An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM

Tareq Y. Al-Naffouri

Abstract— OFDM modulation combines the advantages of high achievable rates and relatively easy implementation. However, for proper recovery of the input, the OFDM receiver needs accurate channel information. In this paper, we propose an expectation-maximization (EM) algorithm for joint channel and data recovery in fast fading environments. The algorithm makes a *collective* use of the data and channel constraints inherent in the communication problem. This comes in contrast to other works which have employed these constraints selectively. The data constraints include pilots, the cyclic prefix, and the finite alphabet restriction, while the channel constraints include sparsity, finite delay spread, and the statistical properties of the channel (frequency and time correlation). The algorithm boils down to a forward-backward (FB) Kalman filter. We also suggest a suboptimal modification that is able to track the channel and recover the data with no latency. Simulations show the favorable behavior of both algorithms compared to other channel estimation techniques.

Index Terms— OFDM, time-variant channels, channel modelling, frequency correlation, time correlation, channel estimation, Kalman filters, expectation-maximization algorithm.

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) is an effective technique for high bit-rate transmission. It has found widespread applications and is already part of many standards (e.g. 802.11a/b/g and 802.11a/e). OFDM combats intersymbol interference by prepending a guard band (cyclic prefix) to the transmitted symbol. This effectively divides the channel into many narrowband ISI-free channels over which parallel streams of data are transmitted. Frequency selectivity can now be mitigated using one tap equalizers. For proper operation of an OFDM receiver, it needs an accurate estimate of the channel state. For rapidly time-variant channels, the receiver faces the additional challenge of performing channel (and data) recovery for each OFDM symbol. In carrying out these two operations, the receiver takes advantage of the rich structure of the underlying communication problem. This structure can be traced back to some inherent constraints on the data or on the channel. Table I lists the most common of these constraints and some of the works that employed them.

A. Approaches to channel estimation in OFDM

Several algorithms have been suggested in literature for channel estimation in OFDM transmission. Each of these algorithms makes use of a subset of the constraints in Table I. These algorithms can be classified into one of the following categories

1) **Training-based estimation:** Pilots are used to perform channel estimation as in [11] and [14].

2) **Blind estimation:** At the other extreme, blind algorithms rely completely on natural constraints inherent in the communication problem to perform channel recovery. For example, [3] used frequency correlation and code, [5] used the cyclostationarity induced by the cyclic prefix and transmitter precoding, [26] used a subspace constraint, and [6] used the cyclic prefix.

3) **Semi-blind estimation:** Semi-blind techniques are a hybrid of blind and training based techniques, utilizing pilots and other natural constraints to perform channel estimation (see for example [3], [19],[21], [22], [24], [27]).

4) **Data-aided channel estimation:** The main and perhaps the only reason to perform channel estimation at the receiver is to use the estimate along with the channel output to recover the transmitted data. One can, in turn, use the detected data to enhance the channel estimate giving rise to an iterative technique for channel and data recovery. With this in mind, it is natural for the two operations of channel and data recovery to be considered jointly, especially since one operation can be used to enhance the performance of the other. This intuitive idea is the basis of joint channel estimation and data detection proposed in [21], [22], [28]. Other works, like [24], [29], and [2], arrived at iterative techniques more rigorously by employing the expectation-maximization (EM) algorithm. The data-aided approach seems the most sensible for channel estimation, especially when the channel is time variant.[1]

The aforementioned works utilize *only a subset* of the constraints on the channel and data. In this paper, however, we present a (data-aided EM) method that can make use of all the constraints in Table I[2]. The method boils down to a forward-backward (FB) Kalman filter which we derive in Section III. One consequence of our approach is the increased storage and latency requirements of the FB-Kalman as it has to process multiple OFDM symbols simultaneously. We thus suggest in Section IV a suboptimal forward-only version (basically a Kalman filter) that is able to perform channel recovery with no latency. These two algorithms are compared and benchmarked in our simulations in Section V. We start by introducing our notation and the system model.

B. Notation

We denote scalars with small-case letters, vectors with small-case boldface letters, and matrices with uppercase boldface letters. We also reserve calligraphic notation (e.g., $\boldsymbol{\mathcal{X}}$ and $\mathcal{X}$) for variables in the frequency domain. The individual entries of a variable like $\underline{h}$ are denoted by $\underline{h}(l)$. When any of these variables becomes a function of time, the time index i appears as a subscript (e.g., we write x_i, $\underline{h}_i(l)$, $\underline{h}_i$, and $\boldsymbol{\mathcal{X}}_i$), and use the notation $\underline{h}_0^T$ to denote the sequence $(\underline{h}_0, \underline{h}_1, \cdots, \underline{h}_T)$.

We also maintain the following conventions: 1) a hat over a variable indicates an estimate of the variable (e.g., $\hat{\underline{h}}$ is an estimate of $\underline{h}$), 2) all vectors are column vectors, 3) all underlined vectors are of length P (e.g., the cyclic prefix $\underline{x}_i$) or of length $P+1$ (e.g., the impulse response $\underline{h}_i$), 4) all vectors with no bars are of length N (e.g., $\boldsymbol{\mathcal{X}}_i$), and 5) all overlined vectors are of length $N+P$ (e.g., $\overline{x}_i$).

II. SYSTEM MODEL

Consider the iid sequence $\mathcal{X}_0^{(T+1)N} = (\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_{(T+1)N})$ to be transmitted. In an OFDM system, the sequence is parsed into a sequence of $T+1$ data symbols $\boldsymbol{\mathcal{X}}_0^T = (\boldsymbol{\mathcal{X}}_0, \boldsymbol{\mathcal{X}}_1, \cdots, \boldsymbol{\mathcal{X}}_T)$ each of length N. Each symbol $\boldsymbol{\mathcal{X}}_i$ undergoes an IDFT operation to produce the time domain symbol $x_i = \sqrt{N}\boldsymbol{Q}^*\boldsymbol{\mathcal{X}}_i$, where $\boldsymbol{Q}$ is the $N \times N$ DFT matrix for which $q_{l,m} = e^{-j\frac{2\pi(l-1)(m-1)}{N}}$. The transmitter then

---

TABLE I
Data and channel constraints used for channel estimation

CONSTRAINTS		ASSUMPTIONS	REFERENCE
Data Constraints		Finite alphabet constraint	[1], [2]
		Code	[3], [4]
		Transmit precoding (e.g., cyclic prefix,silent guard band)	[5], [6], [7], [8], [9], [3]
		Pilots	[10], [11], [12],[13],[14]
Channel Constraints		Finite delay spread	[5], [3], [11]
		Sparsity	[15], [16], [17]
		Frequency correlation	[18],[19], [3]
		Time correlation	[20], [21], [22], [23], [24]
		Uncertainty information	[25], [18]

appends a cyclic prefix (CP) $\underline{\boldsymbol{x}}_i$ (of length P) to $\boldsymbol{x}_i$, resulting finally in a sequence of super-symbols $\overline{\boldsymbol{x}}_0^T$ each of length $N + P$.

We assume that the channel $\underline{\boldsymbol{h}}_i$ (of maximum length $P + 1$) remains fixed over a given OFDM symbol (and its associated CP) but varies from one symbol to the next according to a state-space model

$$\boxed{\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i, \ \ \underline{\boldsymbol{h}}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Pi}_0), \ \ \underline{\boldsymbol{u}}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I})} \quad (1)$$

The matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ in (1) are square matrices of size $P + 1$ and are assumed available to the receiver. In the Appendix, we show how we can construct such a model from the knowledge of the Doppler frequency (time-correlation), the power-delay profile (frequency-correlation), and the receive filter. At the channel output, we obtain a sequence of time-domain super-symbols $\overline{\boldsymbol{y}}_0^T$, which after stripping the cyclic prefix $\underline{\boldsymbol{y}}_i$, produce a sequence of time-domain symbols $\boldsymbol{y}_0^T$. The input/output (I/O) relationship of the OFDM system is best described in the frequency domain

$$\boldsymbol{\mathcal{Y}}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{\mathcal{H}}_i + \boldsymbol{\mathcal{N}}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{Q}_{P+1}\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i \quad (2)$$

where $\boldsymbol{\mathcal{N}}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$ is the additive noise. The second equality in (2) follows from the DFT relationship $\boldsymbol{\mathcal{H}}_i \triangleq \boldsymbol{Q} \begin{bmatrix} \underline{\boldsymbol{h}}_i \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{Q}_{P+1}\underline{\boldsymbol{h}}_i$, where $\boldsymbol{Q}_{P+1}$ consists of the first $P+1$ columns of $\boldsymbol{Q}$. Alternatively, with $\boldsymbol{X}_i \triangleq \text{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{Q}_{P+1}$, we can write

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \boldsymbol{X}_i\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i} \quad (3)$$

A similar relationship exists between the cyclic prefixes at the input and the output

$$\boxed{\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i\underline{\boldsymbol{h}}_i + \underline{\boldsymbol{n}}_i} \quad (4)$$

where $\underline{\boldsymbol{X}}_i$ is a $P \times (P + 1)$ Toeplitz matrix constructed from the vector of cyclic prefixes $[\underline{\boldsymbol{x}}_{i-1}^T, \underline{\boldsymbol{x}}_i^T]^T$. By concatenating (3) and (4), we obtain the total input/output relationship

$$\boxed{\overline{\boldsymbol{\mathcal{Y}}}_i = \overline{\boldsymbol{X}}_i\underline{\boldsymbol{h}}_i + \overline{\boldsymbol{\mathcal{N}}}_i} \quad (5)$$

which incorporates the effect of the OFDM symbol as well as the cyclic prefix observation.

A. Pilot/output relationships

In general, the receiver needs pilots to obtain initial channel estimates. Let the index set $I_\text{p} = \{i_1, i_2, \dots, i_{L_\text{p}}\}$ denote the pilot locations within the OFDM symbol. Also, let $\boldsymbol{X}_{I_\text{p}}$ denote the matrix $\boldsymbol{X}$ pruned of the rows that do not belong to I_p. Then, the pilot/output equation can be derived from the I/O relationship (3) as

$$\boxed{\boldsymbol{\mathcal{Y}}_{iI_\text{p}} = \boldsymbol{X}_{iI_\text{p}}\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_{iI_\text{p}}} \quad (6)$$

III. THE EM ALGORITHM FOR JOINT CHANNEL AND DATA ESTIMATION

A. The EM algorithm

Ideally, we estimate $\underline{\boldsymbol{h}}_i$ using some I/O relationship, e.g. (3), by maximizing the corresponding log-likelihood function

$$\hat{\underline{\boldsymbol{h}}}_i^{\text{MAP}} = \arg\max_{\underline{\boldsymbol{h}}_i} \ \{\ln p(\boldsymbol{\mathcal{Y}}_i | \boldsymbol{X}_i, \underline{\boldsymbol{h}}_i) + \ln p(\underline{\boldsymbol{h}}_i)\}$$

For example, when the channel obeys the I/O relationship (3) (so that $\ln p(\boldsymbol{\mathcal{Y}}_i | \boldsymbol{X}_i, \underline{\boldsymbol{h}}_i) = -\|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\underline{\boldsymbol{h}}_i\|_{\sigma_n^{-2}}^2$ up to some additive constant) and $\underline{\boldsymbol{h}}_i$ is $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Pi})$, (so that $\ln p(\underline{\boldsymbol{h}}_i) = -\|\underline{\boldsymbol{h}}_i\|_{\Pi^{-1}}^2$ up to some additive constant), in this case, the MAP estimate is given by[3]

$$\hat{\underline{\boldsymbol{h}}}_i^{\text{MAP}} = \arg\min_{\underline{\boldsymbol{h}}_i} \left\{ \|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\underline{\boldsymbol{h}}_i\|_{\sigma_n^{-2}}^2 + \|\underline{\boldsymbol{h}}_i\|_{\Pi^{-1}}^2 \right\}$$

In our case, however, the input $\boldsymbol{\mathcal{X}}_i$ (or $\boldsymbol{X}_i$ [4]) is not observable. Thus, we use the expectation-maximization algorithm and maximize instead an averaged form of the log-likelihood function. Specifically, starting from an initial estimate $\hat{\underline{\boldsymbol{h}}}_i^{(0)}$, the estimate $\hat{\underline{\boldsymbol{h}}}_i$ is calculated iteratively, with the estimate at the j-th iteration given by

$$\hat{\underline{\boldsymbol{h}}}_i^{(j)} = \arg\max_{\underline{\boldsymbol{h}}_i} \left\{ E_{X_i | \mathcal{Y}_i, \hat{h}_i^{(j-1)}} \ln p(\boldsymbol{\mathcal{Y}}_i | \boldsymbol{X}_i, \underline{\boldsymbol{h}}_i) + \ln p(\underline{\boldsymbol{h}}_i) \right\}$$

Thus, for the example above, the EM-based estimate (at the j-th iteration) is given by[5]

$$\hat{\underline{\boldsymbol{h}}}_i^{(j)} = \arg\min_{\underline{\boldsymbol{h}}_i} \left\{ \|\boldsymbol{\mathcal{Y}}_i - E[\boldsymbol{X}_i]\underline{\boldsymbol{h}}_i\|_{\sigma_n^{-2}}^2 + \|\underline{\boldsymbol{h}}_i\|_{\text{Cov}[X_i^*]}^2 + \|\underline{\boldsymbol{h}}_i\|_{\Pi^{-1}}^2 \right\}$$

where the two moments of $\boldsymbol{X}_i$ are taken given the output $\boldsymbol{\mathcal{Y}}_i$ and the most recent channel estimate $\hat{\underline{\boldsymbol{h}}}_i^{(j-1)}$. We now derive the EM algorithm for the time-variant case.

B. The EM-based forward-backward Kalman

Consider the OFDM system of section 2, essentially described by the state-space model

$$\begin{aligned} \underline{\boldsymbol{h}}_{i+1} &= \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i & (7) \\ \boldsymbol{\mathcal{Y}}_i &= \boldsymbol{X}_i\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i & (8) \end{aligned}$$

with $\underline{\boldsymbol{h}}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Pi}_0)$ and $\underline{\boldsymbol{u}}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I})$. Given a sequence of $T+1$ input and output symbols $\boldsymbol{X}_0^T$ and $\boldsymbol{\mathcal{Y}}_0^T$, we obtain the MAP estimate of the channel sequence $\underline{\boldsymbol{h}}_0^T$ by maximizing the log-likelihood

$$\mathcal{L} = \ln p(\boldsymbol{\mathcal{Y}}_0^T | \boldsymbol{X}_0^T, \underline{\boldsymbol{h}}_0^T) + \ln p(\underline{\boldsymbol{h}}_0^T)$$

---

[3]We use the weighted norm $\|\underline{\boldsymbol{h}}\|_\Sigma^2$ to denote $\underline{\boldsymbol{h}}^* \Sigma \underline{\boldsymbol{h}}$.

[4]Since $\boldsymbol{X}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{Q}_{P+1}$, conditioning on $\boldsymbol{\mathcal{X}}_i$ can be replaced by conditioning on $\boldsymbol{X}_i$.

[5]The Covariance $\text{Cov}[X_i^*]$ is defined as $\text{Cov}[X_i^*] = E[\boldsymbol{X}_i^* \boldsymbol{X}_i] - E[\boldsymbol{X}_i^*]E[\boldsymbol{X}_i]$.

Now, using (8), we can express the first term of the log-likelihood (up to some additive constant) as

$$\ln p(\boldsymbol{\mathcal{Y}}_0^T | \boldsymbol{\mathcal{X}}_0^T, \underline{\boldsymbol{h}}_0^T) = \sum_{i=0}^{T} \ln p(\boldsymbol{\mathcal{Y}}_i | \boldsymbol{\mathcal{X}}_i, \underline{\boldsymbol{h}}_i)$$

$$= -\sum_{i=0}^{T} \|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}}$$

Similarly, using (7), we can express the second term (again up to some additive constant) as

$$\ln p(\underline{\boldsymbol{h}}_0^T) = \sum_{i=1}^{T} \ln p(\underline{\boldsymbol{h}}_i | \underline{\boldsymbol{h}}_{i-1}) + \ln p(\underline{\boldsymbol{h}}_0)$$

$$= -\sum_{k=1}^{T} \|\underline{\boldsymbol{h}}_k - \boldsymbol{F}\underline{\boldsymbol{h}}_{k-1}\|^2_{\frac{1}{\sigma_u^2} GG^*} - \|\underline{\boldsymbol{h}}_0\|^2_{\Pi_0^{-1}} \quad (9)$$

Combining these two expressions yields

$$\mathcal{L} = -\sum_{i=0}^{T} \|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} - \|\underline{\boldsymbol{h}}_0\|^2_{\Pi_0^{-1}} -$$

$$\sum_{i=1}^{T} \|\underline{\boldsymbol{h}}_i - \boldsymbol{F}\underline{\boldsymbol{h}}_{i-1}\|^2_{\frac{1}{\sigma_u^2} GG^*} \quad (10)$$

Since the channel sequence $\underline{\boldsymbol{h}}_0^T$ is jointly Gaussian, the MAP estimate of the channel sequence given the input and output sequences $\boldsymbol{X}_0^T$ and $\boldsymbol{\mathcal{Y}}_0^T$ is the same as the MMSE estimate given the same sequences. The MMSE estimate itself is obtained by the FB Kalman filter. This allows us to state the following theorem (for a proof, see problem 10.9 in [31]).

Theorem 1: **Channel estimation–Known input case** Consider the state-space model (7)–(8). Given the input and output sequences $\boldsymbol{X}_0^T$ and $\boldsymbol{\mathcal{Y}}_0^T$, the MAP (or equivalently MMSE) estimate of $\underline{\boldsymbol{h}}_0^T$ is obtained by applying the following (forward-backward Kalman) filter to the state-space model (7)–(8)
Forward run: Starting from the initial conditions $\boldsymbol{P}_{0|-1} = \Pi_0$ and $\underline{\boldsymbol{h}}_{0|-1} = \boldsymbol{0}$ and for $i = 1, \ldots, T$, calculate

$$\boldsymbol{R}_{e,i} = \sigma_n^2 \boldsymbol{I}_{N+P} + \boldsymbol{X}_i \boldsymbol{P}_{i|i-1} \boldsymbol{X}_i^* \quad (11)$$

$$\boldsymbol{K}_{f,i} = \boldsymbol{P}_{i|i-1} \boldsymbol{X}_i^* \boldsymbol{R}_{e,i}^{-1} \quad (12)$$

$$\hat{\underline{\boldsymbol{h}}}_{i|i} = (\boldsymbol{I}_{N+P} - \boldsymbol{K}_{f,i} \boldsymbol{X}_i) \hat{\underline{\boldsymbol{h}}}_{i|i-1} + \boldsymbol{K}_{f,i} \boldsymbol{\mathcal{Y}}_i, \quad (13)$$

$$\hat{\underline{\boldsymbol{h}}}_{i+1|i} = \boldsymbol{F} \hat{\underline{\boldsymbol{h}}}_{i|i} \quad (14)$$

$$\boldsymbol{P}_{i+1|i} = \boldsymbol{F}_i \left(\boldsymbol{P}_{i|i-1} - \boldsymbol{K}_{f,i} \boldsymbol{R}_{e,i} \boldsymbol{K}_{f,i}^* \right) \boldsymbol{F}^* + \frac{1}{\sigma_n^2} GG^* \quad (15)$$

Backward run: Starting from $\boldsymbol{\lambda}_{T+1|T} = \boldsymbol{0}$ and for $i = T, T-1, \ldots, 0$, calculate

$$\boldsymbol{\lambda}_{i|T} = \left(\boldsymbol{I}_{P+N} - \boldsymbol{X}_i^* \boldsymbol{K}_{f,i}^* \right) \boldsymbol{F}_i^* \boldsymbol{\lambda}_{i+1|T} +$$

$$\boldsymbol{X}_i \boldsymbol{R}_{e,i}^{-1} \left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \hat{\underline{\boldsymbol{h}}}_{i|i-1} \right) \quad (16)$$

$$\hat{\underline{\boldsymbol{h}}}_{i|T} = \hat{\underline{\boldsymbol{h}}}_{i|i-1} + \boldsymbol{P}_{i|i-1} \boldsymbol{\lambda}_{i|T} \quad (17)$$

The desired estimate is $\hat{\underline{\boldsymbol{h}}}_{i|T}$.

The FB Kalman obtains the MAP estimate of the channel impulse response. In the forward step, the filer obtains the MAP estimate of $\underline{\boldsymbol{h}}_i$ given $\boldsymbol{X}_0^i$. Our aim, however, is to obtain the MAP estimate of $\underline{\boldsymbol{h}}_i$ given the whole sequence $\boldsymbol{X}_0^T$. The backward step obtains the contribution of $\boldsymbol{X}_{i+1}^T$ to the MAP estimate of $\underline{\boldsymbol{h}}_i$.

Theorem 1 allows us to obtain the estimate of $\underline{\boldsymbol{h}}_0^T$ when the input sequence $\boldsymbol{X}_0^T$ is not observable with the help of the expectation maximization (EM) algorithm. Specifically, in the EM algorithm, we maximize the log-likelihood (10) *averaged* over the sequence $\boldsymbol{X}_0^T$. Thus, the j-th iteration of the EM algorithm is now obtained by

maximizing the averaged log-likelihood function $\overline{\mathcal{L}} = E_{X_0^T | \underline{h}_0^T, \mathcal{Y}_0^T} \mathcal{L}$. By inspecting (10), we note that the only term that is modified under expectation is the first summand, and its expectation is given by

$$E \|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} = \|\boldsymbol{\mathcal{Y}}_i - E[\boldsymbol{X}_i] \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} + \|\underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2} \text{Cov}[X_i^*]}$$

$$= \left\| \begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \boldsymbol{0}_{P \times 1} \end{bmatrix} - \begin{bmatrix} E[\boldsymbol{X}_i] \\ \text{Cov}[\boldsymbol{X}_i^*]^{1/2} \end{bmatrix} \underline{\boldsymbol{h}}_i \right\|^2_{\frac{1}{\sigma_n^2}}$$

where the expectations are taken given the most recent estimate $\hat{\underline{\boldsymbol{h}}}_0^{(j-1)}$ and the output symbols $\boldsymbol{\mathcal{Y}}_0^T$. We thus have

$$\overline{\mathcal{L}} = -\sum_{i=0}^{T} \left\| \begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \boldsymbol{0}_{P \times 1} \end{bmatrix} - \begin{bmatrix} E[\boldsymbol{X}_i] \\ \text{Cov}[\boldsymbol{X}_i^*]^{1/2} \end{bmatrix} \underline{\boldsymbol{h}}_i \right\|^2_{\frac{1}{\sigma_n^2}} -$$

$$\sum_{i=1}^{T} \|\underline{\boldsymbol{h}}_i - \boldsymbol{F}\underline{\boldsymbol{h}}_{i-1}\|^2_{\frac{1}{\sigma_u^2} GG^*} - \|\underline{\boldsymbol{h}}_0\|^2_{\Pi_0^{-1}} \quad (18)$$

Note that we can obtain the averaged likelihood (18) from the original likelihood (10) by performing the substitution

$$\boldsymbol{X}_i \longrightarrow \begin{bmatrix} E[\boldsymbol{X}_i] \\ \text{Cov}[\boldsymbol{X}_i^*]^{1/2} \end{bmatrix} \quad \boldsymbol{\mathcal{Y}}_i \longrightarrow \begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \boldsymbol{0}_{P \times 1} \end{bmatrix}$$

We can thus state the following theorem.

Theorem 2: **Channel estimation–Unknown input case** Consider the state-space model (7)–(8) and assume that the receiver does not have access to the transmitted data $\boldsymbol{X}_0^T$. The channel estimate at the j-th iteration $\underline{\boldsymbol{h}}_0^{T(j)}$ of the EM algorithm is obtained by applying the forward-backward Kalman (11)–(17) to the following state-space model

$$\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i \quad (19)$$

$$\begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \boldsymbol{0}_{P \times 1} \end{bmatrix} = \begin{bmatrix} E[\boldsymbol{X}_i] \\ \text{Cov}[\boldsymbol{X}_i^*]^{1/2} \end{bmatrix} \underline{\boldsymbol{h}}_i + \begin{bmatrix} \boldsymbol{\mathcal{N}}_i \\ \underline{\boldsymbol{n}}_i \end{bmatrix} \quad (20)$$

where $\underline{\boldsymbol{n}}_i \sim \mathcal{N}(\boldsymbol{0}_{P \times 1}, \sigma_n^2 \boldsymbol{I})$ is virtual noise that is independent of the physical noise $\boldsymbol{\mathcal{N}}_i$.

To fully implement the EM algorithm, we need to initialize the algorithm and calculate the first and second moments of the input–two steps which we perform next.

C. Initial channel estimation

We can obtain the initial channel estimate from the pilot/output equation (6). We do this by applying the FB Kalman to the state-space model

$$\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i \quad (21)$$

$$\boldsymbol{\mathcal{Y}}_{i I_{\text{p}}} = \boldsymbol{X}_{i I_{\text{p}}} \underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_{i I_{\text{p}}} \quad (22)$$

i.e., by applying the FB Kalman (11)–(17) with the substitutions $\boldsymbol{X}_i \longrightarrow \boldsymbol{X}_{i I_{\text{p}}}$ and $\boldsymbol{\mathcal{Y}}_i \longrightarrow \boldsymbol{\mathcal{Y}}_{i I_{\text{p}}}$

D. Calculating the input moments

Using the relationship $\boldsymbol{X}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i) \boldsymbol{Q}_{P+1}$, we can write

$$E[\boldsymbol{X}_i | \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{H}}_i] = \text{diag}(E[\boldsymbol{\mathcal{X}}_i | \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{H}}_i]) \boldsymbol{Q}_{P+1} \quad (23)$$

$$\text{Cov}[\boldsymbol{X}_i^* | \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{H}}_i] = \boldsymbol{Q}_{P+1}^* \text{Cov}[\boldsymbol{\mathcal{X}}_i^* | \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{H}}_i] \boldsymbol{Q}_{P+1} \quad (24)$$

We can calculate the mean and covariance of $\boldsymbol{\mathcal{X}}_i$ by calculating the first two moments of its individual elements $\mathcal{X}_i(l)$ $l = 1, \cdots N$ [6]. Now to calculate the two moments of $\mathcal{X}_i(l)$, we need to evaluate the pdf $f(\mathcal{X}_i(l)|\mathcal{Y}_i(l), \mathcal{H}_i(l))$. Applying Bayes rule yields

$$
\begin{aligned}
f(\mathcal{X}_i(l)|\mathcal{Y}_i(l), \mathcal{H}_i(l)) &= \frac{f(\mathcal{X}_i, \mathcal{Y}_i|\mathcal{H}_i)}{f(\mathcal{Y}_i|\mathcal{H}_i)} \\
&= \frac{f(\mathcal{X}_i, \mathcal{Y}_i|\mathcal{H}_i)}{\sum_{\mathcal{X}_i = A_1}^{A_M} f(\mathcal{Y}_i, \mathcal{X}_i|\mathcal{H}_i)} \\
&= \frac{f(\mathcal{Y}_i|\mathcal{X}_i, \mathcal{H}_i)f(\mathcal{X}_i|\mathcal{H}_i)}{\sum_{\mathcal{X}_i = A_1}^{A_M} f(\mathcal{Y}_i|\mathcal{X}_i, \mathcal{H}_i)f(\mathcal{X}_i|\mathcal{H}_i)} \\
&= \frac{e^{-\frac{|\mathcal{Y}_i - \mathcal{H}\mathcal{X}_i|^2}{\sigma_n^2}}}{\sum_{j=1}^{M} e^{-\frac{|\mathcal{Y}_i - \mathcal{H}A_j|^2}{\sigma_n^2}}}
\end{aligned}
$$

where we dropped the dependence on l for notational convenience and where, in the deriving the pdf above, we used the fact that $\mathcal{X}_i(l)$ is drawn from the alphabet $A = \{A_1, \ldots, A_M\}$ with equal probability. We can use this to show that

$$
E[\mathcal{X}_i(l)|\mathcal{Y}_i(l), \mathcal{H}_i(l)] = \frac{\sum_{j=1}^{M} A_j e^{-\frac{|\mathcal{Y}_i(l) - \mathcal{H}(l)A_j|^2}{\sigma_n^2}}}{\sum_{j=1}^{M} e^{-\frac{|\mathcal{Y}_i(l) - \mathcal{H}(l)A_j|^2}{\sigma_n^2}}} \quad (25)
$$

$$
E[|\mathcal{X}_i(l)|^2|\mathcal{Y}_i(l), \mathcal{H}_i(l)] = \frac{\sum_{j=1}^{M} |A_j|^2 e^{-\frac{|\mathcal{Y}_i(l) - \mathcal{H}(l)A_j|^2}{\sigma_n^2}}}{\sum_{j=1}^{M} e^{-\frac{|\mathcal{Y}_i(l) - \mathcal{H}(l)A_j|^2}{\sigma_n^2}}} \quad (26)
$$

Equations (23)–(24) represents the soft estimate of the input. We can perform channel estimation using the hard estimate of the input. This is obtained by rounding the first moment in (25)–(26 to the nearest QAM point.

E. Summary of the EM-based FB Kalman

1) Obtain the initial channel estimate $\underline{\boldsymbol{h}}_0^{T^{(0)}}$ by applying the FB Kalman (11)–(17) to the state-space model (21)–(22)
2) Iterate between the expectation and maximization steps for $j = 1, \ldots, N_{\text{iter}}$:
   a) **Expectation:** Compute the first two moments of the input $\boldsymbol{X}_0^T$ given the channel output $\boldsymbol{\mathcal{Y}}_0^T$ and the previous estimate of the channel $\underline{\boldsymbol{h}}_0^{T^{(j-1)}}$ using (25)–(26).
   b) **Maximization:** Obtain the channel estimate $\underline{\boldsymbol{h}}_0^{T^{(j)}}$ by applying the FB Kalman (11)–(17) to the state-space model (19)–(20).

The algorithm can be stopped when the difference between two consecutive estimates $\|\underline{\boldsymbol{h}}_0^{T^{(j)}} - \underline{\boldsymbol{h}}_0^{T^{(j-1)}}\|^2$ is below a certain threshold or when the maximum number of iterations N_{iter} is reached.

IV. THREE EXTENSIONS

A. Using the cyclic prefix observation

The FB Kalman can make use of the CP observation. Here pilot-based estimation remains the same while the EM algorithm is run on the I/O equation (5) which contains the effect of the cyclic prefix.

---

[6]To be precise, calculating the second moment (24) also calls for calculating the cross correlations $E[\mathcal{X}_i(l)\mathcal{X}_i(m)]$ for $l, m = 1, 2, \cdots, N$ as the elements of $\boldsymbol{\mathcal{X}}_i$ are *not* independent given $\boldsymbol{\mathcal{Y}}_i$ and $\boldsymbol{\mathcal{H}}_i$. Calculating these cross moments however is computationally complex and not inline with the OFDM philosophy which achieves equalization on an element by element basis. As such, in calculating $\text{Cov}[\boldsymbol{\mathcal{X}}_i^*|\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{H}}_i]$, we will assume the elements of $\boldsymbol{\mathcal{X}}_i$ are iid given $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{Y}}_i$.

Thus, in this case, we apply the FB Kalman (11)–(17) to the state-space model

$$
\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i \quad (27)
$$

$$
\begin{bmatrix} \overline{\boldsymbol{\mathcal{Y}}}_i \\ \boldsymbol{0}_{P \times 1} \end{bmatrix} = \begin{bmatrix} E[\overline{\boldsymbol{X}}_i] \\ \text{Cov}[\overline{\boldsymbol{X}}_i^*]^{1/2} \end{bmatrix} \underline{\boldsymbol{h}}_i + \begin{bmatrix} \overline{\boldsymbol{\mathcal{N}}}_i \\ \underline{\boldsymbol{n}}_i \end{bmatrix} \quad (28)
$$

The two moments of $\overline{\boldsymbol{X}}_i$ can be obtained from (25)–(26) but the calculations become more cumbersome due to the presence of the CP (see [30]).

B. Using the hard estimate of the input

We can simplify our algorithm by using the hard estimate of the input. In this case, the pilot based estimate remains the same. In the expectation step, however, we only calculate the first moment of the input $\boldsymbol{X}_i$ and round the estimate to the nearest QAM point, $\hat{\boldsymbol{X}}_i$. The channel estimate is subsequently obtained by applying the FB-Kalman (11)–(17) to the state-space model

$$
\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i \quad (29)
$$

$$
\boldsymbol{\mathcal{Y}}_i = \hat{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i \quad (30)
$$

C. Kalman- (forward-only) based estimation

The FB Kalman requires considerable storage and latency. The algorithm needs to wait for all $T + 1$ symbols before it can execute the backward run and hence obtain the channel estimate. One way around this is to reduce the window size T. Alternatively, we can run the filter in the forward direction only (i.e., run (11)–(15)) for both the initial estimation and the EM iteration [7].The algorithm then collapses to the Kalman-based filter proposed in [32] where the data and channel are recovered within one OFDM symbol, i.e. with no latency.

V. SIMULATIONS

We consider an OFDM system that transmits a sequence of 5 symbols, each with 64 carriers and a cyclic prefix of length $P = 15$. The input data is 16 QAM mapped from a binary bit stream through Gray coding. We use 16 pilots in the first symbol and fix the number of pilots in the subsequent symbols to x where $4 \leq x \leq 16$.

The channel IR consists of 16 complex taps (the maximum length possible). The initial IR $\underline{\boldsymbol{h}}_0$ has an exponential delay profile $E[|\underline{h}_0(k)|^2] = e^{-0.2k}$. For $i > 0$, $\underline{\boldsymbol{h}}_i$ is generated according to the dynamical model $\underline{\boldsymbol{h}}_{i+1} = \boldsymbol{F}\underline{\boldsymbol{h}}_i + \boldsymbol{G}\underline{\boldsymbol{u}}_i$ where $\boldsymbol{F}$ and $\boldsymbol{G}$ are diagonal matrices. Specifically, we set $\boldsymbol{F} = f\boldsymbol{I}$ with $f = .7, .8$, or .9 and set the diagonal entries of $\boldsymbol{G}$ as $G(k, k) = \sqrt{(1 - f^2)E[|\underline{h}_0(k)|^2]}$. The state noise $\underline{\boldsymbol{u}}_i$ is iid with unit variance. This choice of parameters ensures that the channel maintains the same delay profile at subsequent time instants. Throughout the simulations, we run the EM algorithm for 4 iterations.

A. Comparing the Kalman and the forward-backward Kalman

In Figure 1, we compare the performance of FB-Kalman and the Kalman based receivers. We carry out this comparison for two levels of time variation ($f = .7, .9$), progressively increasing the number of pilots from $x = 4$ to $x = 16$. As expected, the FB-Kalman consistently outperforms the Kalman filter-based receiver.

---

[7]The Kalman-based algorithm is thus the same algorithm summarized in Subsection III-E. The only difference is that equations (11)–(17) are replaced by the subset of equations (11)–(15).
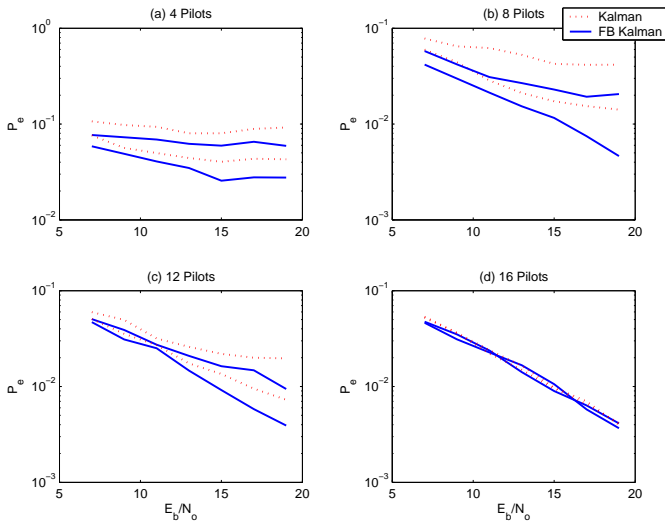
Fig. 1. The FB-Kalman (solid lines) outperforms the Kalman (dotted lines) for two levels of time variation (the top curves correspond to $f = .7$ and the lower to $f = .9$)
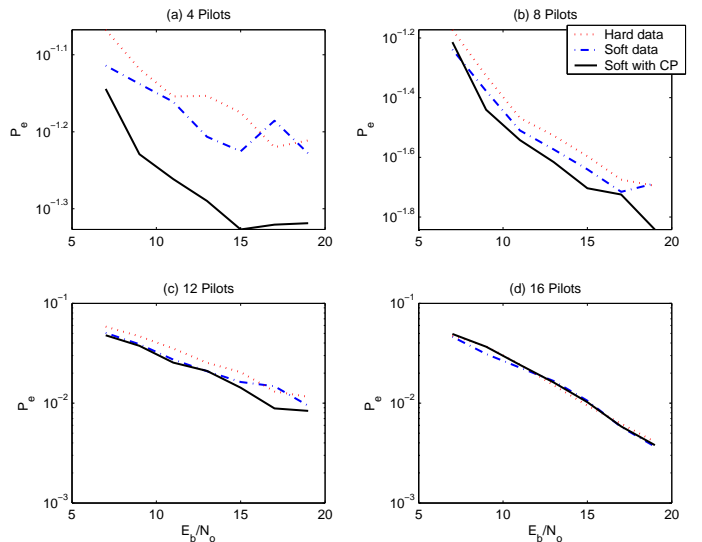

Fig. 2. The FB Kalman-based receiver demonstrates improved BER with increasing levels of signal processing ($f = .7$)
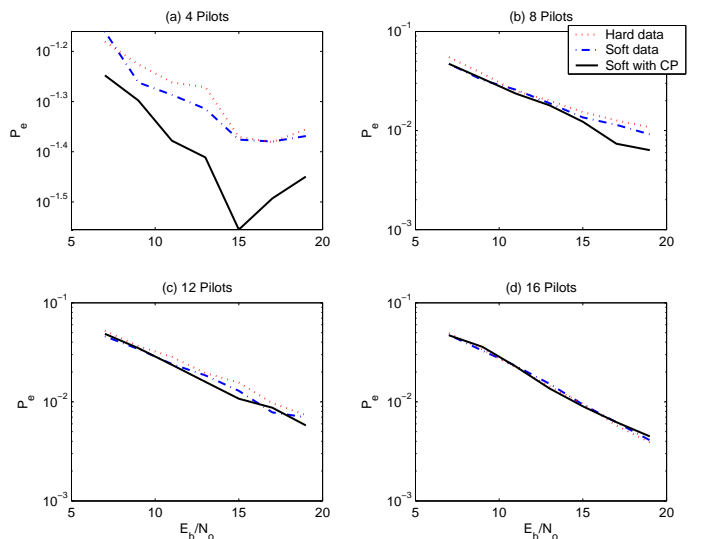

Fig. 3. For $f = .8$, we observe a similar behavior for the FB Kalman as for $f = .7$. For both cases, however, signal processing yields diminishing returns with increasing number of pilots.

B. Effect of increased signal processing

We next consider the effect of increased signal processing on the BER curves for FB-Kalman based receiver. Specifically, we implement this receiver 1) using the CP observation and the soft estimate of the input, 2) using the CP and the hard estimate of the input, and 3) using no CP observation and using the hard estimate of the input. We demonstrate that for $f = .7$ (Figure 2) and $f = .8$ (Figure 3). The two figures show that increasing the level of signal processing pays off producing better BER performance for different number of pilots. However, as we increase the number of pilots signal processing results in diminishing returns.

C. Effect of increasing the number of iterations

Figure 4 demonstrates the effect of increasing the number of EM iterations on the BER performance of the FB-Kalman receiver. We demonstrate that for $f = .9$ and for different number of pilots. As expected, the BER improves as we increase the number of iterations. Notice, however, that with increasing number of pilots the BER curves for 2 and 4 iterations become almost identical demonstrating again the tradeoff between increasing the number of pilots and increased signal processing.

D. Bench marking

Finally, we bench mark the BER performance of the Kalman and FB-Kalman receivers against receivers that have been suggested in literature and also against the known-channel case. Specifically, Figure 5 compares the BER performance of the following five receivers: 1) EM-based least-squares (LS) receiver (i.e. a receiver employing frequency correlation only), 2) the EM-based receiver proposed by Lu, Wang, and Li in [24] [8], 3) the EM-based Kalman receiver 4) the EM-based FB-Kalman receiver, and 4) a receiver with perfect channel knowledge. All receivers implement the same number

---

[8]This algorithm is similar to our (forward-only) Kalman-based algorithm in that it makes use of time and frequency correlation. The two algorithms are still different in that in applying the EM algorithm, this paper assumes that the data is the missing information while [24] assumes the channel to be the missing information.

of pilots. We test these receivers against the dynamically variant channel (1) with $f = .7$. Figure 5 demonstrates that the Kalman and the FB-Kalman receivers outperform the LS receiver and the receiver of [24]. This is especially the case for small number of pilots.

VI. CONCLUSION

In this paper, we considered the problem of semi-blind channel and data recovery in OFDM transmission over time-variant channels. Motivated by the EM approach, the algorithm boils down to a FB Kalman filter. It makes a collective use of the channel and data constraints in Table I. Specifically, the algorithm makes use of the finite alphabet constraints (in (25)–(26)), the data in its soft form (in (19)–(20)), pilots (in (21)–(22)), transmission precoding (in (27)–(28)), finite-delay spread (in that channel estimation is done in the time domain), and frequency- and time-correlation (in (1)). It is also straightforward to incorporate the effect of an outer code and of sparsity (see [30]). We also suggested a relaxed version of the algorithm (a forward-only Kalman) that is able to perform recovery
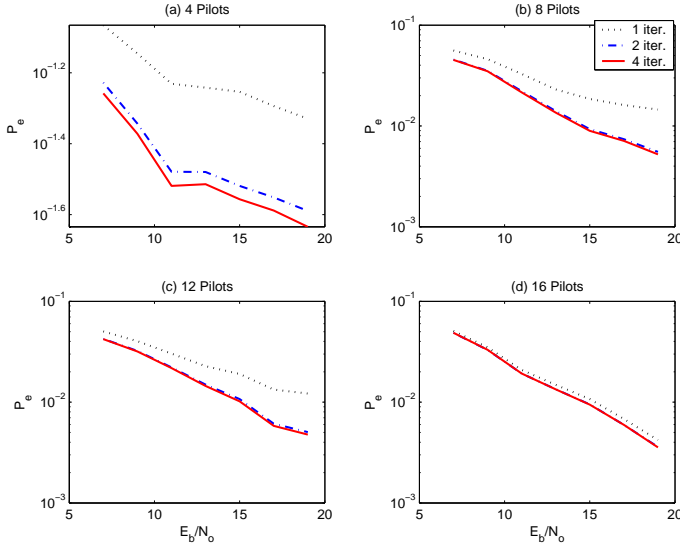
Fig. 4. Increasing the number of EM iterations improves the BER of the FB-Kalman receiver, but the value of these iterations results in diminishing returns
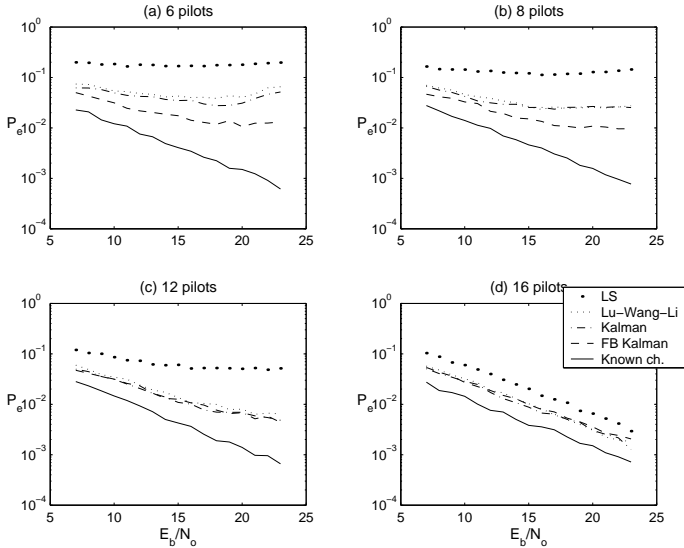

Fig. 5. BER curves comparing a receiver that employs the Kalman filter with one employing the FB-Kalman and another with perfect channel knowledge. The two Kalman receivers employ the same number of pilots with optimum placement.

with no latency and hence avoid the delay and storage shortcomings of the FB-Kalman. Our simulation show the favorable behavior of the FB-Kalman filter. Specifically, simulations demonstrate that increased signal processing always results in better BER behavior.

The Kalman receivers derived here apply to nonstationary channels as well in which the matrices F and G of the state-space model (1) vary with time. The algorithm has already been extended to incorporate the coding constraint on the data [33]. The paper assumes that these parameters are known perfectly at the receiver. However, the receiver can be generalized to estimate the state-space parameters and to be robust to uncertainties in these estimates (e.g., see [25]).

APPENDIX I
CHANNEL MODEL

Here, we show how to derive the state-space model (1) from the power delay profile and the Doppler spread (both of which can be

estimated (e.g., see [34] and [35]) and from the transmit filter. The channel $\underline{h}_i$ is the convolution of the physical channel $\underline{c}_i$ (which consists of $L + 1$ paths arriving at instants $\tau_0, \tau_1 \ldots, \tau_L$) and the receive filter r. Thus, we can write [15]

$$\underline{h}_i = R_i \underline{c}_i \tag{31}$$

where R_i is the receive filter matrix given by

$$R_i = \begin{bmatrix} r(-\tau_0) & r(-\tau_1) & \cdots & r(-\tau_L) \\ r(T - \tau_0) & r(T - \tau_1) & \cdots & r(T - \tau_L) \\ \vdots & \vdots & \vdots & \vdots \\ r(PT - \tau_0) & r(PT - \tau_2) & \cdots & r(PT - \tau_L) \end{bmatrix}$$

Due to the mobile nature of the channel, the physical channel taps $\underline{c}_i(k)$ are time-variant. According to the WSSUS model, the process $\underline{c}_i(k)$ is zero-mean wide-sense stationary complex Gaussian process with autocorrelation $E\left[\underline{c}_i(k)\underline{c}_{i'}(k')\right] = \mathcal{J}_0\left(\alpha_k|i - i'|\right)\delta_{kk'}$ where $\alpha_k = 2\pi f_c(k)(N + P)T$, T is the sampling (baud) rate, $f_c(k)$ is the Doppler frequency associated with the kth tap, and $\mathcal{J}_0$ denotes the zero-order Bessel function of the first kind. We can approximate the time-variant behavior of the tap $\underline{c}_i(k)$ by a first-order AR model (see [21], [22]) $\underline{c}_{i+1}(k) = \mathcal{J}_0\left(\alpha_k\right)\underline{c}_i(k) + \sqrt{(1 - \mathcal{J}_0^2(\alpha_k))E[|\underline{c}_0(k)|^2]}u_i(k)$. The factor $\sqrt{(1 - \mathcal{J}_0^2(\alpha_k))E[|\underline{c}_0(k)|^2]}$ ensures that the tap $\underline{c}_i(k)$ maintains the same power profile for all time. Collecting this recursion for all taps yields

$$\underline{c}_{i+1} = F_c \underline{c}_i + G_c \underline{u}_i \tag{32}$$

where

$$F_c = \operatorname{diag}\left(\ \mathcal{J}_0(\alpha_1), \quad \cdots, \quad \mathcal{J}_0(\alpha_{L+1})\ \right)$$

and

$$G_c = \operatorname{diag}\left(\ \sqrt{(1 - \mathcal{J}_0^2(\alpha_1))E[|\underline{c}_i(1)|^2]}, ..., \sqrt{(1 - \mathcal{J}_0^2(\alpha_{L+1}))E[|\underline{c}_i(L+1)|^2]}\ \right)$$

We can use this dynamical relationship along with (31) to derive a dynamical relationship for the impulse response $\underline{h}$. Specifically, multiplying both sides of (32) by R and noting that $R^\dagger R = I$ [9], we obtain $\underline{h}_{i+1} = F\underline{h}_i + G\underline{u}_i$, where $F = RF_c R^\dagger$ and $G = RG_c$.

REFERENCES

[1] Z. Shengli and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1402–1414, Aug. 2001.
[2] C. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Proc.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.
[3] G. Alrawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic prefix in channel estimation for coded OFDM systems," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 388–390, Jul. 2003.
[4] F. Sanzi and M. C. Necker, "Totlly blind APP channel estimation for mobile OFDM systems," *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 517–519, Nov. 2003.
[5] H. Bölcskei, R. W. Heath, and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Proc.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.

[9]For this to be true, the matrix R has to be tall which will be the case if the sampling rate is high enough so that the number of channel taps $P + 1$ is larger then the number of physical paths.

[6] R.W. Heath and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Proc.*, vol. 47, no. 3, pp. 848–856, Mar. 1999.

[7] G.B. Giannakis, "Filter banks for blind channel identification and equalization," *IEEE Signal Proc. Lett.*, vol. 4, no. 6, pp. 184–187, Jun. 1997.

[8] G. Leus and M. Moonen, "Semi-blind channel estimation for block transmissions with non-zero padding," in *Proc. Asilomar Conf. on Signals, Syst., and Computers*, Nov. 2001, pp. 762–766.

[9] Xiaowen Wang and K. J. Ray Liu, "Adaptive channel estimation using cyclic prefix in multicarrier modulation system," *IEEE Commun. Lett.*, vol. 3, no. 10, pp. 291–293, Oct. 1999.

[10] J.K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels (mobile radio)," *IEEE Trans. Vehicular Tech.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[11] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electr.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.

[12] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Vehicular Tech. Conf.*, May 1997, vol. 3, pp. 1693–1643.

[13] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 2001, pp. 2389–2392.

[14] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," in *Proc. IEEE Vehicular Tech. Conf.*, 1999, vol. 2, pp. 1131–1135.

[15] B. Yang, K. Ben Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[16] I. Kang, M. P. Fitz, and S. B. Gelfand, "Blind estimation of multipath channel parameters: a modal analysis approach," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1140–1150, Aug. 1999.

[17] M.C. Vanderveen, A.-J. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Proc.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.

[18] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[19] O. Edfors, M. Sandell, J. van de Beek, K. S. Wilson, and P. O. Brjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Signal Proc.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[20] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Proc.*, vol. 53, no. 2-3, pp. 211–229, Sept. 1996.

[21] C. Komninakis, C. Fragouli, A. Sayed, and R. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Proc.*, vol. 50, no. 5, pp. 1065–1076, May 2002.

[22] R.A. Iltis, "Joint estimation of PN code delay and multipath using the extended Kalman filter," *IEEE Trans. Commun.*, vol. 38, no. 10, pp. 1677–1685, Oct. 1990.

[23] A. Aghamohammadi, H. Meyr, and G. Ascheid, "Adaptive synchronization and channel parameter estimation using an extended Kalman filter," *IEEE Trans. Commun.*, vol. 37, no. 11, pp. 1212–1219, Nov. 1989.

[24] B. Lu, X. Wang, and Y. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 213–225, Apr. 2002.

[25] A. H. Sayed, "A framework for state-space estimation with uncertain models," *IEEE Trans. Auto. Control.*, vol. 46, no. 7, pp. 998–1013, Jul. 2001.

[26] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. Signal Proc.*, vol. 50, no. 5, pp. 1215–1228, May 2002.

[27] G. Alrawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic prefix in channel estimation for coded OFDM systems," in *Proc. IEEE GlobeCom*, Nov. 2002, pp. 1152–1156.

[28] A. Ghorokhov and J. Linnartz, "Robust OFDM receivers for dispersive time varying channels: equalization and channel acquisition," in *Proc. IEEE Int. Conf. Commun.*, 2002, pp. 470–474.

[29] Y. Xie and C. N. Georghiades, "Two EM-type channel estimation algorithms for OFDM with transmitter diversity," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 106–115, Jan. 2003.

[30] T. Y. Al-Naffouri, Adaptive algorithms for wireless channel estimation, Department of Electrical Engineering, Stanford University, Jan. 2005.

[31] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, Prentice Hall, 2000.

[32] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj, "EM algorithm for channel estimation and equalization in OFDM," in *Proc. IEEE GlobeCom*, 2002, pp. 589–593.

[33] G. Alrawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, "An iterative receiver for coded OFDM systems over time-varying wireless channels," in *Proc. IEEE Int. Conf. Commun.*, May 2003, pp. 3371–3376.

[34] T. Cui C. Tellambura, "Power delay profile and noise variance estimation for OFDM," *IEEE Commun. Lett.*, vol. 49, no. 1, pp. 25–27, Jan. 2006.

[35] Z. Li Jueping Cai; Wentao Song, "Doppler spread estimation for mobile OFDM systems in Rayleigh fading channels," *IEEE Trans. Consumer Electr.*, vol. 49, no. 4, pp. 973–977, Nov. 2006.