS of (4.55) with the unit power constraint $\sum_{k=0}^{Q-1}\lambda_k = N_f$. Using the method of Lagrange optimization, we find the optimal values to be

$$\lambda_k = \frac{N_f}{\sum_{l=0}^{Q-1}\sqrt{z(l)}}\sqrt{z(k)},\ k \in \{0, \ldots, Q-1\} \tag{4.56}$$

where $z(k) \geq 0\ \forall k$ since $\boldsymbol{Z}$ is positive semi-definite. Now, consider the frame sequence

$$t_n^{(i-1)} = \sqrt{\lambda_{\lfloor\frac{n}{N_h}\rfloor}}\delta\big(\langle n\rangle_{N_h}\big) \quad n \in \{0, \ldots, N_f-1\},$$

with $\{\lambda_k\}_{k=0}^{Q-1}$ from (4.56). This sequence satisfies the criterion that all the eigen values of $\boldsymbol{T}_k\boldsymbol{T}_k^H$ are equal to the optimal values found in (4.56) and hence achieves the minimum value of the lower bound (4.55) on the truncated MSE expansion.

We find the the performance of PTP by letting M approach ∞, using an IIR Wiener estimator and transmitting the sequence found from the previous section.

This will give the benchmark for the prediction techniques considered earlier, since we are transmitting an optimal sequence in minimizing the approximate MSE and we are predicting the channel using infinite number of past observations.

4.7 Simulation Results

We consider the system with the parameters $N = 64$ and $N_h = 7$. For simulations, the MSE performance is averaged over at least 1000 channel realizations. The typical values of M and L in our simulations are 1 and 5 respectively.

In Fig. 4.6, we plot MSE versus SNR for different techniques at $f_d = 0.01$. The performance of PW is worse than LCP, PDK and PDW because of the limitation due to ICI. PDK, PDW and LCP perform nearly the same and their performance is close to that of the benchmark PTP scheme.

In Fig. 4.7, we see that the MSE increases with increase in Doppler frequency. This is because, the correlation with past frame observations decreases and the ICI power in the pilot tones of the current frame increases.

In Fig. 4.8, we show the effect of M on the MSE performance. As M increases, the estimation techniques incorporate more observations and hence the MSE performance improves but the improvements are very little.

In Fig. 4.9, we show that effect of L on LCP. Increasing L improves the performance of the Kalman smoothing stage and hence the prediction performance. But increasing the value of L increases the size of matrix inversion and hence the complexity. Also, the performance gains with the increase of L are very little.

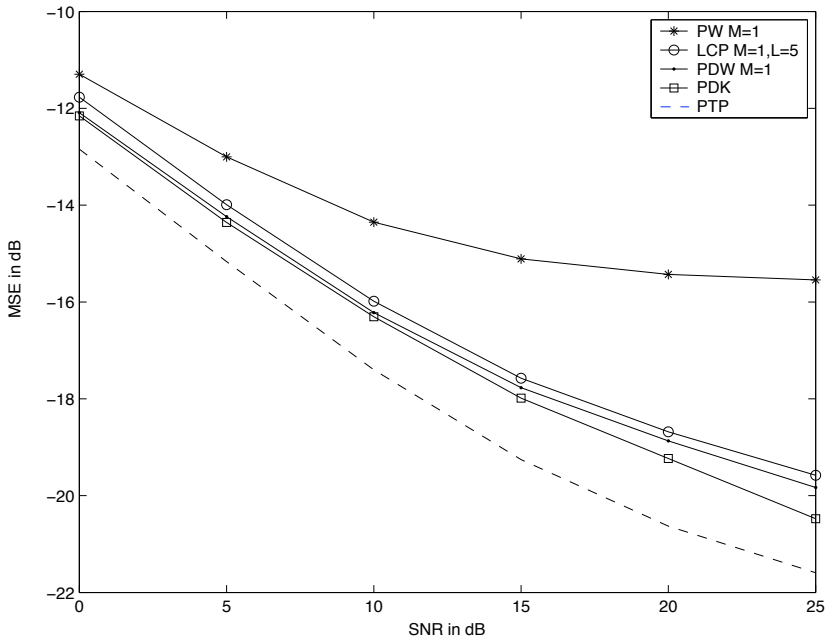

Figure 4.6: Comparison of different techniques at $f_d = 0.01$.

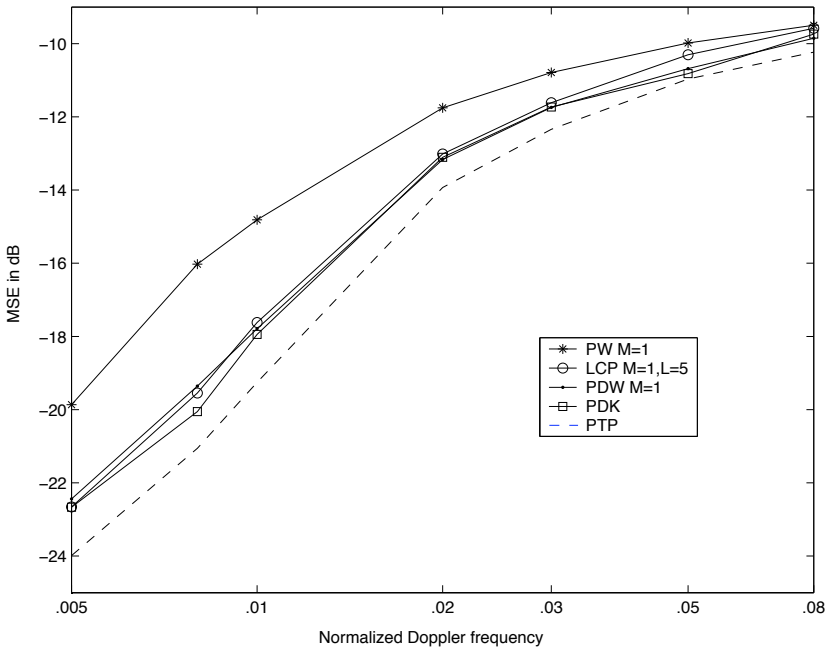

Figure 4.7: Performance for different Doppler frequencies at $SNR = 15dB$

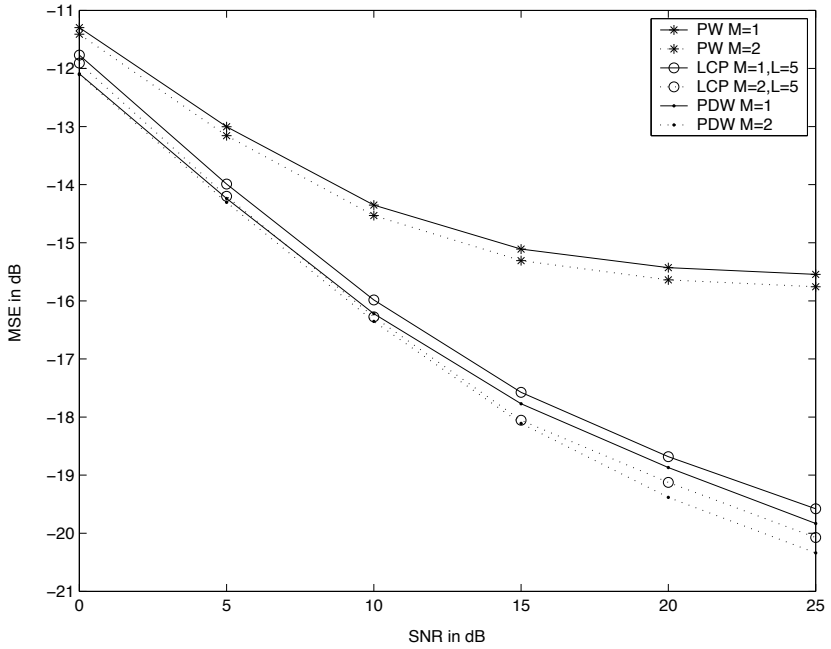

Figure 4.8: Effect of M at $f_d = 0.01$.

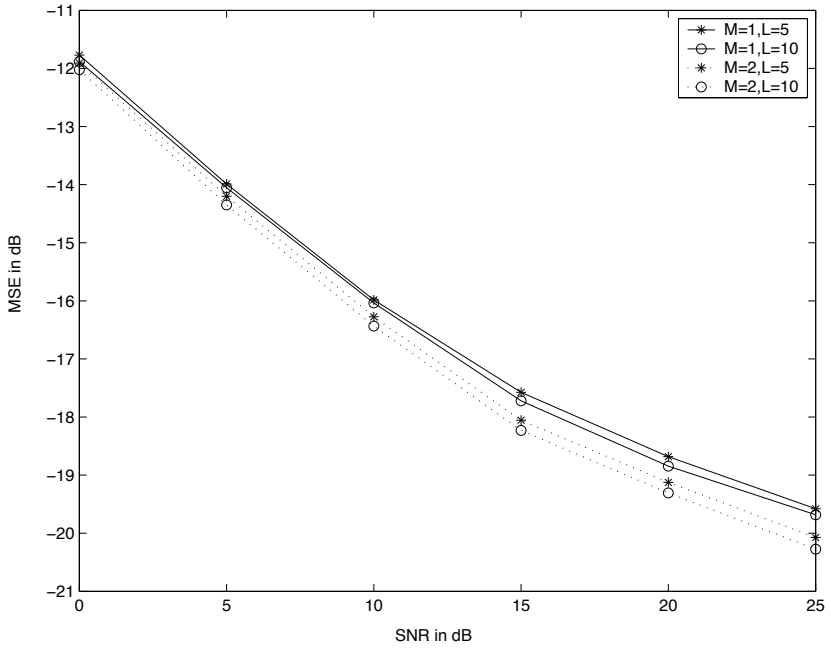

Figure 4.9: Effect of L on LCP at $f_d = 0.01$.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In this thesis, we considered channel identification techniques for a transmission scheme which consists of a sequence of frames with each frame consisting of pilots and data. We considered two systems, the first one having pilots in the time multiplexed fashion, and the second one being a multi-carrier system with frequency multiplexed pilots. We considered scenarios in which the channel is both frequency and time selective.

In Chapter 2, we reviewed the standard optimal estimation techniques. In Chapter 3, we proposed a novel computationally efficient decision-directed estimators for time multiplexed pilots system. We also found the benchmark for the performance of the estimators with the persistent training and prediction scheme. The simulation results presented show the good performance of the proposed techniques.

In Chapter 4, we considered channel tracking techniques for a multi-carrier system with pilot tones. We presented estimation techniques which take the ICI effects into

account. We also presented a pilot tone selection criterion for OFDM in doubly-selective channels and showed its efficacy by theoretical MSE calculations. We also developed a low-complexity estimation technique and presented the simulation results.

5.2 Future Work

The sampling limitation for time multiplexed pilots system and the ICI limitation of the multi-carrier system can be reduced if we could develop joint channel and data estimation techniques. The challenge would be to make these techniques less computationally expensive.

In doubly-selective channels, the time multiplexed pilots suffer from sampling limitation. The 'embedded' pilots are limited by the interference from data. We investigate the choice of the pilot sequence and design of optimal pilot pattern, given the channel fading statistics.

For OFDM systems, increasing the frame length reduces the fraction of overhead due to preamble. But increasing the frame length increases the ICI and hence causes performance degradation. We investigate to find the optimal frame length depending on the channel fading statistics.

Multi-antenna systems are popular since they significantly improve the capacity of the wireless links. We investigate the extension of channel tracking techniques and pilot design issues for MIMO systems.

APPENDIX A

OPTIMALITY RESULTS FOR KRONECKER-DELTA PILOT SEQUENCE

In this appendix, we establish some optimality properties of the Kronecker delta pilot structure considered in our TMP system. We consider the estimation of doubly-selective channel parameters over a frame interval of N time-samples. We focus on estimators which employ outputs of a channel driven by a known unit-variance sequence and corrupted by additive circular white Gaussian noise. Specifically, we are interested in determining the transmitted sequence which minimizes the MSE of channel estimates. For the results presented here, we assume a block-fading channel, where channel variation is negligible within a block of N_h samples, where N_h denotes the channel delay spread.

We now describe the system model used in this appendix. As before, we use $\{y_n\}$ to denote observations, $\{t_n\}$ the transmitted sequence, $\{v_n\}$ the additive noise, and $h_{n,l}$ the time-n channel response to an impulse applied at time $n-d$. From these we define the i^{th}-frame quantities $y_n^{(i)} = y_{iN+n}$, $t_n^{(i)} = t_{iN+n}$, $v_n^{(i)} = v_{iN+n}$, and $h_{n,l}^{(i)} = h_{iN+n,l}$. For convenience, we assume that $N = QN_h$ for some $Q \in \mathbb{Z}$. The i^{th}-frame observation $\boldsymbol{y}^{(i)} = [y_0^{(i)}, \ldots, y_{N-1}^{(i)}]^t$ can be written

$$\boldsymbol{y}^{(i)} \;=\; \boldsymbol{T}^{(i)}\boldsymbol{h}^{(i)} + \boldsymbol{v}^{(i)} \tag{A.1}$$

where $\boldsymbol{h}^{(i)}$ is an $N \times 1$ vector described element-wise as $[\boldsymbol{h}^{(i)}]_l = h^{(i)}_{N_h \lfloor l/N_h \rfloor, \langle l \rangle_{N_h}}$, where $\boldsymbol{v}^{(i)}$ is a circular white Gaussian vector with covariance $\sigma_v^2 \boldsymbol{I}$, and where the block-fading assumption implies $\boldsymbol{T}^{(i)} = \text{blkdiag}\big(\boldsymbol{T}^{(i)}_0, \ldots, \boldsymbol{T}^{(i)}_{Q-1}\big)$ for

$$\boldsymbol{T}^{(i)}_q \;=\; \begin{bmatrix} t^{(i)}_{qN_h} & \cdots & t^{(i)}_{qN_h - N_h + 1} \\ \vdots & \ddots & \vdots \\ t^{(i)}_{qN_h + N_h - 1} & \cdots & t^{(i)}_{qN_h} \end{bmatrix}.$$

We are interested in estimating the i^{th} frame channel coefficients using i^{th} frame observations.

The channel vector $\boldsymbol{h}^{(i)}$ is zero-mean Gaussian, is uncorrelated with the noise, and has covariance $\text{E}\big\{\boldsymbol{h}^{(i)}\boldsymbol{h}^{(i)H}\big\} = \boldsymbol{R}_h$. Employing the WSSUS assumption, we can write $\boldsymbol{R}_h = \boldsymbol{R} \otimes \boldsymbol{I}_{N_h}$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{R}$ the autocorrelation matrix of $[h_{0,0}, h_{N_h,0}, h_{2N_h,0}, \ldots, h_{N-N_h,0}]^t$.

Consider the following persistent Kronecker-Delta (KD) pilot sequence:

$$t_n \;=\; \begin{cases} \sqrt{N_h} & \text{if } \frac{n}{N_h} \in \mathbf{Z} \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

Lemma 1. *The persistent KD sequence (A.2) is among the sequences minimizing the MSE of zero forcing (ZF) estimates for the block-fading channel assumed in this appendix.*

Proof. The ZF estimate of $\boldsymbol{h}^{(i)}$ from $\boldsymbol{y}^{(i)}$ is $\hat{\boldsymbol{h}}^{(i)} = \big(\boldsymbol{T}^{(i)}\big)^{-1}\boldsymbol{y}^{(i)} = \boldsymbol{h}^{(i)} + \big(\boldsymbol{T}^{(i)}\big)^{-1}\boldsymbol{v}^{(i)}$, with i^{th}-frame MSE

$$\begin{aligned} \big(\sigma_e^{(i)}\big)^2 \;&=\; \text{tr}\left(E\left\{\left(\boldsymbol{h}^{(i)} - \hat{\boldsymbol{h}}^{(i)}\right)\left(\boldsymbol{h}^{(i)} - \hat{\boldsymbol{h}}^{(i)}\right)^H\right\}\right) \\ &=\; \text{tr}\left(E\left\{\left(\big(\boldsymbol{T}^{(i)}\big)^{-1}\boldsymbol{v}^{(i)}\right)\left(\big(\boldsymbol{T}^{(i)}\big)^{-1}\boldsymbol{v}^{(i)}\right)^H\right\}\right) \\ &=\; \text{tr}\left(\sigma_v^2\big(\boldsymbol{T}^{(i)}\big)^{-1}\big(\boldsymbol{T}^{(i)}\big)^{-H}\right). \end{aligned} \tag{A.3}$$

Thus, the MSE-minimizing transmitted sequence minimizes $\mathrm{tr}\big(\big(\boldsymbol{T}^{(i)}\big)^{-1}\big(\boldsymbol{T}^{(i)}\big)^{-H}\big)$ subject to the power constraint. Note that, to minimize average MSE, it is sufficient to consider $\boldsymbol{T}^{(i)} = \boldsymbol{T}^{(0)}\ \forall i$, i.e., an N-periodic sequence. Hence, we drop the superscript $(\cdot)^{(i)}$ for the remainder of this proof.

The unit-power constraint can be rewritten $NN_h = \mathrm{tr}\big(\boldsymbol{T}^H\boldsymbol{T}\big) = \sum_{i=0}^{N-1}\lambda_i$, where $\{\lambda_i\}_{i=0}^{i=N-1}$ are the eigenvalues of $\boldsymbol{T}^H\boldsymbol{T}$, and where the structure of $\boldsymbol{T}^H\boldsymbol{T}$ guarantees that $\lambda_i \geq 0\ \forall i$. Note that $\mathrm{tr}\big(\boldsymbol{T}^{-1}\boldsymbol{T}^{-H}\big) = \mathrm{tr}\big(\big[\boldsymbol{T}^H\boldsymbol{T}\big]^{-1}\big) = \sum_{i=0}^{N-1}\lambda_i^{-1}$. Minimizing $\sum_{i=0}^{N-1}\lambda_i^{-1}$ subject to the constraints $\lambda_i \geq 0\ \forall i$ and $\sum_{i=0}^{N-1}\lambda_i = NN_h$ can be accomplished via the method of Lagrange multipliers. This yields $\lambda_0^{opt} = \lambda_1^{opt} = \cdots = \lambda_{N-1}^{opt} = N_h$, or, equivalently, $\boldsymbol{T}_{opt}^H\boldsymbol{T}_{opt} = N_h\boldsymbol{I}_N$. The persistent KD sequence (A.2) leads to $\boldsymbol{T}_q = \sqrt{N_h}\boldsymbol{I}_{N_h}\ \forall q$, and hence is among the minimum-MSE transmitted sequences. □

Lemma 2. *The persistent KD sequence (A.2) is among the N_h-periodic sequences minimizing the MSE of Wiener estimates for the block-fading channel assumed in this appendix.*

Proof. The N_h-periodicity assumption implies that $\boldsymbol{T}_q^{(i)} = \boldsymbol{T}_0^{(0)}\ \forall i, q$, where $\boldsymbol{T}_0^{(0)}$ is a circulant matrix with first column $[t_0, t_1, \ldots, t_{N_h-1}]^t$. Since $\boldsymbol{T}^{(i)} = \boldsymbol{I}_Q \otimes \boldsymbol{T}_0^{(0)}$ is invariant to i, we omit the superscript notation on $\boldsymbol{T}^{(i)}$ and $\boldsymbol{T}_0^{(0)}$. The Wiener estimate of $\boldsymbol{h}^{(i)}$ using $\boldsymbol{y}^{(i)}$ is then given by [15]

$$\hat{\boldsymbol{h}}^{(i)} = \boldsymbol{R}_h\boldsymbol{T}^H\left(\boldsymbol{T}\boldsymbol{R}_h\boldsymbol{T}^H + \sigma_v^2\boldsymbol{I}\right)^{-1}\boldsymbol{y}^{(i)} \tag{A.4}$$

with i^{th}-frame MSE

$$\begin{aligned}(\sigma_e^{(i)})^2 &= \mathrm{tr}\left(\boldsymbol{R}_h - \boldsymbol{R}_h\boldsymbol{T}^H\left(\boldsymbol{T}\boldsymbol{R}_h\boldsymbol{T}^H + \sigma_v^2\boldsymbol{I}\right)^{-1}\boldsymbol{T}\boldsymbol{R}_h\right)\\ &= \mathrm{tr}\left(\left[\boldsymbol{R}_h^{-1} + \sigma_v^{-2}\boldsymbol{T}^H\boldsymbol{T}\right]^{-1}\right).\end{aligned} \tag{A.5}$$

Since $\sigma_e^{(i)}$ is invariant to i we omit its superscript as well. Noting that the structure of $\boldsymbol{R}_h$ implies an eigenvalue decomposition $\boldsymbol{R}_h = (\boldsymbol{U} \otimes \boldsymbol{I}_{N_h})(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_{N_h})(\boldsymbol{U} \otimes \boldsymbol{I}_{N_h})^H$, where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Lambda}$ is diagonal, and that the structure of $\boldsymbol{T}$ implies $\boldsymbol{T}^H\boldsymbol{T} = \boldsymbol{I}_Q \otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)$, the MSE (A.5) can be rewritten as

$$\begin{aligned}
\sigma_e^2 &= \mathrm{tr}\Big(\Big[\big(\boldsymbol{U} \otimes \boldsymbol{I}_{N_h}\big)\big(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_{N_h}\big)\big(\boldsymbol{U}^H \otimes \boldsymbol{I}_{N_h}\big) \\
&\quad + \sigma_v^{-2}\big(\boldsymbol{I}_Q \otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)\big)\Big]^{-1}\Big) \\
&= \mathrm{tr}\Big(\Big[\big(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_{N_h}\big) \\
&\quad + \sigma_v^{-2}\big(\boldsymbol{U}^H \otimes \boldsymbol{I}_{N_h}\big)\big(\boldsymbol{I}_Q \otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)\big)\big(\boldsymbol{U} \otimes \boldsymbol{I}_{N_h}\big)\Big]^{-1}\Big).
\end{aligned}$$

Since $\big(\boldsymbol{U}^H\otimes\boldsymbol{I}_{N_h}\big)\big(\boldsymbol{I}_Q\otimes\big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)\big)\big(\boldsymbol{U}\otimes\boldsymbol{I}_{N_h}\big) = \big(\boldsymbol{U}^H\otimes\boldsymbol{I}_{N_h}\big)\big(\boldsymbol{U}\otimes\big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)\big) = \big(\boldsymbol{U}^H\boldsymbol{U}\big)\otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big) = \boldsymbol{I}_Q \otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)$,

$$\begin{aligned}
\sigma_e^2 &= \mathrm{tr}\Big(\Big[\big(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_{N_h}\big) + \sigma_v^{-2}\big(\boldsymbol{I}_Q \otimes \big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)\big)\Big]^{-1}\Big) \\
&= \sum_{i=0}^{Q-1} \mathrm{tr}\left(\left[\lambda_i^{-1}\boldsymbol{I}_{N_h} + \sigma_v^{-2}\boldsymbol{T}_0^H\boldsymbol{T}_0\right]^{-1}\right) \\
&\geq \sum_{i=0}^{Q-1}\sum_{m=0}^{N_h-1} \frac{1}{\lambda_i^{-1} + \sigma_v^{-2}\big(\boldsymbol{T}_0^H\boldsymbol{T}_0\big)_{m,m}} \qquad \text{(A.6)} \\
&= \sum_{i=0}^{Q-1}\sum_{m=0}^{N_h-1} \frac{1}{\lambda_i^{-1} + N_h\sigma_v^{-2}} \qquad \text{(A.7)}
\end{aligned}$$

where the equality in (A.6) holds iff $\boldsymbol{T}_0^H\boldsymbol{T}_0$ is diagonal and (A.7) follows from the unit power constraint. Note that the persistent KD sequence (A.2) ensures $\boldsymbol{T}_0^H\boldsymbol{T}_0 = N_h\boldsymbol{I}_{N_h}$ and thus attains the MMSE lower bound (A.7). □

BIBLIOGRAPHY

[1] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4 ed., 2001.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Trans. Signal Processing*, vol. 52, Apr. 2004.

[9] 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.

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

[11] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE press, 1993.

[12] 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.

[13] 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.

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

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

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

[17] A. P. Kannu and P. Schniter, "Reduced-complexity decision-directed pilot-aided tracking of doubly-selective channels," in *Proc. Conf. Inform. Science and Systems*, 2004.

[18] A. Graham, *Kronecker Products and Matrix Calculus with Applications*. New York: Halsted Press, John Wiley and Sons, 1981.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.

[20] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," in *IEEE Trans. Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.

[21] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmission over time-varying flat fading channels," *IEEE Trans. Signal Processing*, May 2004. (to appear).

[22] C. Sgraja and J. Lindner, "Estimation of rapid time-variant channels for ofdm using wiener filtering," in *Proc. IEEE Int. Conf. Commun.*, vol. 4, pp. 2390–2395, May 2003.

[23] M. K. Tsatsanis and Z. Xu, "Pilot symbol assisted modulation in frequency selective wireless channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 2353–2365, Aug. 2000.

[24] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[25] S. B. Bulumulla, S. A. Kassam, and S. S. Venkatesh, "A systematic approach to detecting OFDM signals in a fading channel," *IEEE Trans. Commun.*, vol. 48, pp. 725–728, May 2000.

[26] X. Cai and G. B. Giannakis, "Low-complexity ICI suppression for OFDM over time- and frequency-selective Rayleigh fading channels," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Nov. 2002.

[27] S. D'Silva, "On OFDM in doubly-dispersive channels," Master's thesis, The Ohio State University, Dec. 2002.

[28] S. N. Diggavi, "Analysis of multicarrier transmission in time-varying channels," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, pp. 1191–1195, 1997.

[29] B. Hirosaki, "An orthogonally multiplexed QAM system using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. 29, pp. 982–989, July 1981.

[30] Y. H. Kim, I. Song, H. G. Kim, T. Chang, and H. M. Kim, "Performance analysis of a coded OFDM system in time-variant multipath Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 48, no. 5, pp. 1610–1615, 1999.

[31] Y. Li and L. J. Cimini, Jr., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, pp. 401–404, Mar. 2001.

[32] P. Robertson and S. Kaiser, "The effects of Doppler spreads on OFDM(A) mobile radio systems," in *Proc. IEEE Veh. Tech. Conf.*, vol. 1, pp. 329–333, 1999.

[33] M. Russell and G. L. Stüber, "Interchannel interference analysis of OFDM in a mobile environment," in *Proc. IEEE Veh. Tech. Conf.*, vol. 2, pp. 820–824, 1995.

[34] D. Schaufhuber, G. Matz, and F. Hlawatsch, "Kalman tracking of time-varying channels in wireless MIMO-OFDM systems," in *Proc. Asilomar Conf. Signals, Systems and Computers*, 2003.

[35] T. Strohmer and S. Beaver, "Optimal OFDM design for time-frequency dispersive channels," *IEEE Trans. Commun.*, vol. 51, pp. 1111–1122, July 2003.

[36] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Baörjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, July 1998.

[37] X. Wang and K. J. R. Liu, "OFDM channel estimation based on time-frequency polynomial model of fading multipath channels," in *Proc. IEEE Global Telecommunications Conf.*, pp. 212–216, 2001.