AN EM-BASED OFDM RECEIVER FOR TIME-VARIANT CHANNELS

TAREQ Y. AL-NAFFOURI[1], AHMAD BAHAI[1,2], *and* AROGYASWAMI PAULRAJ[1]

[1] Information Systems Lab
Electrical Engineering Department,
Stanford University, CA 94305

[2] Wireless Systems
National Semiconductor
Santa Clara, CA 94538

ABSTRACT

OFDM modulation combines the advantages of high achievable rates and relatively easy implementation. However, for proper recovery of the input, the receiver needs accurate channel information. In this paper, we propose an expectation-maximization (EM) algorithm for joint channel and data recovery. The algorithm makes use of the rich structure of the underlying communication problem- a structure induced by the data and channel constraints. These constraints include pilots, the cyclic prefix (CP), and the finite alphabet constraints on the data, and sparsity, finite delay spread, and the statistical properties of the channel (time and frequency correlation) . Channel identification and equalization is performed optimally and recovery is achieved within the same OFDM symbol using an EM based Kalman filter.

1. 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 including digital audio and video broadcasting (DAB and DVB) in Europe and high speed transmission over digital subscriber lines (DSL) in the United States.

Many techniques have been proposed in literature to estimate and equalize channels for OFDM transmission (see, e.g., [1] and the references therein). These techniques rely on some constraints on the channel or data to perform channel and/or data recovery. Thus, pilots were employed in [2], the cyclic prefix (CP) in [3], coding in [4], frequency correlation in [5], and sparsity in [6]. The aim of this paper is to make a *collective use* of channel constraints (sparsity, finite delay spread, frequency correlation, and time correlation) and data constraints (pilots, finite alphabet constraints, and cyclic prefix) to achieve optimal channel and data recovery within the same symbol and with minimal overhead.

The paper is organized as follows. After introducing our notation in the next section, we perform a careful study in section 3 of the elements of the OFDM transmission. In section 4, we show how to perform optimal MMSE recovery of the input while the dual task of channel recovery is studied in the subsequent section. We end the paper with simulations and some conclusions.

The work of T. Y. Al-Naffouri was partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia.

2. NOTATION

We use regular small-case letters to denote scalars and small-case boldface letters to denote vectors. Calligraphic notation (e.g., $\boldsymbol{\mathcal{X}}$) is used to denote vectors in the frequency domain while uppercase boldface letters are reserved for matrices. When these variables become a function of time, the time index i appears as a subscript. A hat over a variable indicates an estimate of that variable (e.g., as in $\hat{\boldsymbol{\mathcal{X}}}_i$ and $\hat{\underline{\boldsymbol{h}}}$).

Now consider a length-N vector $\boldsymbol{x}_i$. We deal with three derivatives associated with this vector. The first two are obtained by partitioning $\boldsymbol{x}_i$ into a lower part $\underline{\boldsymbol{x}}_i$ (usually known as the cyclic prefix) and an upper part $\tilde{\boldsymbol{x}}_i$. The third derivative, $\overline{\boldsymbol{x}}_i$, is created by concatenating $\boldsymbol{x}_i$ with a copy of its prefix $\underline{\boldsymbol{x}}_i$. The relationship among these variables is summarized by

$$\overline{\boldsymbol{x}}_i = \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \boldsymbol{x}_i \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} \tag{1}$$

This notational convention will be extended to matrices as well. Thus, a matrix $\boldsymbol{Q}$ having N rows will have the natural partitioning

$$\boldsymbol{Q} = \begin{bmatrix} \tilde{\boldsymbol{Q}}_{N-P} \\ \underline{\boldsymbol{Q}}_P \end{bmatrix} \tag{2}$$

where the subscripts stand for the number of rows in each submatrix. Alternatively, the rows belonging to each submtarix could be distinguished by index sets. For example, with $\tilde{I} = \{1, \cdots, N-P\}$ and $\underline{I} = \{N-P+1, \cdots, N\}$, we can rewrite (2) as

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{\tilde{I}} \\ \boldsymbol{Q}_{\underline{I}} \end{bmatrix} \tag{3}$$

3. ESSENTIAL ELEMENTS OF OFDM TRANSMISSION

Consider the sequence $\{\mathcal{X}(i)\}$ that we wish to transmit. Data are collected and transmitted in symbols $\boldsymbol{\mathcal{X}}_i$ of length N. In an OFDM system, the symbol vector $\boldsymbol{\mathcal{X}}_i$ undergoes an IDFT operation to produce the transform vector $\boldsymbol{x}_i$

$$\boldsymbol{x}_i = \frac{1}{\sqrt{N}} \boldsymbol{Q} \boldsymbol{\mathcal{X}}_i \tag{4}$$

where $\boldsymbol{Q}$ is the unitary DFT matrix $\boldsymbol{Q} = \left[e^{j\frac{2\pi}{N}lm}\right]$. This induces the underlying sequence $\{x(i)\}$. If this sequence is transmitted through a nonideal channel $\underline{\boldsymbol{h}}$, which we take as FIR of maximum length $P+1$, it will be subject to intersymbol interference (ISI). To go around this, a guard band is inserted between any consecutive symbols, $\boldsymbol{x}_{i-1}$ and $\boldsymbol{x}_i$. In particular, to each symbol, we append a cyclic prefix of length P as done in (1). Thus, instead of transmitting $\boldsymbol{x}_i$, we transmit $\overline{\boldsymbol{x}}_i$ defined in (1). The concatenation of these symbols in turn produces the underlying sequence $\{\overline{x}(i)\}$.

When passed through the channel $\underline{\boldsymbol{h}}$, the sequence $\{\overline{x}(i)\}$ produces the output sequence $\{\overline{y}(i)\}$. Similarly, we split the output into symbols of length $M = N + P$, and further split each symbol into a length-N symbol $\boldsymbol{y}_i$ and a prefix associated with it $\underline{\boldsymbol{y}}_i$, i.e.

$$\overline{\boldsymbol{y}}_i = \begin{bmatrix} \underline{\boldsymbol{y}}_i \\ \boldsymbol{y}_i \end{bmatrix} \tag{5}$$

The prefix $\underline{\boldsymbol{y}}_i$ absorbs all ISI that takes place between the adjacent symbols $\overline{\boldsymbol{x}}_{i-1}$ and $\overline{\boldsymbol{x}}_i$. The remaining part $\boldsymbol{y}_i$ of the symbol depends on the ith input symbol only. This can be seen from the input/output relationship

$$\begin{bmatrix} \boldsymbol{y}_{i-1} \\ \underline{\boldsymbol{y}}_i \\ \boldsymbol{y}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{n}_{i-1} \\ \underline{\boldsymbol{n}}_i \\ \boldsymbol{n}_i \end{bmatrix} + \tag{6}$$

$$\begin{bmatrix} \multicolumn{2}{c}{\overline{\boldsymbol{H}}} & \boldsymbol{O}_{N\times P} & \boldsymbol{O}_{N\times N} \\ \boldsymbol{O}_{P\times N} & \underline{\boldsymbol{H}}_{\mathrm{U}} & \underline{\boldsymbol{H}}_{\mathrm{L}} & \boldsymbol{O}_{P\times N} \\ \boldsymbol{O}_{N\times N} & \boldsymbol{O}_{N\times P} & \multicolumn{2}{c}{\overline{\boldsymbol{H}}} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_{i-1} \\ \tilde{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix}$$

where $\boldsymbol{n}$ is the output noise which we take to be white Gaussian with variance σ_n^2. The matrices $\overline{\boldsymbol{H}}$, $\underline{\boldsymbol{H}}_{\mathrm{U}}$, and $\underline{\boldsymbol{H}}_{\mathrm{L}}$ are convolution (Toeplitz) matrices of proper sizes created from the vector $\underline{\boldsymbol{h}}$. Because of the redundancy in the input, the convolution in (6) can be decomposed into two distinct constituent convolution operations or subchannels as we discuss now.

3.1. Circular Convolution (Subchannel)

From (6), we can parse the subsystem of equations

$$\boldsymbol{y}_i = \overline{\boldsymbol{H}} \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} = \overline{\boldsymbol{H}}\, \overline{\boldsymbol{x}}_i + \boldsymbol{n}_i \tag{7}$$

This shows that $\boldsymbol{y}_i$ is created solely from $\overline{\boldsymbol{x}}_i$ through convolution. Moreover, the existence of a cyclic prefix in $\overline{\boldsymbol{x}}_i$ renders this convolution cyclic, and we can write

$$\boxed{\boldsymbol{y}_i = \boldsymbol{h} \circledast \boldsymbol{x}_i + \boldsymbol{n}_i} \tag{8}$$

where $\boldsymbol{h}$ is a length-N zero-padded version of $\underline{\boldsymbol{h}}$

$$\boldsymbol{h} = \begin{bmatrix} \underline{\boldsymbol{h}} \\ \boldsymbol{O}_{(N-P-1)\times 1} \end{bmatrix} \tag{9}$$

In the frequency domain, the cyclic convolution (8) reduces to the element-by-element operation

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \mathrm{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{N}}_i} \tag{10}$$

where $\boldsymbol{\mathcal{H}}$, $\boldsymbol{\mathcal{X}}_i$, $\boldsymbol{\mathcal{Y}}_i$, and $\boldsymbol{\mathcal{N}}_i$ are the DFT's of $\boldsymbol{h}$, $\boldsymbol{x}_i$, $\boldsymbol{y}_i$, and $\boldsymbol{n}_i$, respectively

$$\boldsymbol{h} = \boldsymbol{Q}\boldsymbol{\mathcal{H}},\ \boldsymbol{x}_i = \frac{1}{\sqrt{N}}\boldsymbol{Q}\boldsymbol{\mathcal{X}}_i,\ \boldsymbol{y}_i = \frac{1}{\sqrt{N}}\boldsymbol{Q}\boldsymbol{\mathcal{Y}}_i,\ \boldsymbol{n}_i = \frac{1}{\sqrt{N}}\boldsymbol{Q}\boldsymbol{\mathcal{N}}_i \tag{11}$$

It is easy to show that $\underline{\boldsymbol{h}}$ and $\boldsymbol{\mathcal{H}}$ are related by the partial DFT relationship

$$\boldsymbol{\mathcal{H}} = \tilde{\boldsymbol{Q}}^*_{P+1}\underline{\boldsymbol{h}} \tag{12}$$

This can be used to rewrite (10) in the time-frequency form

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \mathrm{diag}(\boldsymbol{\mathcal{X}}_i)\tilde{\boldsymbol{Q}}^*_{P+1}\underline{\boldsymbol{h}} + \boldsymbol{\mathcal{N}}_i} \tag{13}$$

Let I_P denote the index set of the pilot symbols, then we can form the following pilot/output relationship

$$\boxed{\boldsymbol{\mathcal{Y}}_{iI_p} = \mathrm{diag}(\boldsymbol{\mathcal{X}}_i)_{I_p}\tilde{\boldsymbol{Q}}^*_{P+1}\underline{\boldsymbol{h}} + \boldsymbol{\mathcal{N}}_{iI_p}} \tag{14}$$

3.2. Linear Convolution (Subchannel)

From (6), we can also extract a relationship between the input and output prefixes. This can be used to show that the input prefix sequence $\{\underline{x}(i)\}$ is related to the output prefix sequence $\{\underline{y}(i)\}$ through linear convolution with the channel $\underline{h}$, i.e.

$$\boxed{\underline{y}(i) = \underline{h}(i) * \underline{x}(i) + \underline{n}(i)} \tag{15}$$

We could write (15) in matrix form as

$$\boxed{\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i\underline{\boldsymbol{h}} + \underline{\boldsymbol{n}}_i} \tag{16}$$

where $\underline{\boldsymbol{X}}_i$ is a Toeplitz matrix defined by

$$\begin{aligned} \underline{\boldsymbol{X}}_i &= \begin{bmatrix} \underline{x}_i(0) & \underline{x}_{i-1}(P-1) & \cdots & \underline{x}_{i-1}(0) \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & \underline{x}_{i-1}(1) \\ \vdots & \cdots & \ddots & \vdots \\ \underline{x}_i(P-1) & \underline{x}_i(P-2) & \cdots & \underline{x}_{i-1}(P-1) \end{bmatrix}. \\ &= \boldsymbol{X}_{i\mathrm{L}} + \boldsymbol{X}_{i\mathrm{U}} \end{aligned} \tag{17}$$

and where $\boldsymbol{X}_{i\mathrm{L}}$ ($\boldsymbol{X}_{i\mathrm{U}}$) is the lower (upper) triangular part of $\boldsymbol{X}_i$ formed from the cyclic prefix $\underline{\boldsymbol{x}}_i$ ($\underline{\boldsymbol{x}}_{i-1}$).

3.3. Total Channel

The sequence $\{\overline{y}(i)\}$ at the channel output is related naturally to the input sequence $\{\overline{x}(i)\}$ through linear convolution with the channel

$$\overline{y}(i) = h(i) * \overline{x}(i) + \overline{n}(i) \tag{18}$$

We can express this in matrix form by concatenating (13) and (16) to get

$$\boxed{\overline{\boldsymbol{\mathcal{Y}}}_i = \overline{\boldsymbol{X}}_i\underline{\boldsymbol{h}} + \overline{\boldsymbol{\mathcal{N}}}_i} \tag{19}$$

where, in line with the notational convention (1),

$$\overline{\boldsymbol{X}}_i \triangleq \begin{bmatrix} \operatorname{diag}(\boldsymbol{\mathcal{X}}_i)\tilde{\boldsymbol{Q}}^*_{P+1} \\ \underline{\boldsymbol{X}}_i \end{bmatrix}, \quad \overline{\boldsymbol{\mathcal{Y}}}_i = \begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \underline{\boldsymbol{y}}_i \end{bmatrix}, \quad \overline{\boldsymbol{\mathcal{N}}}_i = \begin{bmatrix} \boldsymbol{\mathcal{N}}_i \\ \underline{\boldsymbol{n}}_i \end{bmatrix} \quad (20)$$

3.4. Sparsity and Time Variance

Wireless channels are usually sparse. Moreover, the taps of a typical channel fade in value at a much faster rate than in location. We could thus safely assume the taps to be known in location but not in value [6]. So let I_c denote the index set of the active taps in $\underline{\boldsymbol{h}}$, the various input/output relationships can be rewritten in terms of $\underline{\boldsymbol{h}}_{I_c}$. In particular, the relationships (13) and (19) take the form

$$\boldsymbol{\mathcal{Y}}_{iI_p} = \operatorname{diag}(\boldsymbol{\mathcal{X}}_i)_{I_p}\tilde{\boldsymbol{Q}}^*_{I_c}\underline{\boldsymbol{h}}_{I_c} + \boldsymbol{\mathcal{N}}_{iI_p} \quad (21)$$

$$\overline{\boldsymbol{\mathcal{Y}}}_i = \overline{\boldsymbol{X}}_i \boldsymbol{I}^*_{I_c}\underline{\boldsymbol{h}}_{I_c} + \overline{\boldsymbol{\mathcal{N}}}_i \quad (22)$$

Here, as per our notation, the matrix $\boldsymbol{I}_{I_c}$ is created from the size $P+1$ identity matrix $\boldsymbol{I}$ by keeping the rows indexed by the set I_c.

To avoid intercarrier interference, we assume that the channel remains time invariant over any one OFDM symbol and the associated cyclic prefix. However, we also assume that from one symbol to the next, the active taps follow a state-space model

$$\underline{\boldsymbol{h}}_{i+1\,I_c} = \boldsymbol{F}\underline{\boldsymbol{h}}_{i\,I_c} + \boldsymbol{G}\boldsymbol{u}_i \quad (23)$$

This model allows for maximum time variance without introducing intercarrier interference.

In the next two sections, we show how to perform optimal data detection given the channel estimate and how to perform optimal channel estimation given some information about the data.

4. MEAN-SQUARE ESTIMATION OF DATA

Data is best estimated from the circular channel. The decoupled nature of this channel makes it possible to perform optimal data recovery with low complexity. Thus, given the channel gain in the lth bin, $\mathcal{H}_i(l)$, the MMSE estimate of $\mathcal{X}_i(l)$ is given by

$$\hat{\mathcal{X}}_i(l) \triangleq E\left[\mathcal{X}_i(l)|\mathcal{X}_i(l)\right] = \frac{\sum_{j=1}^{j=|A|} A_j e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}_i(l)A_j|^2}{\sigma_n^2}}}{\sum_{j=1}^{|A|} e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}_i(l)A_j|^2}{\sigma_n^2}}} \quad (24)$$

where $A = \{A_1, A_2, \ldots, A_{|A|}\}$ is the input alphabet. Since $\boldsymbol{x}_i$ and $\boldsymbol{\mathcal{X}}_i$ are linearly related by (11), so are their MMSE estimates and we can write

$$\hat{\boldsymbol{x}}_i = \frac{1}{N}\boldsymbol{Q}\hat{\boldsymbol{\mathcal{X}}}_i \qquad \hat{\underline{\boldsymbol{x}}}_i = \frac{1}{N}\underline{\boldsymbol{Q}}_P\hat{\boldsymbol{\mathcal{X}}}_i \quad (25)$$

We can as easily calculate the conditional second-order moment of the input which will be useful for channel estimation further ahead

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

4.1. First and Second Moments of $\overline{\boldsymbol{X}}_i$

For the purpose of channel estimation, we need to calculate the first and second order moments of $\overline{\boldsymbol{X}}_i$. From (17) and (20), we can express the first moment of $\overline{\boldsymbol{X}}_i$ as

$$E\left[\overline{\boldsymbol{X}}_i\right] \triangleq \begin{bmatrix} E[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i)]\tilde{\boldsymbol{Q}}^*_{P+1} \\ E\left[\underline{\boldsymbol{X}}_{i\mathrm{L}} + \underline{\boldsymbol{X}}_{i\mathrm{U}}\right] \end{bmatrix} \quad (27)$$

$$= \begin{bmatrix} \operatorname{diag}(\hat{\boldsymbol{\mathcal{X}}}_i)\tilde{\boldsymbol{Q}}^*_{P+1} \\ \hat{\underline{\boldsymbol{X}}}_{i\mathrm{L}} + \hat{\underline{\boldsymbol{X}}}_{i\mathrm{U}} \end{bmatrix} \quad (28)$$

where $\hat{\underline{\boldsymbol{X}}}_{i\mathrm{L}}$ ($\hat{\underline{\boldsymbol{X}}}_{i\mathrm{U}}$) is constructed from $\hat{\underline{\boldsymbol{x}}}_i$ ($\hat{\underline{\boldsymbol{x}}}_{i-1}$) just as $\underline{\boldsymbol{X}}_{i\mathrm{L}}$ ($\underline{\boldsymbol{X}}_{i\mathrm{U}}$) is constructed from $\underline{\boldsymbol{x}}_i$ ($\underline{\boldsymbol{x}}_{i-1}$).

The covariance of $\overline{\boldsymbol{X}}_i^*$ is not as easy to calculate.[1] From the defining expressions (17) and (20), we deduce

$$\operatorname{Cov}\left[\overline{\boldsymbol{X}}_i^*\right] = \tilde{\boldsymbol{Q}}_{P+1}\operatorname{Cov}\left[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i^*)\right]\tilde{\boldsymbol{Q}}^*_{P+1} + \operatorname{Cov}\left[\underline{\boldsymbol{X}}_i^*\right]$$
$$= \tilde{\boldsymbol{Q}}_{P+1}\operatorname{Cov}\left[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i^*)\right]\tilde{\boldsymbol{Q}}^*_{P+1} + \operatorname{Cov}\left[\underline{\boldsymbol{X}}_{i\mathrm{L}}^*\right] + \operatorname{Cov}\left[\underline{\boldsymbol{X}}_{i\mathrm{U}}^*\right]$$

Now the covariance $\operatorname{Cov}\left[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i^*)\right]$ is a diagonal matrix whose diagonal elements are simply the variances

$$\operatorname{Cov}[\mathcal{X}_i(l)] = E[|\mathcal{X}_i(l)|^2] - |\hat{\mathcal{X}}_i(l)|^2 \quad (29)$$

which can be calculated from (24) and (26). The elements of the covariance matrix $C_{\mathrm{L}} \triangleq \operatorname{Cov}\left[\underline{\boldsymbol{X}}_{i\mathrm{L}}^*\right]$ are calculated recursively from

$$C_{\mathrm{L}}(j,k) = C_{\mathrm{L}}(j+1,k+1) + \underbrace{E[\underline{x}_i^*(P-j)\underline{x}_i(P-k)] - \hat{\underline{x}}_i^*(P-j)\hat{\underline{x}}_i(P-k)}_{\text{covariance evaluated in (33)}} \quad (30)$$

The recursion is run (backward) for $j,k = 1,2,\cdots,P$ starting from the boundary conditions

$$C_{\mathrm{L}}(P+1,l) = C_{\mathrm{L}}(l,P+1) = 0, \quad l = 1,2,\cdots,P \quad (31)$$

The covariances that appear in (30) are the entries of the covariance $\operatorname{Cov}[\underline{\boldsymbol{x}}_i]$ and are collectively calculated from

$$\operatorname{Cov}[\underline{\boldsymbol{x}}_i] = \frac{1}{N}\underline{\boldsymbol{Q}}_P\operatorname{Cov}[\boldsymbol{\mathcal{X}}_i]\underline{\boldsymbol{Q}}_P^* \quad (32)$$

$$= \frac{1}{N}\underline{\boldsymbol{Q}}_P\operatorname{Cov}[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i)]\underline{\boldsymbol{Q}}_P^* \quad (33)$$

where (32) follows from the partial IDFT relationship $\underline{\boldsymbol{x}}_i = (1/\sqrt{N})\underline{\boldsymbol{Q}}_P\boldsymbol{\mathcal{X}}_i$, and where the diagonal elements of the covariance $\operatorname{Cov}[\operatorname{diag}(\boldsymbol{\mathcal{X}}_i)]$ of (33) have already been calculated in (29). Similarly, we can show that the covariance $C_{\mathrm{U}} \triangleq \operatorname{Cov}\left[\underline{\boldsymbol{X}}_{i\mathrm{U}}^*\right]$ satisfies the recursion

$$C_{\mathrm{U}}(j+1,k+1) = C_{\mathrm{U}}(j,k) + E[\underline{x}_{i-1}^*(P-j)\underline{x}_{i-1}(P-k)] - \hat{\underline{x}}_{i-1}^*(P-j)\hat{\underline{x}}_{i-1}(P-k)$$

The recursion is kick-started from the initial conditions

$$C_{\mathrm{U}}(1,l) = C_{\mathrm{U}}(l,1) = 0, \quad l = 1,2,\cdots,P \quad (34)$$

[1] The covariance of a matrix $\boldsymbol{A}$ is defined as $\operatorname{Cov}[\boldsymbol{A}] \triangleq E[\boldsymbol{A}\boldsymbol{A}^*] - E[\boldsymbol{A}]E[\boldsymbol{A}^*]$.

5. CHANNEL ESTIMATION

Our aim in this paper is to perform joint channel and data recovery. Given the input and output data, the IR $\underline{\boldsymbol{h}}_i$ is optimally estimated by maximizing the log-likelihood function $\log\, p(\underline{\boldsymbol{h}}_i|\overline{\boldsymbol{\mathcal{Y}}}_i, \overline{\boldsymbol{X}}_i)$ which amounts to solving the least-squares problem

$$\hat{\underline{\boldsymbol{h}}}_i = \arg\max_{\underline{\boldsymbol{h}}_i} \|\overline{\boldsymbol{\mathcal{Y}}}_i - \overline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} \tag{35}$$

Since the input is not available, we average the log likelihood over the input $\overline{\boldsymbol{X}}_i$ resulting in the regularized least-squares problem

$$\begin{aligned}\hat{\underline{\boldsymbol{h}}}_i &= \arg\max_{\underline{\boldsymbol{h}}_i} E_{\mathcal{X}_i} \|\overline{\boldsymbol{\mathcal{Y}}}_i - \overline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} &(36)\\ &= \arg\max_{\underline{\boldsymbol{h}}_i} \|\overline{\boldsymbol{\mathcal{Y}}}_i - E\left[\overline{\boldsymbol{X}}_i\right] \underline{\boldsymbol{h}}_i\|^2_{\frac{1}{\sigma_n^2}} + \|\underline{\boldsymbol{h}}\|^2_{\frac{1}{\sigma_n^2}\mathrm{Cov}[\overline{\boldsymbol{X}}_i^*]} &(37)\end{aligned}$$

Thus, starting from some initial channel estimate, we alternate between data detection using (24) and (26) and channel estimation from (36) in an iterative process known as the <u>E</u>xpectation <u>M</u>aximization (EM) algorithm. With a priori frequency-correlation information ($\boldsymbol{\Pi}_0$), one can add the regularizing term $\|\underline{\boldsymbol{h}}\|^2_{\Pi_0^{-1}}$ to the objective functions (35) and (36) to enhance the quality of the estimates. More details and elaborate discussions can be found in [8]

In the present case, the impulse response, or the active part of it, $\underline{\boldsymbol{h}}_{i I_c}$, is related to $\underline{\boldsymbol{h}}_{i-1 I_c}$ through the state-space model

$$\underline{\boldsymbol{h}}_{i+1 I_c} = \boldsymbol{F}\underline{\boldsymbol{h}}_{i I_c} + \boldsymbol{G}\boldsymbol{u}_i \tag{38}$$

This fact together with the input/output equation

$$\overline{\boldsymbol{\mathcal{Y}}}_i = \overline{\boldsymbol{X}}_i \boldsymbol{I}_{I_c}^* \underline{\boldsymbol{h}}_{i I_c} + \overline{\boldsymbol{\mathcal{N}}}_i \tag{39}$$

can be used to optimally estimate $\underline{\boldsymbol{h}}_{i I_c}$. Specifically, when the input is fully available, we employ the Kalman filter [7]

$$\begin{aligned}\boldsymbol{P}_{i|i-1} &= \begin{cases}\boldsymbol{\Pi}_0 & i=0\\ \boldsymbol{F}\boldsymbol{P}_{i-1}\boldsymbol{F}^* + \sigma_u^2 \boldsymbol{G}\boldsymbol{G}^* & i \geq 1\end{cases} &(40)\\ \boldsymbol{R}_{e,i} &= \sigma_n^2 \boldsymbol{I}_{N+P} + \overline{\boldsymbol{X}}_i \boldsymbol{I}_{I_c}^* \boldsymbol{P}_{i|i-1} \boldsymbol{I}_{I_c} \overline{\boldsymbol{X}}_i^* &(41)\\ \boldsymbol{K}_{f,i} &= \boldsymbol{P}_{i|i-1} \boldsymbol{I}_{I_c} \overline{\boldsymbol{X}}_i^* \boldsymbol{R}_{e,i}^{-1} &(42)\\ \hat{\underline{\boldsymbol{h}}}_{i I_c} &= \left(\boldsymbol{I} - \boldsymbol{K}_{f,i}\overline{\boldsymbol{X}}_i \boldsymbol{I}_{I_c}^*\right)\boldsymbol{F}\hat{\underline{\boldsymbol{h}}}_{i-1 I_c} + \boldsymbol{K}_{f,i}\overline{\boldsymbol{\mathcal{Y}}}_i &(43)\\ \boldsymbol{P}_i &= \boldsymbol{P}_{i|i-1} - \boldsymbol{K}_{f,i}\boldsymbol{R}_{e,i}\boldsymbol{K}_{f,i}^* &(44)\end{aligned}$$

where $\hat{\underline{\boldsymbol{h}}}_{-1 I_c} = 0$. In practice, however, the input is only partially known and the Kalman filter has to be modified accordingly; we distinguish between three possibilities

5.1. Input Partially Known (Pilots)

Here, we replace the input/output equation (39) with the pilot/output relationship

$$\boldsymbol{\mathcal{Y}}_{i I_p} = \mathrm{diag}(\boldsymbol{\mathcal{X}}_i)_{I_p} \bar{\boldsymbol{Q}}_{I_c}^* \underline{\boldsymbol{h}}_{i I_c} + \boldsymbol{\mathcal{N}}_{i I_p} \tag{45}$$

As a result, the Kalman filter (40)-(44) applies with the following change of variables

$$\overline{\boldsymbol{X}}_i \boldsymbol{I}_{I_c}^* \to \mathrm{diag}\left(\boldsymbol{\mathcal{X}}_i\right)_{I_p} \boldsymbol{Q}_{I_c}^* \quad \overline{\boldsymbol{\mathcal{Y}}}_i \to \boldsymbol{\mathcal{Y}}_{i I_p} \quad \boldsymbol{I}_{N+P} \to \boldsymbol{I}_{|I_p|} \tag{46}$$

where I_p is the index set of the pilot locations and $|I_p|$ is their number.

5.2. Input in Detected Form

Here, the detected input is available to us in the form of an MMSE estimate and an estimate of the input energy (see (24) and (26)). In this case, we can derive an EM-based Kalman filter that is similar to (40)-(44) with the following change of variables (see [9] for details)

$$\overline{\boldsymbol{X}}_i \to \begin{bmatrix} E\left[\overline{\boldsymbol{X}}_i\right] \\ \mathrm{Cov}\left[\overline{\boldsymbol{X}}_i^*\right]^{1/2}\end{bmatrix}, \overline{\boldsymbol{\mathcal{Y}}}_i \to \begin{bmatrix}\overline{\boldsymbol{\mathcal{Y}}}_i \\ 0_{P\times 1}\end{bmatrix}, \boldsymbol{I}_{N+P} \to \boldsymbol{I}_{N+2P} \tag{47}$$

5.3. No Input Information

For the first symbol, we need some pilot information to kick-start the estimation process. For the subsequent symbols, however, we can do away with pilots and rely on the previous channel estimate to initialize the estimation process for the current OFDM symbol. We can thus assume that we have no input/output equation and accordingly, we modify Kalman filter (40)-(44) by the change of variables

$$\overline{\boldsymbol{X}}_i \boldsymbol{I}_{I_c}^* \to \boldsymbol{O} \tag{48}$$

5.4. Summary of the Algorithm

In what follows, we summarize the channel estimation and equalization algorithm. Consider a sequence of OFDM symbols passing through a time-variant channel according to (23). The first symbol is assigned pilots in the frequency bins I_p while the subsequent symbols might not have any. For each symbol, we perform two operations

<u>Initialization:</u> Obtain the initial estimate using the Kalman filter (40)-(44) together with the change of variables

$$\begin{cases}(46) & \text{In the presense of pilots}\\ (48) & \text{with no pilots (applies for } i \geq 1 \text{ only)}\end{cases}$$

<u>Iterations:</u>

- Obtain the frequency response $\hat{\boldsymbol{\mathcal{H}}}_i$ using (12)
- Obtain the estimate of the first and second moments of the input (using (24) and (26)) and of the matrix $\overline{\boldsymbol{X}}_i$ (using (27)-(34))
- Estimate $\hat{\underline{\boldsymbol{h}}}_{i I_c}$ using the Kalman filter (40)-(44) together with the substitutions (47)
- Repeat

6. SIMULATIONS

We consider an OFDM system that transmits a sequence of 5 symbols each with 128 carriers and a cyclic prefix of length $P = 15$. The input data is 16 QAM transmitted at an SNR of 14 dB. The first symbol contains 16 pilots that are equally spaced while the subsequent 4 symbols carry none. The channel IR consists of 16 complex taps (the maximum length possible). The initial IR $\underline{h}_0$ has an exponential delay profile $E[|\underline{h}_0(k)|^2] = e^{-0.2k}$. For $i \geq 0$, $\underline{h}_i$ is generated according to the state-space model (38) with $F = .99I$ and with G diagonal such that $G(i,i) = \sqrt{(1-(0.99)^2)E[|\underline{h}_0(k)|^2]}$. The state noise u_i is iid with unit variance.

For each symbol, we iterate the EM-algorithm a number times. The whole simulation is run for 50 runs. Fig. 1 shows a typical learning curve for the channel variation over the 5 symbols. For each OFDM symbol, the MSE is largest at the start of each symbol and decreases after a few iterations although pilots are employed for the first symbol only.

We next measure the effect of the cyclic prefix and the soft estimates of the input. We thus compare the EM algorithm with a version that does not use the cyclic prefix observation and with one that makes use of the hard input estimate as opposed to the soft estimate as mandated by the expectation step. We compare these three scenarios by plotting the aggregate MSE $\sum_{i=1}^{5} \|\hat{\underline{h}}_i - \underline{h}_i\|^2$ vs. iteration in Fig. 2. The difference in gain demonstrates the gain provided by more sophisticated signal processing.

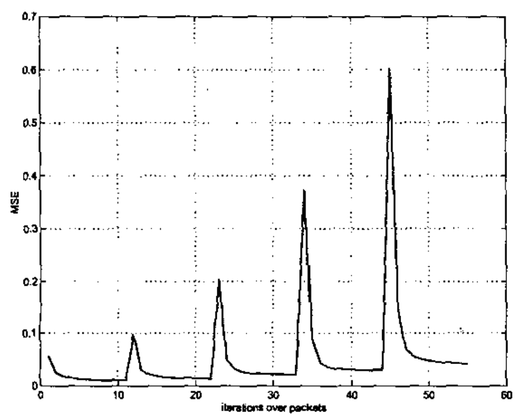

Figure 1: Learning Curve along 5 OFDM symbols

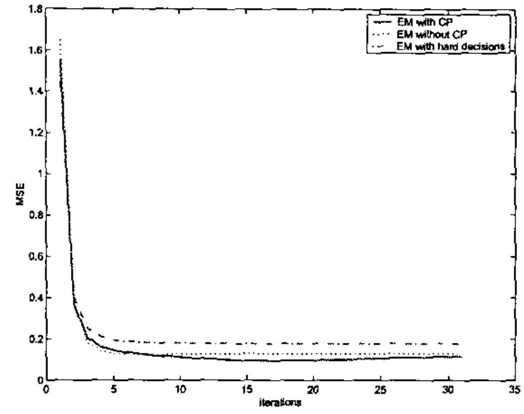

Figure 2: Learning curves showing the contributions of CP and soft input

7. CONCLUSION

In this paper, we designed an OFDM receiver for time-variant channels. Specifically, the receiver uses the pilots to kick start channel estimation and subsequently iterates between that and data recovery. In doing so, the receiver utilizes the data constraints (which includes the cyclic prefix, pilots, and the finite alphabet nature of the data) and employs the data estimates in soft format. The receiver also makes use of the various constraints on the channel (which includes sparsity and finite delay spread information as well as time and frequency correlation). Channel estimation boils down to an EM-based Kalman filter and is always done with zero latency. If increased latency is not an issue, one can enhance channel estimation by employing future as well as past symbols using an EM-based forward backward Kalman filter.

8. REFERENCES

[1] W. Zhendao and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Magazine,* vol. 17, no. 3, pp. 29–48, May 2000.

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

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

[4] V. Mignone and A. Morello, "CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. Commun.*, vol. 44, no. 9 , pp. 1144 -1151, Sep. 1996.

[5] 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.

[6] B. Yang, K. Ben Letaief, R. Cheng, 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.

[7] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation,* Prentice Hall, NJ, 2000.

[8] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj, "Semi-blind channel identification and equalization in OFDM: An expectation-maximimzation approach," *to appear in VTC 2002,* Vancouvor, Canada.

[9] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj, "An EM-based OFDM Receiver," *to be submitted to IEEE Trans. Signal Processing.*