BLIND MAXIMUM-LIKELIHOOD CHANNEL AND DATA RECOVERY IN OFDM

Tareq Y. Al-Naffouri[1], Ahmad Bahai[2], Arogyaswami Paulraj[1]

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

[2] 2900 Semiconductor Drive
National Semiconductor
Santa Clara, CA 95052

Abstract

OFDM modulation combines the advantages of high achievable rates and relatively easy implementation. In this paper, we show how to perform exact maximum-likelihood channel and data recovery in OFDM transmission. Our approach relies on decomposing the OFDM channel into two subchannels (cyclic and linear) that share the same input and are characterized by the same channel parameters. This fact enables us to estimate the channel parameters from one subchannel and substitute the estimates into the other thus obtaining a nonlinear relationship involving the input and the output data only. The relationship does not show any channel dependence whatsoever and can be exhaustively searched for the maximum-likelihood estimate of the input. This shows that OFDM systems are completely identifiable using output data only irrespective of the location of the channel zeros as long as the delay spread is less than the length of the cyclic prefix.

1. INTRODUCTION

There has been increasing interest in OFDM as it combines the advantages of high achievable rates and easy implementation. This is reflected by the many standards that considered and adopted OFDM including those for digital audio and video broadcasting, high speed modems over digital subscriber lines, and local area wireless broadband systems. [1]

OFDM divides the communication channel into independent subchannels by appending a cyclic prefix (CP) to the data block transmitted through the channel. This turns out to a very convenient structure that lends itself to exploiting the various constraints imposed by the transmitter and the channel. As such, many techniques have been proposed in literature to estimate and equalize channels for OFDM transmission (see, e.g., [1], [2], [3], [4], and the references therein). In this paper, we propose a method for blind channel and data recovery in OFDM. Our method has the following main advantages:

1. Channel identification and equalization is performed from output data only irrespective of the channel zero locations, [1] and without the need for a training sequence or a priori channel information. The channel order is also immaterial as long as it is less than or equal to the length of the CP.
2. Both channel and data recovery are performed in the maximum-likelihood (ML) sense thus taking care of the presence of noise in an optimal manner.
3. While the quality of the data estimate is dependent on the the SNR, as it should, it is completely independent of the quality of the channel estimate. This is demonstrated by showing that the ML data estimate is the minimum of a criterion that depends on the output data only.

1.1. A Sketch of our Approach

The presence of a cyclic prefix at the input transforms the linear OFDM channel into two parallel subchannels:

[1] This comes contrary to the common belief that OFDM can not be equalized for channels with zeros on the FFT grid [1], [2], and [5]

1. A cyclic subchannel that relates the ith input and output packets and thus is free of any ISI effects. This subchannel is best described in the frequency domain.

2. A linear subchannel that carries the burden of ISI and that relates the input and output prefixes through linear convolution. This subchannel is best studied in the time domain.

It can be shown that the two subchannels are characterized by the same set of parameters (or impulse response (IR)) and are driven by the same stream of data. They only differ in the way in which they operate on the data (i.e., linear vs. circular convolution). This fact enables us to estimate the IR from one subchannel and eliminate its effect from the other, thus obtaining a nonlinear relationship that involves the input and output data only. This relationship can in turn be optimized for the ML data estimate; something that can be achieved through exhaustive search (in the worst case scenario).

1.2. Organization of the Paper

This 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 an OFDM channel decomposing it into a cyclic subchannel described in the frequency domain and a linear subchannel described in the time domain. We bridge the gap between the two descriptions by introducing what we call time-frequency characterization. In section 4, we show how this characterization can be used to perform joint ML recovery of channel and data information. We also address the implications of our findings on the identifiability of OFDM channels from output data only, and touch on extensions to zero-padded OFDM.

2. NOTATION

We use regular small-case letters to denote scalars and small-case boldface letters to denote vectors. Caligraphic notation (e.g. $\boldsymbol{\mathcal{X}}$) is used to denote vectors in the transform (frequency) domain. When these variables become a function of time, the time index i appears between parentheses for scalars (e.g. $x(i)$) and as a subscript for vectors (e.g. $\boldsymbol{h}_i$). Uppercase boldface letters are reserved for matrices.

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 an upper (prefix) vector $\underline{\boldsymbol{x}}_i$ and a lower (usually longer) vector $\tilde{\boldsymbol{x}}_i$ so that

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

The third derivative, $\overline{\boldsymbol{x}}_i$, is created by concatenating $\boldsymbol{x}_i$ with a copy of its prefix $\underline{\boldsymbol{x}}_i$. Thus, we have

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

This notational convention of underlined and overlined variables will be extended to matrices as well. Thus, in line with the above notation, a matrix $\boldsymbol{Q}$ having N rows will have the natural partitioning

$$\boldsymbol{Q} = \begin{bmatrix} \underline{\boldsymbol{Q}} \\ \tilde{\boldsymbol{Q}} \end{bmatrix} \tag{2}$$

where the number of rows in $\underline{\boldsymbol{Q}}$ and $\tilde{\boldsymbol{Q}}$ are understood from the context. When this is not clear enough, the number of rows will appear as a subscript and we write

$$\boldsymbol{Q} = \begin{bmatrix} \underline{\boldsymbol{Q}}_L \\ \tilde{\boldsymbol{Q}}_{N-L} \end{bmatrix} \tag{3}$$

3. ESSENTIAL ELEMENTS OF OFDM TRANSMISSION

Consider the sequence $\{\mathcal{X}(i)\}$ that we wish to transmit. Data are collected and transmitted in packets $\boldsymbol{\mathcal{X}}_i$ of length N. In an OFDM system, the packet vector $\boldsymbol{\mathcal{X}}_i$ undergoes an IDFT operation to produce the transform vector $\boldsymbol{x}_i$. The two vectors are thus related by the unitary transformation

$$\boldsymbol{x}_i = \boldsymbol{Q}\boldsymbol{\mathcal{X}}_i \tag{4}$$

where $\boldsymbol{Q}$ is the 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 length $L+1$, it will be subject to intersymbol interference (ISI). To go around this, a guard band is inserted between any consecutive packets, $\boldsymbol{x}_{i-1}$ and $\boldsymbol{x}_i$. In particular, to each packet, we append a cyclic prefix of length L as done in (1). Thus, instead of transmitting $\boldsymbol{x}_i$, we transmit the length $N+L$ packet $\overline{\boldsymbol{x}}_i$ as defined in (1). The concatenation of these packets 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 $\{\overline{y}(i)\}$. Motivated by the packet structure of the input, it is convenient to deal with the output in the form of packets of length $M = N+L$, and further split each packet into a length-N packet $\boldsymbol{y}_i$ and a prefix associated with it $\underline{\boldsymbol{y}}_i$, i.e.

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

This is a natural way to partition the output because the prefix $\underline{\boldsymbol{y}}_i$ actually absorbs all ISI that takes place between the adjacent packets $\overline{\boldsymbol{x}}_{i-1}$ and $\overline{\boldsymbol{x}}_i$. Moreover, the remaining part $\boldsymbol{y}_i$ of the packet depends on the ith input packet only. These facts and more can be seen from the input/output relationship

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

$$\left[\begin{array}{cc|cc} \multicolumn{2}{c|}{\overline{\boldsymbol{H}}} & \boldsymbol{O}_{N\times L} & \boldsymbol{O}_{N\times N} \\ \cdots & & \cdots & \\ \boldsymbol{O}_{L\times N} & \underline{\boldsymbol{H}}^B & \underline{\boldsymbol{H}}^F & \boldsymbol{O}_{L\times N} \\ \cdots & & \cdots & \\ \boldsymbol{O}_{N\times N} & \boldsymbol{O}_{N\times L} & \multicolumn{2}{c}{\overline{\boldsymbol{H}}} \end{array}\right] \begin{bmatrix} \underline{\boldsymbol{x}}_{i-1} \\ \tilde{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix}$$

where $\boldsymbol{n}$ is the output noise which we take to be white Gaussian. The matrices $\overline{\boldsymbol{H}}$, $\underline{\boldsymbol{H}}^B$, and $\underline{\boldsymbol{H}}^F$ are convolution (Toeplitz) matrices of proper sizes created from the vector $\underline{\boldsymbol{h}}$. In particular, we have

$$\underline{\boldsymbol{H}}^B = \begin{bmatrix} \underline{h}(0) & & & & \\ \underline{h}(1) & \underline{h}(0) & & & \\ \underline{h}(2) & \underline{h}(1) & \underline{h}(0) & & \\ \vdots & \ddots & \ddots & \ddots & \\ \underline{h}(L-1) & \cdots & \underline{h}(2) & \underline{h}(1) & \underline{h}(0) \end{bmatrix} \tag{7}$$

$$\underline{\boldsymbol{H}}^F = \begin{bmatrix} \underline{h}(L) & \underline{h}(L-1) & \cdots & \underline{h}(1) \\ & \underline{h}(L) & \cdots & \underline{h}(2) \\ & & \ddots & \vdots \\ & & & \underline{h}(L) \end{bmatrix} \tag{8}$$

Because of the redundancy in the input, the convolution in (6) can be decomposed into two distinct constituent convolution operations or subchannels. This decomposition is essential for channel and data recovery, which is the center of attention in this paper. In what follows, we shall describe each of these operations separately.

3.1. Circular Convolution (Subchannel)

From (6), we can write

$$\boldsymbol{y}_i = \overline{\boldsymbol{H}} \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} = \overline{\boldsymbol{H}}\,\overline{\boldsymbol{x}}_i + \boldsymbol{n}_i \tag{9}$$

Thus, $\boldsymbol{y}_i$ is created solely from $\overline{\boldsymbol{x}}_i$ through convolution and hence is free of inter-block interference. Moreover, the existence of a cyclic prefix in $\overline{\boldsymbol{x}}_i$ renders this convolution cyclic. In other words, we can rewrite (9) as

$$\boxed{\boldsymbol{y}_i = \boldsymbol{h} \circledast \boldsymbol{x}_i + \boldsymbol{n}_i} \tag{10}$$

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

$$\boldsymbol{h} = \begin{bmatrix} \underline{\boldsymbol{h}} \\ \boldsymbol{O}_{(N-L-1)\times 1} \end{bmatrix} \tag{11}$$

In the frequency domain, the cyclic convolution (10) reduces to the element-by-element operation

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}} \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i} \tag{12}$$

where $\boldsymbol{\mathcal{H}}$, $\boldsymbol{\mathcal{X}}_i$, and $\boldsymbol{\mathcal{Y}}_i$ are the DFT's of $\boldsymbol{h}$, $\boldsymbol{x}_i$, and $\boldsymbol{y}_i$, respectively. In other words, we have

$$\boldsymbol{h} = \boldsymbol{Q}\boldsymbol{\mathcal{H}}, \quad \boldsymbol{x}_i = \boldsymbol{Q}\boldsymbol{\mathcal{X}}_i, \quad \text{and} \quad \boldsymbol{y}_i = \boldsymbol{Q}\boldsymbol{\mathcal{Y}}_i \tag{13}$$

We can also show that

$$\underline{\boldsymbol{h}} = \underline{\boldsymbol{Q}}_{L+1}\boldsymbol{\mathcal{H}} \quad \underline{\boldsymbol{x}}_i = \tilde{\boldsymbol{Q}}_L \boldsymbol{\mathcal{X}}_i \tag{14}$$

which we will find useful further ahead.

3.2. Linear Convolution (Subchannel)

From (6), we can also extract a constituent relationship between the input and output prefixes

$$\underline{\boldsymbol{y}}_i = \begin{bmatrix} \underline{\boldsymbol{H}}^B & \underline{\boldsymbol{H}}^F \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_{i-1} \\ \underline{\boldsymbol{x}}_i \end{bmatrix} + \underline{\boldsymbol{n}}_i \tag{15}$$

This can be used to show that the underlying prefix sequences $\{\underline{x}(i)\}$ and $\{\underline{y}(i)\}$ are related together with the channel $\underline{h}$ through a linear convolution [6], i.e.

$$\boxed{\underline{y}(i) = \underline{h}(i) * \underline{x}(i) + \underline{n}(i)} \tag{16}$$

Remarks

- By now, the convenience of the notation adopted herein should be appreciated. Underlining (e.g., $\underline{\boldsymbol{x}}_i$) and overlining (e.g., $\overline{\boldsymbol{x}}_i$) reflect the relation of the vectors involved to the original vector ($\boldsymbol{x}_i$ in this case). More importantly, this notation extends naturally to the channel and output variables, and enables us to write fundamental relationships like (10), (12), and (16) almost by inspection.

- Notice that the cyclic relationship (12) is expressed in terms of the frequency domain quantities $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{H}}$ while (15) is expressed in terms of the time domain quantities $\underline{\boldsymbol{x}}_i$ and $\underline{\boldsymbol{h}}$. To perform ML recovery, we need to combine the estimates obtained from (10) and (15). However, as a prerequisite, we need to directly express either of these relationships in the transform domain. In what follows, we digress to express (15) in a time-frequency form. [2]

3.3. Time-Frequency Representation of Linear Convolution

Consider the linear convolution (15). It is difficult to express this relation totally in the frequency domain and simultaneously maintain a succinct matrix structure. We resort instead to a hybrid time-frequency form wherein either the channel or the data are expressed in the frequency domain, but not both.

To express (15) in terms of $\boldsymbol{\mathcal{H}}$, we need as a prerequisite to interchange the roles of $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{h}}$ in (15). In other words, we need to express it in the form

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

Starting from (16), it is easy to show that the matrix $\underline{\boldsymbol{X}}_i$ actually takes the form:

$$\underline{\boldsymbol{X}}_i = \underline{\boldsymbol{X}}_{i-1}^B + \underline{\boldsymbol{X}}_i^F \tag{17}$$

where (compare with (7) and (8))

$$\underline{\boldsymbol{X}}_{i-1}^B = \begin{bmatrix} 0 & \underline{x}_{i-1}(L-1) & \cdots & \underline{x}_{i-1}(1) & \underline{x}_{i-1}(0) \\ 0 & 0 & \cdots & \underline{x}_{i-1}(2) & \underline{x}_{i-1}(1) \\ \vdots & \vdots & \cdots & \underline{x}_{i-1}(2) & \underline{x}_{i-1}(1) \\ 0 & 0 & \cdots & 0 & \underline{x}_{i-1}(L-1) \end{bmatrix}$$

$$\underline{\boldsymbol{X}}_i^F = \begin{bmatrix} \underline{x}_i(0) & 0 & \cdots & & 0 \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ \underline{x}_i(L-1) & \cdots & \underline{x}(1) & \underline{x}_i(0) & 0 \end{bmatrix}$$

This fact together with the FFT relationship (14)

$$\underline{\boldsymbol{h}} = \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{H}}$$

yields the desired time-frequency form

$$\boxed{\begin{aligned} \underline{\boldsymbol{y}}_i &= \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{H}} + \underline{\boldsymbol{n}}_i \\ &= \left(\underline{\boldsymbol{X}}_{i-1}^B \underline{\boldsymbol{Q}}_{L+1} + \underline{\boldsymbol{X}}_i^F \underline{\boldsymbol{Q}}_{L+1} \right) \boldsymbol{\mathcal{H}} + \underline{\boldsymbol{n}}_i \end{aligned}} \tag{18}$$

[2] Alternatively, we could express (12) in a time-frequency form. This is not as attractive, however, since it effectively means loosing the diagonal form of (12). Another important feature about (12) is that it is directly expressed in terms of the row data $\boldsymbol{\mathcal{X}}_i$.

4. MAXIMUM-LIKELIHOOD ESTIMATION

Consider the frequency domain description of the cyclic subchannel (12)

$$\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}} \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i \tag{19}$$

To obtain the ML estimate of $\boldsymbol{\mathcal{H}}$, we assume that the sequence $\boldsymbol{\mathcal{X}}_i$ is deterministic and perform an element-by-element division of (19) by $\boldsymbol{\mathcal{X}}_i$ to get

$$\boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}} + \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{N}}_i \tag{20}$$

where

$$\boldsymbol{D}_{\mathcal{X}}^{-1} = \mathrm{diag}(\boldsymbol{\mathcal{X}}_i) \tag{21}$$

Equivalently, we can write (20) as

$$\boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{N}}_i' \tag{22}$$

where $\boldsymbol{\mathcal{N}}'$ is Gaussian distributed with zero mean and autocorrelation matrix

$$\boldsymbol{R}_{n'} = \sigma_n^2 \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{D}_{\mathcal{X}}^{-*} = \sigma_n^2 |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tag{23}$$

4.1. ML Estimation of $\boldsymbol{\mathcal{H}}$

The maximum-likelihood estimate of $\boldsymbol{\mathcal{H}}$ can now be obtained by solving the system of equations (22) in the least-squares (LS) sense subject to the constraint

$$\tilde{\boldsymbol{Q}}_{N-L-1} \boldsymbol{\mathcal{H}} \triangleq \tilde{\boldsymbol{Q}} \boldsymbol{\mathcal{H}} = \boldsymbol{0} \tag{24}$$

We can show that the ML estimate is given by [7]

$$\hat{\boldsymbol{\mathcal{H}}}^{\mathsf{ML}} = \left[\boldsymbol{I} - \boldsymbol{R}_{n'} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} \boldsymbol{R}_{n'} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i \tag{25}$$

or in view of (23),

$$\hat{\boldsymbol{\mathcal{H}}}^{\mathsf{ML}} = \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i \tag{26}$$

4.2. Eliminating the Channel Effect

The ML estimate (26) was obtained solely from the cyclic convolution subchannel. The observations of the linear subchannel have not been used so far. Upon replacing $\boldsymbol{\mathcal{H}}$ in the time-frequency form (18) (corresponding to the linear subchannel)

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{H}} + \underline{\boldsymbol{n}}_i$$

with its ML estimate (26), we obtain

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i + \underline{\boldsymbol{n}}_i$$

This is an input/output relationship that does not depend on any channel information whatsoever. Since the data is assumed deterministic, maximum-likelihood estimation is the optimum way to detect it, i.e. we minimize

$$\hat{\boldsymbol{\mathcal{X}}}_i^{\mathsf{ML}} = \arg\min_{\mathcal{X}_i} \tag{27}$$

$$\left\| \underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i \right\|^2$$

This is a nonlinear least-squares problem in the data. In the worst case scenario, it can be solved by an exhaustive search over all possible sequences $\boldsymbol{\mathcal{X}}_i$.

To gain more insight into this problem, we now treat the case of constant modulus data which leads to more explicit developments.

4.3. ML Estimation in the Constant Modulus Case

In the constant modulus case, we have

$$\boldsymbol{D}_{\mathcal{X}}^{-2} = \frac{1}{\mathcal{E}_X} \boldsymbol{I} \tag{28}$$

As a consequence, we can also write

$$\boldsymbol{D}_{\mathcal{X}}^{-1} = \frac{1}{\mathcal{E}_X} \boldsymbol{D}_{\mathcal{X}} \tag{29}$$

Thus, the ML estimate of $\boldsymbol{\mathcal{H}}$ (26) simplifies to

$$\begin{aligned} \hat{\boldsymbol{\mathcal{H}}}^{\mathsf{ML}} &= \frac{1}{\mathcal{E}_X} \left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}} \boldsymbol{\mathcal{Y}}_i & (30) \\ &= \frac{1}{\mathcal{E}_X} \left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^* \tilde{\boldsymbol{Q}} \right] \boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i & (31) \end{aligned}$$

where in (31), we used the fact that $\tilde{\boldsymbol{Q}}$ is a left-inverse of $\tilde{\boldsymbol{Q}}^*$ – a consequence of the unitary nature of $\boldsymbol{Q}$

$$\boldsymbol{I} = \boldsymbol{Q}\boldsymbol{Q}^* = \begin{bmatrix} \underline{\boldsymbol{Q}}_{L+1} \\ \tilde{\boldsymbol{Q}}_{N-L-1} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{Q}}_{L+1}^* & \tilde{\boldsymbol{Q}}_{N-L-1}^* \end{bmatrix} \tag{32}$$

Just as we did in the general case, we now replace the effect of $\boldsymbol{\mathcal{H}}$ in the linear convolution subchannel, as expressed in (18), by its ML estimate to get

$$\begin{aligned} \underline{\boldsymbol{y}}_i &= \frac{1}{\mathcal{E}_X} \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \left[\boldsymbol{I} - \tilde{\boldsymbol{Q}}^* \tilde{\boldsymbol{Q}} \right] \boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i + \underline{\boldsymbol{n}}_i & (33) \\ &= \frac{1}{\mathcal{E}_X} \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i + \underline{\boldsymbol{n}}_i & (34) \end{aligned}$$

where in going to (34), we used the fact that

$$\underline{\boldsymbol{Q}}_{L+1} \tilde{\boldsymbol{Q}} = \underline{\boldsymbol{Q}}_{L+1} \tilde{\boldsymbol{Q}}_{N-L-1} = \boldsymbol{0}$$

which can be deduced from (32). The ML estimate of $\boldsymbol{\mathcal{X}}_i$ is now obtained by performing the minimization

$$\hat{\boldsymbol{\mathcal{X}}}_i^{\mathsf{ML}} = \arg\min_{\mathcal{X}_i} \left\| \underline{\boldsymbol{y}}_i - \frac{1}{\mathcal{E}_X} \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i \right\|^2 \tag{35}$$

Notice that the only unknowns in this minimization are $\underline{\boldsymbol{X}}_i$ and $\boldsymbol{\mathcal{X}}_i$, i.e. the input data sequence. This minimization is nothing but a *nonlinear least-squares* problem in the data. In the worst case scenario, we can obtain the ML estimate through an exhaustive search.

Remarks

1. ML identification of FIR channels from output observations is a long standing problem [8]. In fact, it remains an open problem even in the noiseless case. Here we have shown how joint ML channel identification and data detection can be achieved for an FIR system when driven by a special (OFDM deterministic finite alphabet) input.

2. By following the derivations closely, we can that one critical step in our derivation is the deterministic assumption on the input (not its finite alphabet or repetitive nature). For it is this what makes the LS channel estimation of (26) and (31) equivalent to ML identification.

3. Nevertheless, a deterministic assumption on the data is more realistic than a statistical assumption, and becomes even more so for short data records (packets) for which the statistical assumptions don't substantiate. [9]

4. Another critical step in our derivation is the incorporation of *all* output samples (including the output prefix) produced by the input packet. If, on the other hand, the output prefix is discarded, even channel identifiability becomes at stake [1], [2], [5].

5. We would like to stress that the repetitive nature of the data (i.e. the presence of the cyclic prefix) has no bearing on our ability to perform ML recovery. In fact, by following a similar approach, we can still perform ML recovery for zero-padded OFDM transmission (as opposed to CP-padded one).

6. The above derivation dismisses the myth that CP-OFDM can not be equalized for channels with zeros on the FFT grid [1], [2], [5]. This would be true if the cyclic subchannel (12) only is used for data recovery. If, on the other hand, the linear and

cyclic subchannels are used, the location of the channel zeros become a non-issue in blind equalization.

7. What further distinguishes our approach from other equalization approaches is its up-front acknowledgment of the presence of noise. This comes contrary to many other multichannel equalization techniques wherein exact zero forcing solutions are sought while the noisy case is dealt with in an add-hoc manner [8], [10].

8. Not only is channel information not explicitly used in equalization but even the *channel order* becomes immaterial for that purpose. Thus, although (27) and (35) are explicitly dependent on L, this dependence reflects the length of the cyclic prefix, not the channel length. The only requirement for (27) and (35) to be valid is for the channel length to be less than or equal to the cyclic prefix.

9. Exponentially complex algorithms for blind channel-estimation/data-detection are not unusual. A finite alphabet procedure with exponential complexity has recently been proposed by Shengli and Giannakis [11].

5. CONCLUSION

In this paper, we demonstrated how to perform blind ML channel and data recovery in OFDM transmission. In particular, it was shown that the ML data estimate is the solution of an integer nonlinear-least squares problem. This proves that data recovery is possible from output data only irrespective of the channel zero locations and irrespective of the quality of the channel estimates or of its exact order. The algorithm developed in this paper entails exponential complexity whose reduction is the subject of future research.

6. REFERENCES

[1] Z. Wang and G. B. Giannakis, "Wireless Multicarrier Communications: Where Fourier Meets Shannon," IEEE Signal Process. Mag., vol. 17, No. 3, pp. 29-48, May 2000.

[2] B. Muquet, M. de Courville, G. B. Giannakis, Z. Wang, and P. Duhamel, "Reduced Complexity Equalizers for Zero-Padded OFDM Transmissions," Proc. of Intl. Conf. on Acoust. Speech and Signal Proc., vol. 5, pp. 2973-2976, Istanbul, Turkey, June 5-9, 2000.

[3] R. W. Heath Jr. and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," em IEEE Trans. Signal Process., vol 47, pp. 774-780, Mar. 1999.

[4] B. Muquet, M. de Courville, P. Duhamel, and V. Buzenae, "A subspace based blind and semi-blind channel identification method for OFDM systems," Proc. IEEE Workshop for Signal Process. Advances in Wireless Comm., Annapolis, MD, May 1999, pp. 243-246.

[5] B. Muquet, M. de Courville, P. Duhamel and G. B. Giannakis, " OFDM with Trailing Zeros versus OFDM with Cyclic Prefix: Links, Comparisons and Application to the HiperLAN/2 System," Proc. of Intl. Conf. on Communications, pp. 1049 -1053, New Orleans, LA, June 18-22, 2000.

[6] X. Wang and R. Liu, "Adaptive channel estimation using cyclic prefix in multicarrier modulation system," *IEEE Comm. Lett.,* vol. 3, no. 10, Oct. 1999.

[7] T. Soderstrom and P. Soitca, "System Identification," Prentice-Hall, 1989.

[8] P. Stoica, T. McKelvey, and J. Mari, "MA Estimation in Polynomial Time," IEEE Transactions on Signal Processing, vol. 48, pp. 1999-2012, Jul 2000.

[9] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," IEEE Transactions on Signal Processing, vol. 44, pp. 661-671, Mar. 1996.

[10] G. B. Giannakis and C. Tepedelenlioglu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," IEEE Transactions on Signal Processing, vol. 47, pp. 62-74, January 1999.

[11] S. Zhou and G. B. Giannakis, " Finite-Alphabet based Channel Estimation for OFDM and related Multi-Carrier Systems," IEEE Transactions on Communications, vol. 49, no. 8, pp. 1402-1414, August 2001.