Asymptotically Efficient Reduced Complexity Frequency Offset and Channel Estimators for Uplink MIMO-OFDMA Systems

Serdar Sezginer, Student Member, IEEE, Pascal Bianchi*, Member, IEEE

Abstract

In this paper, we address the joint data-aided estimation of frequency offsets and channel coefficients in uplink MIMO-OFDMA systems. As the Maximum Likelihood (ML) estimator is impractical in this context, we introduce a family of suboptimal estimators with the aim of exhibiting an attractive tradeoff between performance and complexity. The estimators do not rely on a particular subcarrier assignment scheme (SAS) and are thus valid for a large number of OFDMA systems. As far as complexity is concerned, the computational cost of the proposed estimators is shown to be significantly reduced compared to existing estimators based on ML. As far as performance is concerned, the proposed suboptimal estimators are shown to be asymptotically efficient, i.e., the covariance matrix of the estimation error achieves the Cramér-Rao bound when the total number of subcarriers increases. Simulation results sustain our claims.

Index Terms

Asymptotic efficiency, frequency synchronization, channel estimation, MIMO, OFDMA.

EDICS category: SSP-PERF; SPC-MULT

I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) has recently become very popular in wireless communications and already been included in IEEE 802.16 specifications for fixed and mobile broadband wireless access. In an OFDMA system, each user modulates a certain group of subcarriers, following

S. Sezginer and P. Bianchi are with the Telecommunications Department, Supélec, Plateau de Moulon, F-91192 Gif-sur-Yvette, France (e-mail: {serdar.sezginer, pascal.bianchi}@supelec.fr, tel: +33 (0)1 69 85 14 55, fax: +33 (0)1 69 85 14 69).

* Corresponding Author.
a given subcarrier assignment scheme (SAS). The signal transmitted by a given user is impaired by a frequency selective channel and by a certain frequency offset. Along with its attractive features, the drawbacks associated with Orthogonal Frequency Division Multiplexing (OFDM) are directly inherited by OFDMA: in OFDM, it is well-known that the presence of frequency offsets introduces intercarrier interference (ICI) which is damaging in terms of symbol detection. In addition to ICI, inaccurate frequency offset estimation furthermore introduces multiple-access interference (MAI) in uplink OFDMA systems which degrades as well the overall system’s performance. Accurate estimation of the unknown frequency offsets and channel coefficients is therefore a crucial issue in OFDMA.

Estimation of frequency offset and/or channel coefficients for (single-user) OFDM systems has been thoroughly investigated in the literature (see, e.g., [1], [2], and references therein). Moreover, a number of works has been devoted to the analysis of the performance of frequency offset estimators in a single user context [3]–[5]. In the single user case, it is well-known that accurate frequency offset and channel estimation can be achieved with very reasonable complexity. In the uplink OFDMA case, frequency offset and channel estimation is unfortunately a more difficult task. In particular, parameter estimation based on the direct maximization of the log-likelihood function is impractical. Therefore, suboptimal algorithms have been proposed. Most of these estimators are non data-aided and are designed for a specific SAS. In case of clustered SAS, i.e., when each user modulates a group of contiguous subcarriers, [6] proposes an estimator based on the presence of non-modulated subcarriers between groups of modulated subcarriers. Other examples include estimators based on a cyclic prefix method proposed in [7] and a Kurtosis maximization presented in [2]. In case of the equispaced (also called the interleaved) SAS, i.e., when adjacent subcarriers are modulated by different users, [8] proposes a subspace method based on the periodic structure of the signals transmitted by each user.

Although non data-aided techniques for synchronization and channel estimation have recently taken considerable attention, the recently standardized multicarrier systems employ training blocks and thus encourage the design of data-aided estimators. In parallel, current trends for transmission in OFDMA systems bring to the fore the importance of more flexible estimators which do not rely on particular SAS. Recent works [9] and [10] investigate the data-aided estimation of frequency offsets in OFDMA uplink for general SAS. In [9], the estimation of carrier frequency and timing offsets of a new user entering an OFDMA system is addressed, assuming that the other users have already been synchronized. On the other hand, [10] proposes an ML-based alternating-projection algorithm. At each iteration of the algorithm, the procedure mainly consists in estimating the parameter of only one user while keeping the estimates of the other users at their most updated values. This method allows to replace the multi-
dimensional exhaustive search initially required by the direct likelihood maximization with a succession of one-dimensional exhaustive searches. The complexity is therefore reduced compared to the rigorous ML estimator whereas the performance remains close to the performance of the ML estimator. However, the algorithm of [10] is still computationally demanding. For instance, the algorithm is difficult to implement in situations where the number of subcarriers is large. It is therefore of practical interest to propose suboptimal estimators which are likely to be implemented in such contexts, i.e., which have reasonable complexity and, on the other hand, which have a performance close to the ML performance.

In this paper, we consider an uplink OFDMA system with $K$ users in a Multiple-Input Multiple-Output (MIMO) context. We address the issue of the estimation of the set of $K$ frequency offsets $\omega = [\omega_1, \ldots, \omega_K]^T$ corresponding to each of these $K$ users, and the set of $K$ vectors of channel coefficients. The signal model is described in Section II. We briefly recall the principle of the ML estimator in Section III. It is shown that the computational burden of the ML estimator is mainly due to the fact that the evaluation of the log-likelihood function $\hat{\omega} \rightarrow J_{N}^{\text{ML}}(\hat{\omega})$ requires the inversion of a certain matrix for each trial value $\hat{\omega}$ of the set of frequency offsets $\omega$. The proposed family of estimators is introduced in Section IV. It is based on the observation that the costly matrix inverse involved in the ML estimation can be approximated by an other matrix which is simpler to evaluate. This approximation is motivated by previous works [11] and is particularly accurate when the number of subcarriers is large. Although the proposed estimation method is suboptimal, we prove in Section V that the proposed estimator is asymptotically efficient. In other words, its performance coincides with the performance of the ML algorithm provided that the number of subcarriers is large enough. Simulation results are presented in Section VII.

II. Signal Model

We consider an uplink MIMO-OFDMA transmission. We assume that $K$ users share $N$ subcarriers. Each user has $N_T$ transmit antennas. One symbol sequence is sent by each transmit antenna $t$ ($t = 1, \ldots, N_T$) of each user $k$ ($k = 1, \ldots, K$) using an OFDM modulator. The OFDM symbol transmitted by user $k$ at a given antenna $t$ in the frequency domain is represented by sequence $s_{N,k}(t,0), \ldots, s_{N,k}(t,N-1)$. We omit the block index for the sake of simplicity. In the sequel, we assume that for each $k$ and for each $t$, sequence $(s_{N,k}(t,j))_j$ is known by the receiver (training sequence). It is worth noting that in usual OFDMA systems, only a subset of the $N$ available subcarriers is effectively modulated by a given user $k$, following a given SAS. For each $j = 0, \ldots, N-1$, we simply consider that $s_{N,k}(t,j) = 0$ in the case where subcarrier $j$ is not modulated by user $k$. However, we do not specify any SAS. In our model, training sequences $(s_{N,k}(t,j))_j$ and $(s_{N,k}(t',j))_j$ sent at different antennas $t$ and $t'$ are possibly different. For a given user $k$ and a given
antenna $t$, we denote by $(a_{N,k}^{(t)}(n))_n$ the inverse discrete Fourier transform of sequence $(s_{N,k}^{(t)}(j))_j$:

$$a_{N,k}^{(t)}(n) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} s_{N,k}^{(t)}(j)e^{2\pi j n j/N}$$

for each integer $n$. Cyclic prefix is added to the above time-domain version of the OFDM block and the resulting sequence is transmitted over a multipath channel.

We denote by $N_R$ the number of receive antennas at the base station. For each $r = 1, \ldots, N_R$, the complex envelope of the signal received by antenna $r$ is sampled at symbol rate. After cyclic prefix removal, the corresponding received samples can be written for each $n = 0, \ldots, N - 1$ as

$$y_N^{(r)}(n) = \sum_{k=1}^{K} e^{j\omega_k n} \sum_{l=1}^{N_T} \sum_{i=0}^{L-1} h_k^{(t,r)}(l) a_{N,k}^{(t)}(n-l) + v^{(r)}(n).$$

For each user $k$, parameter $\omega_k$ is defined as $\omega_k = 2\pi \delta f_k T$, where $\delta f_k$ denotes the frequency offset corresponding to user $k$ and where $T$ denotes the sampling period. Parameter $h_k^{(t,r)}(l)$ represents $l$th tap of the channel impulse response between $t$th transmit antenna of user $k$ and $r$th receive antenna of the base station. Each channel is assumed to have no more than $L$ nonzero taps, where integer $L$ does not depend on $k$ and does not exceed the length of the cyclic prefix. Sequence $(v^{(r)}(n))_n$ denotes a white Gaussian noise of variance $\sigma^2$. Note that equation (2) implicitly assumes that all users are quasi-synchronous in time: all delays of signals transmitted by all users are within the length of cyclic prefix. In equation (2), we also assume that the (angular) frequency offset $\omega_k$ is constant with respect to (w.r.t.) antenna pairs $(t, r)$. We mention that in certain MIMO systems, different frequency offsets may be associated with each transmit-receive antenna pair. This case is usually considered in macro-diversity systems [12]. In the present paper, we consider the classical case (see, e.g., [2] and references therein) where $\omega_k$ is constant w.r.t. antenna pairs $(t, r)$. In the sequel, it is convenient to make use of a compact matrix representation of (2). To that end, we introduce the following notations. Define $h_k^{(t,r)} = [h_k^{(t,r)}(0), \ldots, h_k^{(t,r)}(L-1)]^T$ and $h_k^{(r)} = [h_k^{(1,r)}T, \ldots, h_k^{(N_R,r)}T]^T$, where $(\cdot)^T$ represents the transpose operator. Stacking all $N$ samples $y_k^{(r)}(n)$ received by antenna $r$ into one column vector $y_N^{(r)} = [y_N^{(r)}(0), \ldots, y_N^{(r)}(N-1)]^T$, one obtains:

$$y_N^{(r)} = \sum_{k=1}^{K} \Gamma_N(\omega_k) A_{N,k} h_k^{(r)} + v_N^{(r)},$$

where $\Gamma_N(\omega_k) = \text{diag}(1, e^{j\omega_k}, \ldots, e^{j\omega_k(N-1)})$ and $v_N^{(r)} = [v^{(r)}(0), \ldots, v^{(r)}(N-1)]^T$. Here, the matrices $A_{N,k}$ are defined as $A_{N,k} = [A_{N,k}^{(1)}, \ldots, A_{N,k}^{(N_R)}]$ where for each antenna $t$, $A_{N,k}^{(t)}$ is an $N \times L$ matrix containing the time-domain training sequence sent at the $t$th transmit antenna of user $k$. More precisely, $A_{N,k}^{(t)} = [(a_{N,k}^{(t)}(i-j))_{0 \leq i \leq N-1}^{0 \leq j \leq L-1}$. Note that $A_{N,k}^{(t)}$ is a circulant matrix because of the cyclic prefix insertion.
at the transmitter side. We finally stack the samples received by all antennas into a single $NN_R \times 1$ vector $y_N = [y_N^{(1)^T}, \ldots, y_N^{(Nn)^T}]^T$ given by
\[
y_N = \sum_{k=1}^{K} [\mathbf{I}_{N_R} \otimes (\Gamma_N(\omega_k) \mathbf{A}_{N,K})] \mathbf{h}_k + \mathbf{v}_N, \tag{4}
\]
where $\mathbf{I}_{N_R}$ denotes the $N_R \times N_R$ identity matrix, $\otimes$ stands for the Kronecker product, and $\mathbf{v}_N = [\mathbf{v}_N^{(1)^T}, \ldots, \mathbf{v}_N^{(Nn)^T}]^T$ is an additive noise vector with independent complex circular Gaussian random entries of variance $\sigma^2$. Here, vector $\mathbf{h}_k = [\mathbf{h}_k^{(1)^T}, \ldots, \mathbf{h}_k^{(Nn)^T}]^T$ contains all channel coefficients of a given user $k$. Denoting by $\omega = [\omega_1, \ldots, \omega_K]^T$ the vector containing all frequency offsets and by $\mathbf{h} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]^T$ the vector of channel coefficients, (4) can be written under the following compact form
\[
y_N = \mathbf{Q}_N(\omega) \mathbf{h} + \mathbf{v}_N, \tag{5}
\]
where $\mathbf{Q}_N(\omega)$ represents the following $N_R N \times K N_R N_T L$ matrix
\[
\mathbf{Q}_N(\omega) = [\mathbf{I}_{N_R} \otimes (\Gamma_N(\omega_1) \mathbf{A}_{N,1}), \ldots, \mathbf{I}_{N_R} \otimes (\Gamma_N(\omega_K) \mathbf{A}_{N,K})]. \tag{6}
\]

In the sequel, we investigate the data-aided ML estimation of the frequency offsets and the channel parameters associated with all $K$ users.

### III. Exact ML Estimation of Frequency Offsets and Channel Parameters

We describe the exact ML estimator of the unknown deterministic parameter vector $[\omega, \mathbf{h}]^T$. The log-likelihood function for the unknown parameters $\omega, \mathbf{h}$ has the form
\[
\Lambda_N(\tilde{\omega}, \tilde{\mathbf{h}}) = -\frac{1}{\sigma^2} \| y_N - \mathbf{Q}_N(\tilde{\omega}) \tilde{\mathbf{h}} \|^2 + C_N, \tag{7}
\]
where $\tilde{\omega} = [\tilde{\omega}_1, \ldots, \tilde{\omega}_K]^T$ and $\tilde{\mathbf{h}} = [\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_K]^T$ are trial values of $\omega$ and $\mathbf{h}$, respectively. Here, $C_N$ designates a constant which does not depend on the parameter vector. Notation $\| \mathbf{x} \|^2$ stands for $\mathbf{x}^H \mathbf{x}$ where $\mathbf{x}$ is any column vector. The ML estimates $\hat{\omega}_N^{ML}, \hat{\mathbf{h}}_N^{ML}$ of parameters $\omega, \mathbf{h}$ are defined as the argument of the maximum of the log-likelihood function. The latter maximization can be simplified as follows. For any fixed value of $\tilde{\omega}$, the log-likelihood $\Lambda_N(\tilde{\omega}, \tilde{\mathbf{h}})$ is maximum when $\tilde{\mathbf{h}}$ coincides with
\[
\hat{\mathbf{h}}_N(\tilde{\omega}) = (\mathbf{Q}_N^H(\tilde{\omega}) \mathbf{Q}_N(\tilde{\omega}))^{-1} \mathbf{Q}_N^H(\tilde{\omega}) y_N. \tag{8}
\]
Therefore, the maximum of (7) coincides with the maximum of function $\hat{\omega} \rightarrow \Lambda_N(\hat{\omega}, \hat{\mathbf{h}}(\hat{\omega}))$. Substituting the above expression of $\hat{\mathbf{h}}$ in (7), we conclude that the log-likelihood function is maximum for $\hat{\omega}_N^{ML} = \arg \max_{\omega} J_N^{ML}(\omega)$, where
\[
J_N^{ML}(\omega) = y_N^H \mathbf{Q}_N(\omega) \left( \mathbf{Q}_N^H(\omega) \mathbf{Q}_N(\omega) \right)^{-1} \mathbf{Q}_N^H(\omega) y_N. \tag{9}
\]
In practice, the exact ML estimation of the whole set of parameters can thus be divided in two steps. Firstly, ML estimates of the frequency offsets are obtained by maximization of (9). Secondly, ML estimates of the channel coefficients are obtained by

\[ \hat{h}_{ML}^N = \hat{h}^N(\hat{\omega}_{ML}^N), \]

where function \( \hat{h}(\hat{\omega}) \) is defined as in (8).

However, the determination of \( \hat{\omega}_{ML}^N \) requires not only a \( K \) dimensional search but also an \( KN_TN_RL \times KN_TN_RL \) matrix inversion for each trial value \( \tilde{\omega} \). This makes the rigorous ML estimation impractical. Hence, simpler estimators are needed having similar performance to the ML estimator with lower complexity.

IV. PROPOSED ESTIMATOR

In this section, we first provide some preliminary remarks which allow to motivate the introduction of the proposed estimate. Basically, our approach consists in replacing the inner factor \( (Q^H_N(\tilde{\omega})Q_N(\tilde{\omega}))^{-1} \) of the righthand side of (9) with an appropriate approximation of the latter matrix. This approximation should be chosen such that the corresponding “approximated ML criterion” becomes easier to compute than the exact ML criterion \( J_{ML}^N(\tilde{\omega}) \). On the other hand, the approximation should be fine enough so that the estimation remains (almost) optimal in the ML sense. Thus, our aim is to define a simplified estimation algorithm whose performance becomes identical to the performance of the exact ML algorithm when the number \( N \) of subcarriers increases.

**Remark 1:** In our model, as \( N \) tends to infinity we assume that i) the number \( K \) of users remains constant and ii) the number of antennas remains constant. We also assume that when \( N \) tends to infinity, the overall bandwidth is constant. In other words, the sampling rate \( \frac{1}{T} \) remains constant and as a result, the subcarrier spacing \( \frac{1}{NT} \) decreases to zero.

A. The Main Idea

Here, we recall some results of [11] and explain how these results can be used to exhibit a simple approximation of \( (Q^H_N(\tilde{\omega})Q_N(\tilde{\omega}))^{-1} \). Consider a given (fixed) value of \( \tilde{\omega} \). Since our aim is to find an equivalent form for matrix \( (Q^H_N(\tilde{\omega})Q_N(\tilde{\omega}))^{-1} \) as \( N \) tends to infinity, we now study the asymptotic behavior of \( Q^H_N(\tilde{\omega})Q_N(\tilde{\omega}) \). Using (6), one obtain

\[
\frac{1}{N}Q^H_N(\tilde{\omega})Q_N(\tilde{\omega}) = \begin{bmatrix}
I_{N_H} \otimes U_{N,1,1} & \cdots & I_{N_H} \otimes U_{N,1,K} \\
\vdots & & \vdots \\
I_{N_H} \otimes U_{N,K,1} & \cdots & I_{N_H} \otimes U_{N,K,K}
\end{bmatrix},
\]

where for each \( k, l = 1, \ldots, K \),

\[
U_{N,k,l} = \frac{1}{N}A^H_{N,k} \Gamma_N(\tilde{\omega}_l - \tilde{\omega}_k)A_{N,l}.
\]
It has been shown in [11] that, under some mild assumptions which will be detailed in the sequel,

- if $k \neq l$, $U_{N,k,l}$ tends to the null matrix,
- if $k = l$, $U_{N,k,k}$ tends to a certain deterministic matrix $R_k$ which is related to the statistics of the training sequence $(s_{N,k}(j))_j$ of user $k$.

Before defining more accurately the above limit $R_k$, we briefly explain why the above lemma has an important consequence. Using this result, one immediately deduces that matrix $\frac{1}{N} Q^H_N(\tilde{\omega}) Q_N(\tilde{\omega})^{-1}$ converges to a block diagonal matrix $\text{diag}(I_{N \times R_1}, \ldots, I_{N \times R_K})$. Consequently, one could easily think about replacing the complicated matrix $(Q^H_N(\tilde{\omega}) Q_N(\tilde{\omega}))^{-1}$ in criterion $(9)$ with the asymptotically equivalent matrix $\frac{1}{N} \text{diag}(I_{N \times R_1^{-1}}, \ldots, I_{N \times R_K^{-1}})$. Although the estimator which is proposed in this paper turns out to be slightly more involved, it is based on the same kind of idea. Before introducing this novel estimator in Section IV-C, it is necessary to recall in more details the result of [11] which provides the limit of matrices $U_{N,k,l}$ as $N$ tends to infinity.

### B. A Former Result

From now on, we make the following assumptions on the training sequences transmitted by all users.

**Assumption 1:** For a given antenna $t$ of a given user $k$, $(s_{N,k}(j))_j$ is a sequence of independent random variables with zero mean.

However, we do *not* assume that training symbols are identically distributed. In particular, the variance $E[|s_{N,k}(j)|^2]$ of the $j$th training symbol depends on $j$. As we consider the OFDMA context, a certain number of subcarriers may not be modulated by user $k$. If $j$ is one of these subcarriers, we simply consider that $E[|s_{N,k}(j)|^2] = 0$. Furthermore, training sequences $(s_{N,k}(j))_j$ and $(s_{N,k}(j))_j$ transmitted by two different antennas $t$ and $t'$ of a given user $k$ are possibly correlated (due to the possible use of a beamformer). Therefore, the cross-correlation $E[s_{N,k}(j)x_{N,k}(j)^*]$ may be nonzero, where $x^*$ denotes the conjugate of $x$.

**Assumption 2:** Training sequences sent by two different users $k \neq l$ are independent.

**Assumption 3:** 16th-order moments\(^1\) of random variables $(s_{N,k}(j))_j$ are uniformly bounded, i.e.,

$$\sup_N \max_j E\left[ |s_{N,k}(j)|^{16} \right] < M$$

for each $t$, where $M$ is a constant independent of $N$.

\(^1\)the assumption is somewhat stronger than in [11] where it was only assumed that the 8th-order moments are bounded. In this paper, stronger Assumption 3 is needed for the purpose of the asymptotic analysis of the proposed estimate.
We now introduce the following tool in order to be able to characterize the limit of matrices $U_{N,k,l}$ as $N$ tends to infinity. For each $k = 1, \ldots, K$, define the following matrix-valued measure [13] $\mu_{N,k}$ defined for any Borel set $A$ of $[0, 1]$ by

$$\mu_{N,k}(A) = \frac{1}{N} \sum_{j=0}^{N-1} E \left[ s_{N,k}(j) s_{N,k}(j)^H \right]^* \mathcal{I}_A \left( \frac{j}{N} \right),$$

where $\mathcal{I}_A$ stands for the indicator function of set $A$ (i.e., $\mathcal{I}_A(f) = 1$ if $f \in A$, $\mathcal{I}_A(f) = 0$ otherwise) and where vector $s_{N,k}(j) = [s_{N,k}^{(1)}(j), \ldots, s_{N,k}^{(N_T)}(j)]^T$ contains training symbols sent by all antennas of user $k$ at a given subcarrier $j$. We assume as in [11] that

**Assumption 4:** For each $k$, there is a matrix-valued measure $\mu_k$ such that $\mu_{N,k}$ converges weakly to $\mu_k$ as $N$ tends to infinity.

We are now able to study the asymptotic behavior of matrices $U_{N,k,l}$ as $N$ tends to infinity.

**Lemma 1:** Define vector $e(f) = [1, e^{2\pi f}, \ldots, e^{2\pi f(L-1)}]^T$ for each $f \in [0, 1]$. For each $k, l = 1, \ldots, K$

$$\lim_{N \to \infty} U_{N,k,l} = \delta(k-l)R_k \; a.s.$$  \hspace{1cm} (13)

where notation $a.s.$ stands for almost surely and where coefficient $\delta(k-l)$ is equal to 1 if $k = l$ and to zero otherwise. Here, matrix $R_k$ denotes the following $LNT \times LNT$ matrix

$$R_k = \int_0^1 \mu_k(df) \otimes [e(f)e(f)^H].$$  \hspace{1cm} (14)

We refer to [11] for a proof of the above lemma. Using (10), the following result is the immediate consequence of Lemma 1.

**Lemma 2:** For each $\hat{\omega}$, $\frac{1}{N} Q^H_N(\hat{\omega})Q_N(\hat{\omega})$ converges almost surely to a deterministic matrix $R$ which is independent of $\hat{\omega}$ and defined by

$$R = \text{diag} \left( I_{N_T} \otimes R_1, \ldots, I_{N_T} \otimes R_K \right).$$  \hspace{1cm} (15)

Thanks to the above lemma, we are now able to provide a relevant approximation of the exact ML criterion.

**C. Proposed estimate of frequency offsets**

Our aim is now to make use of Lemma 2 in order to propose a relevant and simple approximation of criterion $J^\text{ML}_N(\hat{\omega})$ defined by (9). We recall that the main obstacle in the computation of the exact ML estimate is the calculation of the inverse matrix $\left( Q^H_N(\hat{\omega})Q_N(\hat{\omega}) \right)^{-1}$ in (9). Using Lemma 2, it is straightforward to show that $\left( \frac{1}{N} Q^H_N(\hat{\omega})Q_N(\hat{\omega}) \right)^{-1}$ converges a.s. to $R^{-1}$ as $N$ tends to infinity. Therefore, a simple idea would be to simply replace the inner factor of the righthand side of (9) by its limit $R^{-1}$.
Following such an idea, the corresponding suboptimal estimate of $\omega$ would coincide with the maximum of the following cost function

$$y_N^H Q_N(\hat{\omega}) R^{-1} Q_N^H(\hat{\omega}) y_N.$$  \hfill (16)

Note that matrix $R^{-1}$ can be calculated beforehand. Thus, such a criterion does not involve any matrix inversion during the estimation step. Estimation of $\omega$ becomes practical. Unfortunately, as discussed in the sequel, the performance of such an estimate is far from achieving the performance of the exact ML estimate, even in the asymptotic regime (i.e., when $N$ is large). In particular, a detailed asymptotic analysis would reveal that it is not asymptotically efficient. This unfortunate behavior is due to the fact that the substitution of $R$ with matrix $\frac{1}{N} Q_N^H(\hat{\omega}) Q_N(\hat{\omega})$ is actually a too sharp approximation. In the sequel, we propose a finer estimate.

Due to Lemma 2, matrix $Q_N^H(\hat{\omega}) Q_N(\hat{\omega})$ verifies:

$$\frac{1}{N} Q_N^H(\hat{\omega}) Q_N(\hat{\omega}) = R + E_N(\hat{\omega}),$$  \hfill (17)

where $E_N(\hat{\omega})$ converge almost surely to zero as $N$ tends to infinity. Intuitively, since $E_N(\hat{\omega})$ is close to zero for sufficiently large values of $N$, it is reasonable to approximate matrix $(R + E_N(\hat{\omega}))^{-1}$ by its first order expansion [14] $R^{-1} - R^{-1} E_N(\hat{\omega}) R^{-1}$ and to construct a “simplified” ML criterion by substituting this first order expansion with the inner factor of (9). The proposed cost function is thus defined as follows:

$$J_N(\hat{\omega}) = y_N^H Q_N(\hat{\omega}) \left( R^{-1} - R^{-1} E_N(\hat{\omega}) R^{-1} \right) Q_N^H(\hat{\omega}) y_N$$  \hfill (18)

for each trial vector $\hat{\omega}$, where $E_N(\hat{\omega})$ is defined by (17). The proposed estimate of $\omega$ can be obtained by maximization of $J_N(\hat{\omega})$. For large values of $N$, it is reasonable to believe that this novel criterion is nearly equal to the ML criterion. Using the definition of $E_N(\hat{\omega})$, we now simplify the expression of $J_N(\hat{\omega})$.

$$J_N(\hat{\omega}) = y_N^H Q_N(\hat{\omega}) \left( 2R^{-1} - \frac{1}{N} R^{-1} Q_N^H(\hat{\omega}) Q_N(\hat{\omega}) R^{-1} \right) Q_N^H(\hat{\omega}) y_N$$

$$= -N \mathcal{J}_N(\hat{\omega}) + N y_N^H y_N,$$  \hfill (19)

where

$$\mathcal{J}_N(\hat{\omega}) = \left\| \left( \frac{1}{N} Q_N(\hat{\omega}) R^{-1} Q_N^H(\hat{\omega}) - I_{NN} \right) y_N \right\|^2.$$  \hfill (20)

Note that the maximization of (18) is of course equivalent to the minimization of (20). Finally, the proposed estimate of the set of frequency offsets is obtained as follows.

**Proposed estimate**:

$$\hat{\omega}_N = \arg \min_{\hat{\omega}} \mathcal{J}_N(\hat{\omega}).$$  \hfill (21)
where $J_N(\tilde{\omega})$ is given by (20).

Remark 2: A whole family of criteria can be derived from the same kind of idea. Generally speaking, instead of using the first order expansion of the inverse matrix $(R + E_N(\tilde{\omega}))^{-1}$, one may approximate this inversion with its $Mth$ order expansion $(R + E_N(\tilde{\omega}))^{-1} = \sum_{m=0}^{M} (-1)^m (R^{-1}E_N(\tilde{\omega}))^m R^{-1}$. Then, after some algebra,

$$J_N^{(M)}(\tilde{\omega}) = (-1)^M y_N^H \left( \frac{1}{N} Q_N(\tilde{\omega}) R^{-1} Q_N^H(\tilde{\omega}) - I_{NN} \right)^{M+1} y_N. \quad (22)$$

We refer to [15] for the detailed derivation of the above expression of $J_N^{(M)}(\tilde{\omega})$. Intuitively, if $M > M'$, the criterion $J_N^{(M)}$ is expected to lead to a better performance than $J_N^{(M')}$, because the $Mth$-order expansion of $(\frac{1}{N} Q_N^H(\tilde{\omega}) Q_N(\tilde{\omega}))^{-1}$ is a more accurate approximation than the $M'$th-order expansion. In the sequel, we will mainly focus on the most simple one of these criteria, namely, the criterion $J_N = J_N^{(1)}$ obtained with $M = 1$ and given by (20). If this simple criterion is shown to perform well, it is a fortiori reasonable to conjecture that the other (more involved) criteria also do. This claim will be confirmed by simulations.

D. Proposed estimate of channel coefficients

Once vector $\omega$ has been obtained via the above minimization of $J_N$, it is straightforward to compute an estimate of channel coefficients using a relation similar to (8). The most immediate way to estimate is of course to use directly expression (8) with $\tilde{\omega} = \hat{\omega}_N$, where $\hat{\omega}_N$ is the proposed estimate (21) of frequency offsets. Such a procedure is of course likely to be implemented and provides very satisfying results. However, it still requires the computation of matrix $(Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N))^{-1}$. In certain situations, this can still be too computationally demanding, especially when the number $K$ of users and/or the number $L$ of channel coefficients is significant even for SISO case.

Using the same idea as in the previous Section, we propose a simpler channel estimate based on the approximation of the above matrix $(Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N))^{-1}$. As in the previous section, we replace $(Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N))^{-1}$ by its first order expansion

$$\left( \frac{1}{N} Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N) \right)^{-1} \simeq R^{-1} - R^{-1} \left( \frac{1}{N} Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N) - R \right) R^{-1}. \quad (23)$$

Based on (8), this leads to the following estimate of channel coefficients:

$$\hat{h}_N = \frac{1}{N} \left( 2R^{-1} - \frac{1}{N} R^{-1} Q_N^H(\tilde{\omega}_N) Q_N(\tilde{\omega}_N) R^{-1} \right) Q_N^H(\tilde{\omega}_N) y_N. \quad (23)$$

Of course, matrix $R^{-1}$ can be calculated beforehand, so that in practice, no matrix inversion is required for the computation of the above channel estimate. Denoting by $\hat{h}_{N,k}^{(r)}$ the estimate of $h_{k}^{(r)}$ for each antenna
r = 1, . . . , NR of each user k = 1, . . . , K, equation (23) is equivalent to
\[
\hat{h}_{N,k}^{(r)} = \frac{1}{N} \mathbf{R}_k^{-1} \mathbf{A}_{N,k}^H \Gamma_N^H (\hat{\omega}_N) \left( 2 \mathbf{I}_N - \sum_{l=1}^{K} \mathbf{T}_N (\hat{\omega}_{N,l}) \right) y_N^{(r)},
\]
(24)
where \( \mathbf{T}_N (\hat{\omega}_{N,l}) = \frac{1}{N} \Gamma_N (\hat{\omega}_{N,l}) \mathbf{A}_{N,l} \mathbf{R}_l^{-1} \mathbf{A}_{N,l}^H \Gamma_N^H (\hat{\omega}_{N,l}) \). It is worth mentioning that, using an approach similar to the previous paragraph, the inverse matrix \( (\mathbf{Q}_N^H (\hat{\omega}_N) \mathbf{Q}_N (\hat{\omega}_N))^{-1} \) can as well be approximated with it Mth order expansion with the aim of constructing an even more accurate estimate of the channel coefficient.

V. ASYMPTOTIC STUDY OF THE PROPOSED ESTIMATOR

We now study the asymptotic behavior of the estimation error associated with the frequency offsets \( \hat{\omega}_{N,k} - \omega_k \) and with the channel coefficients \( \hat{\mathbf{h}}_{N,k} - \mathbf{h}_k \) as the number \( N \) of subcarriers tends to infinity. The estimates of the frequency offsets and the channel coefficients are respectively given by (21) and (24). As the estimate of the channel is a simple function (24) of the estimate of the frequency offsets, it is natural to focus on \( \hat{\omega}_N \) at first. Once the asymptotic behavior of \( \hat{\omega}_N \) has been characterized, the asymptotic study of channel estimates becomes possible.

A. The main result

The main result is given in the following theorem. Here, we denote by \( \hat{\theta}_{N,k} \) the estimate of the real parameter vector \( \theta_k = [\omega_k, \mathbf{h}_{k,R}^T, \mathbf{h}_{k,I}^T]^T \) and by \( \hat{\theta}_N = [\theta_{N,1}^T, \ldots, \theta_{N,K}^T]^T \) the estimate of the whole \( K(1+2LN_RN_T) \times 1 \) parameter vector \( \theta = [\theta_1^T, \ldots, \theta_K^T]^T \) including all users. In the above definition, \( \mathbf{h}_{k,R} \) and \( \mathbf{h}_{k,I} \) respectively represent the real and the imaginary parts of vector \( \mathbf{h}_k \). We denote the whole set of training symbols of all users by \( \mathbf{s}_N = [\mathbf{s}_{N,1}^T(0), \ldots, \mathbf{s}_{N,R}^T(N-1), \ldots, \mathbf{s}_{N,K}^T(0), \ldots, \mathbf{s}_{N,K}^T(N-1)]^T \). In the sequel, we study the asymptotic behavior of the normalized estimation error vector \( \mathbf{W}_N (\hat{\theta}_N - \theta) \), where \( \mathbf{W}_N \) is the \( K(1+2LN_RN_T) \times K(1+2LN_RN_T) \) diagonal normalization matrix \( \mathbf{W}_N = \text{diag}(\mathbf{w}_N^T, \ldots, \mathbf{w}_N^T) \) where \( \mathbf{w}_N^T \) denotes the row vector \( \mathbf{w}_N^T = [N^{3/2}, N^{1/2}, \ldots, N^{1/2}] \) of length \( 1+2LN_RN_T \).

Theorem 1: For almost any realization of the training symbols \( \mathbf{s}_N \), the normalized estimation error \( \mathbf{W}_N (\hat{\theta}_N - \theta) \) converges in distribution to a Gaussian random vector of zero mean and covariance matrix \( \Sigma \) defined by \( \Sigma = \text{diag}(\Sigma_1, \ldots, \Sigma_K) \) where for each \( k, \Sigma_k \) is the \( (1+2LN_RN_T) \times (1+2LN_RN_T) \) matrix equal to

\[
\Sigma_k = \frac{\sigma_k^2}{2} \begin{bmatrix}
\frac{12}{\gamma_k} & \frac{6 h_{k,R}^T}{\gamma_k} & \frac{-6 h_{k,R}^T}{\gamma_k} \\
\frac{6 h_{k,R}^T}{\gamma_k} & 3 \text{Re} \left[ (\mathbf{I}_{N_R} \otimes \mathbf{R}_k^{-1}) \right] & -3 \text{Im} \left[ (\mathbf{I}_{N_R} \otimes \mathbf{R}_k^{-1}) \right] \\
\frac{-6 h_{k,R}^T}{\gamma_k} & -3 \text{Im} \left[ (\mathbf{I}_{N_R} \otimes \mathbf{R}_k^{-1}) \right] & \frac{6 h_{k,R}^T}{\gamma_k}
\end{bmatrix},
\]
(25)

November 25, 2006
More precisely, the conditional distribution of $W_N (\hat{\theta}_N - \theta)$ w.r.t. $s_N$ converges to the Gaussian distribution of zero mean and covariance matrix $\Sigma$ with probability one. The proof of Theorem 1 is provided in Section V-C. We now make the following comments.

**Comments**

- The proposed estimate is almost surely asymptotically normal.
- Theorem 1 implies in particular that for each user $k$,

$$\lim_{N \to \infty} N^3 E_N \left[ (\omega_{N,k} - \omega_k)^2 \right] = \frac{6\sigma^2}{\gamma_k} \ a.s. \quad (26)$$

$$\lim_{N \to \infty} N E_N \left[ \|h_{N,k} - h_k\|^2 \right] = N R \sigma^2 \text{tr} \left( R_k^{-1} \right) + \frac{3\sigma^2 h_k^H h_k}{2 \gamma_k} \ a.s. \quad (27)$$

where $E_N[\cdot]$ designates the conditional expectation w.r.t. the training sequence $s_N$. In particular, the MSE associated with the frequency offset of a given user $k$ tends to zero at convergence speed $\frac{1}{N^3}$, while the MSE associated with the channel coefficients tends to zero at convergence speed $\frac{1}{N}$.

- For large values of $N$, the covariance matrix of the normalized estimation error $W_N (\hat{\theta}_N - \theta)$ converges to a block diagonal matrix. This means that the estimation errors on the parameters of two different users $k \neq l$ become independent as $N$ tends to infinity. Furthermore, it is worth noting that the mean square errors corresponding to the parameters of a given user $k$ does not depend on the number $K$ of users. Therefore, as long as $N$ is large enough, the performance of the proposed estimate is not affected by the presence of other users: it is identical to the performance that one would have observed if only one user was transmitting.

**B. Asymptotic efficiency of the proposed estimate**

The proposed estimate of $\omega$ is suboptimal compared to the rigorous ML estimate, because it is based on an approximation of the log-likelihood function. However, it is explained in this subsection that the performance of the estimate is asymptotically optimal. More precisely, when the number $N$ of subcarriers is large enough, there is no degradation of the performance when using the proposed suboptimal estimate instead of the ML estimate. This result is the immediate consequence of Theorem 1 along with the results of [11].

Denote by $\text{CRB}_N(s_N)$ the Cramér-Rao Bound (CRB) associated with $\theta$. We recall that the CRB can be interpreted as a lower bound on the covariance matrix of any unbiased estimate $\hat{\theta}_N$ of $\theta$:

$$E_N \left[ (\hat{\theta}_N - \theta)(\hat{\theta}_N - \theta)^T \right] \succ \text{CRB}_N(s_N), \quad (28)$$
where notation $A \succeq B$ means that the matrix $A - B$ is a non-negative square matrix. In [11], the following result has been shown:

$$\lim_{N \to \infty} W_N \text{CRB}_N(s_N) W_N = \Sigma \quad \text{a.s.}$$

(29)

where $\Sigma$ is precisely the matrix defined in Theorem 1. In other words, the limit $\Sigma$ of the covariance matrix of the estimation error corresponding to the proposed estimate coincides with the asymptotic CRB. As a result, the proposed estimate is almost surely asymptotically efficient. This remark is of practical importance. Indeed, in spite of the fact that the proposed estimator is suboptimal compared to the ML estimator, it turns out that its performance becomes equal to the performance of the ML algorithm when $N$ increases. There is no loss of optimality when using the proposed simple estimator instead of the rigorous ML algorithm, provided that $N$ is large enough.

C. Proof of Theorem 1

In the present section, we provide the proof of Theorem 1.

1) Step 1. Consistency of the estimate of frequency offsets: The first step of the asymptotic analysis consists in showing that the estimate of $\hat{\omega}_N$ converges to $\omega$ provided that the number of training symbols increases. This property is established by Theorem 2 below. Note that it is necessary to prove the consistency of the proposed estimate in order to be able to further study the asymptotic behavior of the estimation error. Due to the lack of space, the following statement is not proved in the present paper but the proof is available in [15]. Note that the proof is based on results previously introduced by [16].

Theorem 2: The following two properties hold:

$$\lim_{N \to \infty} \hat{\omega}_N = \omega \quad \text{a.s.}$$

$$\lim_{N \to \infty} N(\hat{\omega}_N - \omega) = 0 \quad \text{a.s.}$$

Here, recall that notation a.s. stands for almost surely.

2) Step 2. Asymptotic study of the estimate of frequency offsets: We now study the behavior of the estimation error $\hat{\omega}_N - \omega$ as the number $N$ of subcarriers tends to infinity. According to Assumption 4, for each Borel set $A$ in $[0, 1]$, $\mu_{N,k}(A) \to \mu_k(A)$ as $N$ tends to infinity. In order to say that our asymptotic results are valid, we need furthermore to ensure the convergence of $\mu_{N,k}(A)$ to $\mu_k(A)$ “quickly enough” (that is, at least at speed $\frac{1}{\sqrt{N}}$).

Assumption 5: For each $k$ and $(t, t')$, denote by $|\mu_{N,k}^{(t, t')} - \mu_k^{(t, t')}|$ the total variation of $\mu_{N,k}^{(t, t')} - \mu_k^{(t, t')}$. We assume that $\sqrt{N}|\mu_{N,k}^{(t, t')} - \mu_k^{(t, t')}|([0, 1]) \leq C$ for certain constant $C$. 

November 25, 2006 DRAFT
Due to [17], this implies that for a given function $F$ continuous on $[0,1]$, 
\[
\sqrt{N} \left( \int F(f) \mu_N^{(t,t')} (df) - \int F(f) \mu_k^{(t,t')} (df) \right) \leq C',
\]
where $C'$ is a certain constant (which depends on $F$). Note that such a technical assumption encompasses usual subcarrier and power allocation schemes.

We now study the asymptotic behavior of the estimation error. We define the gradient vector of cost function $J_N(\omega)$ as $\nabla J_N(\omega) = [\frac{\partial J_N(\omega)}{\partial \omega_1}, \ldots, \frac{\partial J_N(\omega)}{\partial \omega_K}]^T$ and the Hessian matrix associated with $J_N(\omega)$ as the $K \times K$ matrix $H_N(\omega) = \left[ \frac{\partial^2 J_N(\omega)}{\partial \omega_k \partial \omega_l} \right]_{k,l=1,\ldots,K}$. We then make use of the Taylor-Lagrange expansion of $\nabla J_N(\omega)$ at the true value $\omega$ of the parameter vector. There exists a real number $\hat{\omega}_N$ such that the (normalized) estimation error $N\sqrt{N} (\hat{\omega}_N - \omega)$ can be written as
\[
N\sqrt{N} (\hat{\omega}_N - \omega) = - \left( \frac{1}{N^3} H_N(\hat{\omega}_N) \right)^{-1} \left( \frac{1}{N\sqrt{N}} \nabla J_N(\omega) \right),
\]
where $\hat{\omega}_N$ can be written as $\hat{\omega}_N = \lambda \hat{\omega}_N + (1 - \lambda) \omega$ for a certain $\lambda \in [0,1]$. Therefore, the asymptotic analysis of the estimation error reduces to the separate study of the gradient vector and the Hessian matrix associated with $J_N$. We now provide the following lemmas.

**Lemma 3:** The gradient vector of $J_N$ at point $\omega = \omega$ verifies
\[
\frac{1}{N\sqrt{N}} \nabla J_N(\omega) = - \zeta_N + o_{as}(1),
\]
where components of random vector $\zeta_N = [\zeta_{N,1}, \ldots, \zeta_{N,K}]^T$ are defined for each $k = 1,\ldots,K$ by
\[
\zeta_{N,k} = \text{Im} \left[ h_k^H \left( I_{N_r} \otimes \frac{A_{N,k}^H (I_N - \frac{2}{N} D_N) \Gamma_{N}^{H} (\omega_k)}{\sqrt{N}} \right) \nu_N \right],
\]
where $D_N = \text{diag}(0,1,\ldots,N-1)$.

**Lemma 4:** The normalized Hessian matrix of $J_N$ at point $\omega_N$ converges almost surely to a diagonal matrix as $N$ tends to infinity:
\[
\lim_{N \to \infty} \frac{1}{N^3} H_N(\omega_N) = \frac{1}{6} \text{diag} (\gamma_1,\ldots,\gamma_K) \text{ a.s.}
\]
where for each $k = 1,\ldots,K$, $\gamma_k$ represents the following deterministic constant:
\[
\gamma_k = h_k^H (I_{N_r} \otimes R_k) h_k.
\]

The proofs of the above lemmas are provided in the Appendices II and III. We now give some insights on the consequences of these lemmas. Lemmas 3 and 4 indicate that the normalized estimation error $N\sqrt{N} (\hat{\omega}_N - \omega)$ has the same asymptotic behavior as vector $[\frac{6}{\gamma_k} \zeta_{N,1}, \ldots, \frac{6}{\gamma_k} \zeta_{N,K}]^T$. Indeed, it is straightforward to show from Lemmas 3 and 4 that
\[
N\sqrt{N} (\hat{\omega}_{N,k} - \omega_k) = \frac{6}{\gamma_k} \zeta_{N,k} + o_{as}(1).
\]
Note that for each \( k \), \( \zeta_{N,k} \) depends only on the training sequence of user \( k \) (via matrix \( \mathbf{A}_{N,k} \)). Recall that the training sequences of two distinct users \( k \neq l \) are independent. Therefore, we already have the insight that the estimation errors \( N \sqrt{N} (\hat{\omega}_{N,k} - \omega_k) \) and \( N \sqrt{N} (\hat{\omega}_{N,l} - \omega_l) \) corresponding to users \( k \) and \( l \) respectively become independent as \( N \) tends to infinity.

3) Step 3. Asymptotic study of the estimate of channel coefficients: We now study the asymptotic behavior of the estimation error associated with the proposed estimate of channel coefficients given by (24). We first focus on the estimate of the channel coefficients \( \mathbf{h}_k^{(r)} \) corresponding to a given user \( k \) and associated with a given receive antenna \( r \). For the purpose of the asymptotic analysis, it is convenient to write \( \mathbf{h}_k^{(r)} \) as follows:

\[
\mathbf{h}_k^{(r)}(r) = \mathbf{h}_k^{(r)} + \frac{\mathbf{R}_k^{-1}}{\sqrt{N}} \left[ \frac{\mathbf{A}_{N,k}^H \Gamma_{N}^H (\omega_{N,k}) \mathbf{v}_N^{(r)}}{\sqrt{N}} + \sum_{l=1}^{K} \Delta_{N,k,l} \mathbf{h}_l^{(r)} + \epsilon_{N,k}^{(r)} \right].
\]

Here, the quantities \( \Delta_{N,k,l} \) and \( \epsilon_{N,k}^{(r)} \) are given by \( \Delta_{N,k,l} = \mathbf{P}_{N,k,l}(\omega_l - \hat{\omega}_{N,k}) - \mathbf{P}_{N,k,l}(\hat{\omega}_{N,l} - \hat{\omega}_{N,k}) \) and

\[
\epsilon_{N,k}^{(r)} = -\frac{\mathbf{R}_k^{-1}}{\sqrt{N}} \sum_{l=1}^{K} \mathbf{P}_{N,k,l} (\hat{\omega}_{N,l} - \hat{\omega}_{N,k}) \mathbf{R}_l^{-1} \left( \sum_{l'=1}^{K} \mathbf{P}_{N,l,l'} (\hat{\omega}_{N,l'} - \hat{\omega}_{N,l}) \mathbf{h}_{l'}^{(r)} + \frac{\mathbf{A}_{N,l}^H \Gamma_{N}^H (\hat{\omega}_{N,l}) \mathbf{v}_N^{(r)}}{\sqrt{N}} \right)
\]

where, for each \( x \), \( \mathbf{P}_{N,k,l}(x) = \sqrt{N} \left( \frac{1}{N} \mathbf{A}_{N,k}^H \Gamma_{N}(x) \mathbf{A}_{N,l} - \delta(k-l) \mathbf{R}_k \right) \). Using an approach similar to Appendix II, it can be shown that \( \epsilon_{N,k}^{(r)} \) converges almost surely to zero as \( N \) tends to infinity. We now further simplify the terms \( \Delta_{N,k,l} \). To that end, we provide the following lemma.

**Lemma 5:** For each \( k, l = 1, \ldots, K, \)

\[
\Delta_{N,k,l} = -\delta(k-l) \frac{3i \zeta_{N,k}}{\gamma_k} \mathbf{R}_k + o_{\text{as}}(1).
\]

The proof of the above lemma is omitted due to lack of place but a sketch of the proof is available in [15]. It can also be shown using a similar proof that vector \( \frac{1}{\sqrt{N}} \mathbf{R}_k^{-1} \mathbf{A}_{N,k}^H \Gamma_{N}^H (\hat{\omega}_{N,k}) \mathbf{v}_N^{(r)} \) has the same asymptotic behavior as \( \frac{1}{\sqrt{N}} \mathbf{R}_k^{-1} \mathbf{A}_{N,k}^H \Gamma_{N}^H (\omega_k) \mathbf{v}_N^{(r)} \). We finally obtain the following expression of the normalized estimation error.

\[
\sqrt{N} (\hat{\mathbf{h}}_{N,k}^{(r)} - \mathbf{h}_k^{(r)}) = \frac{\mathbf{R}_k^{-1} \mathbf{A}_{N,k}^H \Gamma_{N}^H (\omega_k)}{\sqrt{N}} \mathbf{v}_N^{(r)} - \frac{3i \zeta_{N,k}}{\gamma_k} \mathbf{h}_k^{(r)} + o_{\text{as}}(1).
\]

Due to the presence of random variable \( \zeta_{N,k} \) in both expressions (34) and (35), we have the insight that the estimation error corresponding to channel coefficients \( \mathbf{h}_k \) is correlated to the estimation error corresponding to frequency offset \( \omega_k \). The last step of the asymptotic analysis consists in putting all pieces (34) and (35) together, and to provide a compact asymptotic expression of the performance associated with the whole set of parameters \([\omega, \mathbf{h}]\).
4) Step 4. Asymptotic performance of the global estimate: Our aim is to study the asymptotic behavior of the vector-valued estimation error \( \hat{\theta}_N - \theta \). Equations (34) and (35) suggest to rather study the normalized estimation error \( W_N \left( \hat{\theta}_N - \theta \right) \). Using (34) and (35), it is straightforward to obtain an asymptotically equivalent expression of the normalized estimation error \( W_N \left( \hat{\theta}_N - \theta \right) \). Replacing \( \zeta_{N,k} \) with its definition (31), we obtain directly that

\[
W_N \left( \hat{\theta}_N - \theta \right) = \text{Im} \left[ Z_N v_N \right] + o_{\text{as}}(1),
\]

where \( Z_N \) is the following block matrix:

\[
Z_{N,k} = \begin{bmatrix}
\frac{6}{7} \Phi_{N,k} \\
\frac{3}{7} h_k \Phi_{N,k} \\
\Psi_{N,k} - \frac{3}{7} h_k R h_k H \Phi_{N,k} \\
\end{bmatrix},
\]

and where for each \( k \),

\[
\Phi_{N,k} = (\frac{2}{7} R_k \Gamma^H_N(\omega_k) \Phi_{N,k} + \frac{3}{7} R_k (I_N - 2\frac{2}{7} D N) \Gamma^H_N(\omega_k)) \sqrt{N},
\]

\[
\Psi_{N,k} = (\frac{2}{7} R_k \Gamma^H_N(\omega_k) \Phi_{N,k} - \frac{3}{7} R_k h_k R h_k H \Phi_{N,k} ) \sqrt{N}.
\]

We deduce from (35) that the normalized estimation error \( W_N \left( \hat{\theta}_N - \theta \right) \) has the same asymptotic behavior as random vector \( \text{Im} \left[ Z_N v_N \right] \). It can be easily shown that, for a fixed value of training symbols (i.e., a fixed value of matrices \( A_{N,k} \)), the latter \( \text{Im} \left[ Z_N v_N \right] \) has a Gaussian distribution. In other words, the conditional distribution of \( \text{Im} \left[ Z_N v_N \right] \) w.r.t. the set of training symbols \( s_N \) is a Gaussian distribution (whose covariance matrix depends on \( s_N \)). This claim can be easily shown by noticing that for a fixed value of \( s_N \), \( \text{Im} \left[ Z_N v_N \right] \) is a linear function of random vectors \( \Gamma^H_N(\omega_k) v^{(r)}_N \) for all \( r = 1, \ldots, N_R \). Since vector \( v^{(r)}_N \) has i.i.d. zero mean complex circular Gaussian entries, then the above vectors have as well i.i.d. zero mean complex circular Gaussian entries. Thus, \( \text{Im} \left[ Z_N v_N \right] \) is Gaussian distributed. Consequently, for a fixed value of training symbols \( s_N \), \( \text{Im} \left[ Z_N v_N \right] \) converges in distribution to Gaussian distributed random vector whose covariance matrix \( \Sigma \) coincides with the limit of the covariance matrix of \( \text{Im} \left[ Z_N v_N \right] \). In other words,

\[
\Sigma = \lim_{N \to \infty} E \left[ \left. \text{Im} \left[ Z_N v_N \right] \text{Im} \left[ Z_N v_N \right]^T \right| s_N \right].
\]

The last technical point is to calculate the above limit to prove Theorem 1. This is done in Appendix IV.

VI. PRactical implementation and Computational Complexity

In order to implement the estimator of \( \omega \) based on criterion \( J_N \) (or on its generalized version \( J^{(M)}_N \) given in (22)), we propose to make use of a Newton search algorithm. The frequency offsets estimate \( \hat{\omega}_N \) at the
ith iteration is updated as follows: \( \hat{\omega}_N^{i+1} = \hat{\omega}_N^i + H_N^{-1}(\hat{\omega}_N^i) \nabla J_N(\hat{\omega}_N^i) \), where \( H_N^{-1}(\hat{\omega}_N^i) \) and \( \nabla J_N(\hat{\omega}_N^i) \) respectively represent the inverse Hessian matrix and the gradient vector associated with \( J_N \). Of course, a more simple gradient search algorithm is also likely to be used in order to avoid the computation of the inverse Hessian matrix. The above Newton (or gradient) search algorithm requires a relevant initial estimate \( \hat{\omega}_N^0 \) in order to converge toward the desired value. For example, such an initial estimate can be obtained by minimizing criterion \( J_N(\tilde{\omega}) \) w.r.t. \( \tilde{\omega} \) on a grid. Simpler criteria such as the suboptimal criterion (16) or the “single user” estimator of [18] can also be proposed for this task. We refer to [10] for a discussion on such issues. We now briefly investigate the complexity of the above Newton search algorithm and compare the latter with existing estimates. Gradient vector \( \nabla J_N \) can be computed using the formula given in Appendix II. We first remark that the matrix \( R^{-1} \) which is involved in this expression can be computed beforehand, so that it generates no additional computational burden. It can be shown that at each iteration of the Newton algorithm, the evaluation of the Gradient and the Hessian matrix both require less than \( O(NK^2(LN_T)^2) \) operations. The overall complexity of the proposed algorithm is thus bounded by \( O(n_i(NK^2(LN_T)^2 + K^3)) \), where \( n_i \) denotes the number of iterations of the Newton algorithm and where the presence of the term \( K^3 \) is due to the inversion of the Hessian matrix. Note that in practice, only a few iterations are needed (typically \( n_i = 2, 3, 4 \)). In the case where \( M > 1 \), the complexity analysis of the estimator based on the generalized criterion \( J_N^{(M)} \) is more difficult. However, simulations indicate that the computational burden remains at the same order at least for \( M = 3 \) and 5.

It is interesting to compare the complexity of the proposed algorithm with existing estimators. As the rigorous ML estimator is known to be impractical anyway, it is more interesting to draw a comparison with respect to more practical estimators such as the ML-based algorithm proposed by [10]. In this paper, authors use the so-called alternating projection frequency estimation (APFE) procedure in order to estimate the argument of the maximum of a simplified version of the log-likelihood criterion. The performance of such an algorithm is shown to be very attractive. Unfortunately, the computational burden of the APFE method is still very significant so that this estimator may be impractical in certain situations, especially for large values of \( N \). For each iteration and for each user \( k = 1, \ldots, K \), the APFE algorithm requires to evaluate a certain cost function (which depends on \( k \)) for each \( \tilde{\omega}_k \) on a grid of size \( n_g \). It can be shown that the complexity of the APFE method is about \( O(n_i(n_gK + K^4)(N(LN_T)^2 + (LN_T)^3)) \). In practice, the complexity of the above algorithm depends on the number \( n_g \) of points in the grid which can be considerable, especially when \( N \) is large.

Note finally that the proposed channel estimate can as well be implemented with low computational cost. Unlike the estimators [9][10] based on expression (8) and which require the inversion of a \( KN_R N_T L \times (26) \).
\( K N_R N_T L \) matrix, the proposed channel estimate \((23)\) avoids such an inversion.

VII. Simulation Results

We consider an uplink OFDMA system with QPSK signaling. We investigate either the SISO case or the MIMO case with \( N_R = N_T = 2 \). For each user and for each transmit-receive antenna pair, we consider a multipath fading channel with 8 independent paths. Complex gains associated with each path are assumed to be circular complex Gaussian random variables with zero mean and unit variance. For each user \( k \), the value of \( \delta f_k \) is randomly chosen in the interval \([\frac{-0.4}{N_T}, \frac{0.4}{N_T}]\). In the sequel, without loss of generality, we focus on the results corresponding to the first user \( k = 1 \) and suppose that average transmitted powers \( P_k \) are equal for all users. All results are averaged over 1000 realizations of the training sequences, the frequency offsets and the channel parameters. The SAS which is considered in the simulations is the following. We assume that the total bandwidth is divided into groups of four consecutive subcarriers. Each group is modulated by only one user. Groups of subcarriers are assigned to the different users following an equispaced assignment scheme. For instance, user 1 modulates the first group, the \((K + 1)\)th group, the \((2K + 1)\)th group, etc. All subcarriers are modulated with equal power and training sequences sent at different antennas are uncorrelated. The proposed estimates of the frequency offsets are obtained by minimization of criterion \( J(M) \) defined by \((22)\) for \( M = 1 \) and \( M = 3 \). The estimates are computed using the Newton algorithm depicted in Section VI and are compared with the AFPE proposed by [10]. Both algorithms are used with \( n_i = 2 \) iterations as in [10]. Estimates of channel parameters are obtained using \((23)\) when \( M = 1 \) and using the generalized version of \((23)\) when \( M = 3 \) as discussed in Section IV-D. Finally, the performance is compared with the exact and asymptotic CRB derived in [11].

We first study the behavior of the mean square error (MSE) as the number \( N \) of subcarriers increases. Figure 1(a) represents the (empirical evaluation of the) MSE \( E[(\hat{\omega}_{N,1} - \omega_1)^2] \) as a function of \( N \), when \( \frac{P_k}{\sigma^2} = 10 \)dB and when \( N_R = N_T = 1 \). Here, since the main focus is on the accuracy of the fine search of the parameters, Newton algorithms are initialized with the true value of the parameters. We observe that the proposed estimator fits to the CRB, even for moderate values of \( N \). As expected, the estimation of the parameters based on criterion \( J^{(M)} \) with \( M = 3 \) provides more accuracy compared to \( M = 1 \). This is due to the fact that criterion \( J^{(3)} \) is finer approximation of the log-likelihood criterion. Similar observations can be drawn from Figure 1(b) which represents the MSE of the channel estimate \( E[\|\hat{h}_{N,1} - h_1\|^2] \): again, the estimates fit to the CRBs. This sustains the claim that the proposed estimators are asymptotically efficient.

In the \( 2 \times 2 \) MIMO case, Figures 2(a) and 2(b) indicate that the proposed estimates are as well close to the CRB. However, when the number of antennas or the number of users increases, we observe that the
asymptotic regime is reached for a larger number of subcarriers. Again, the increase in the value of the expansion order $M$ results in a faster convergence to the CRB.

Figures 3, 4 represent the performance as a function of the signal to noise ratio (SNR) $\frac{P_1}{\sigma^2}$ when the number $N$ of subcarriers is constant. Figure 3 is obtained for $N = 1024$ for both SISO and MIMO case and for either 2 or 4 users. It again sustains the theoretical claim that the MSE of the estimate are close to the CRB, especially when $M \geq 3$. However, when $M = 1$, we remark that the performance of the estimate does not fit to the CRB for high values of the SNR $\frac{P_1}{\sigma^2}$. This can be intuitively explained as follows. The asymptotic study shows that the (normalized) estimation error $N\sqrt{N}(\hat{\omega}_N,1 - \omega_1)$ coincides with $rac{\hat{\omega}}{\omega_1}\zeta_{N,1} + o_{as}(1)$. For moderate values of the SNR, the term $o_{as}(1)$ is negligible compared to $\zeta_{N,1}$. However, $\zeta_{N,1}$ is a linear function of the noise vector. Thus, at high SNR, components of the noise vector are close to zero, so that the modulus of $\zeta_{N,1}$ may be very slight: for large but finite $N$, the modulus of $\zeta_{N,1}$ may be about as small as the second term $o_{as}(1)$, which is thus no longer negligible. Therefore, at high SNR, this intuitively indicates that the asymptotic regime is reached for larger values of $N$ compared to the low SNR case.

Finally, in Figure 4, we compare the proposed estimates with the APFE method. Here, both Newton algorithms and the APFE method are initialized using the result of a preliminary coarse search. The simple suboptimal “single user” estimator of [18] is used for this coarse search as suggested in Section VI. The results of the fine search are presented after the elimination of the possible outliers produced by the coarse search. For instance, when $N = 256$ and $\frac{P_1}{\sigma^2} = 15$dB, the percentage of outliers which have been eliminated in order to plot the results is equal to %0.6 for APFE and %0.1 for the proposed method. Note that the number of outliers increases at low SNR. We refer to [3] for a detailed discussion on this issue. Note that the APFE method is difficult to implement for large values of $N$ as discussed in Section VI. Therefore, we consider the case $N = 256$. We also focus on the SISO case in order to implement the APFE method described in [10]. Figure 4(a) represents the MSE corresponding to frequency offset estimates while Figure 4(b) represents the MSE corresponding to channel estimates. As far as the APFE algorithm is concerned, the channel estimation is performed using expression (8) (only replacing $\hat{\omega}$ in the latter with the APFE estimate of $\omega$). Although the value of $N$ is moderate, Figures 4(a) and 4(b) show that when $M = 3$, the proposed estimator performs similarly to the APFE method: both estimators fit to the CRB curve. As discussed previously, the estimator based on the first order expansion $M = 1$ does not fit to the CRB. The asymptotic regime is not reached for $N = 256$ in this range of SNRs. For the sake of completeness, Figures 4(a) and 4(b) also include the MSE of the estimates obtained from the single user estimator [18]. As expected, such estimates are far from achieving the CRB since they are not specially
designed for the multi-user context.

VIII. CONCLUSIONS

A family of estimators of frequency offsets and MIMO channel coefficients has been proposed in the case where $K$ users are transmitting toward a single receiver using an orthogonal frequency division multiple access scheme. The estimators do not rely on a particular SAS and can be used in a general OFDMA context. The implementation of such estimators can be achieved with very reasonable complexity compared to existing estimators based on the maximization of the likelihood function. In addition to their reduced complexity, the estimators are shown to be asymptotically efficient. Indeed, when the number $N$ of subcarriers increases, the covariance matrix of the estimation error becomes identical to the Cramèr-Rao bound. Simulations sustain our theoretical claims and illustrate the attractive performance of the estimators even for moderate values of the number of subcarriers.

APPENDIX I

SKETCH OF THE PROOF OF THE CONSISTENCY OF THE ESTIMATE

The proof is based on the following preliminary lemma.

Lemma 6: Consider a given antenna pair $t, t'$. For each $p, q = 0, \ldots, L - 1$, define

$$ e_{p,q}^{N,k,l}(n) = a_{N,k,t}(n-p)^* a_{N,l,t'}(n-q) - E[a_{N,k,t}(n-p)^* a_{N,l,t'}(n-q)]. $$

Define for each real number $\delta$ and for each $u = 0, 1, 2$,

$$ S_{p,q,u}^{N,k,l}(\delta) = \frac{u + 1}{N u + 1} \sum_{n=0}^{N-1} n^u e_{p,q}^{N,k,l}(n) e^{in\delta}. $$

Then, $\sup_{\delta} \left| S_{p,q,u}^{N,k,l}(\delta) \right|$ converges almost surely to zero as $N$ tends to infinity.

The proof of the above lemma is not provided here due to the lack of space but more details are available in [15]. In the sequel, we focus on the case $N_T = N_R = 1$ but the proof can be extended without any difficulty to the MIMO case. For each $N$, define the estimation error associated with the $k$th user as $\delta_{N,k} = \omega_k - \hat{\omega}_{N,k}$ modulo $2\pi$. Our aim is to prove that for each $k = 1, \ldots, K$, $N\delta_{N,k}$ converges to zero with probability one. To that end, we first prove the following lemma.

Lemma 7: The following result holds:

$$ \lim_{N \to \infty} \frac{1}{N} \left( J_N(\hat{\omega}_N) - J_N(\omega) \right) + \sum_{k=1}^{K} h_k^H R_k h_k \left( 1 - |q_N(\delta_{N,k})|^2 \right) = 0 \ a.s. $$

where $q_N(\delta) = \frac{1}{N} \sum_{n=0}^{N-1} e^{in\delta}$. 

DRAFT November 25, 2006
The proof of the above lemma is based on preliminary Lemma 6. Again, the proof is omitted due to the lack of space, but is available in [15]. Since $|q_N(\delta)| \leq 1$ for every $\delta$, it is clear that both terms of the righthand side of (38) are non-negative numbers. Therefore, both terms converge to zero. In other words, for each $k = 1, \ldots, K$, $h_k^H R_k h_k (1 - |q_N(\delta_{N,k})|^2)$ tends almost surely to zero. As a consequence, for each $k$, we obtain that $\lim_{N \to \infty} q_N(\delta_{N,k}) = 1$ a.s. Following [16], we now make use of the following lemma from [19]:

**Lemma 8:** Consider a real-valued sequence $c_N \in (-\pi, \pi]$ such that $\lim_{N \to \infty} c_N = c$.

- if $c \neq 0$, then $q_N(c_N)$ tends to 0,
- if $c = 0$ and $N|c_N - c| \to \infty$, then $q_N(c_N)$ tends to 0,
- if $c = 0$ and $N|c_N - c| \to \beta \in \mathbb{R}$, then $q_N(c_N)$ tends to $e^{i \frac{\delta \sin(\beta/2)}{\beta/2}}$.

Now consider a sequence $\delta_{N,k}$ such that $q_N(\delta_{N,k})$ tends to 1 as $N$ tends to infinity. As $\delta_{N,k}$ belongs to the bounded interval $(-\pi, \pi]$, there exist a convergent subsequence $\delta_{\varphi(N),k}$ extracted from $\delta_{N,k}$. The above lemma along with condition $q_{\varphi(N)}(\delta_{\varphi(N),k}) \to 1$ implies that $\delta_{\varphi(N),k}$ converges to zero. Thus, every convergent subsequence extracted from $\delta_{N,k}$ converges to zero. Therefore, $\delta_{N,k}$ converges to zero. We now show that $N \delta_{N,k}$ is a bounded sequence. If $N \delta_{N,k}$ is not bounded, one can extract a subsequence $\varphi(N) \delta_{\varphi(N),k}$ such that $\varphi(N) \delta_{\varphi(N),k}$ tends to infinity as $N \to \infty$. Using again Lemma 8, we conclude that $q_{\varphi(N)}(\delta_{\varphi(N),k})$ tends to 0. This is in contradiction with $q_N(\delta_{N,k}) \to 1$. Thus, sequence $N \delta_{N,k}$ is bounded. Again, consider a sequence $\varphi(N) \delta_{\varphi(N),k}$ extracted from $N \delta_{N,k}$ such that $\varphi(N) \delta_{\varphi(N),k}$ tends to a certain $\beta \in \mathbb{R}$, then $q_{\varphi(N)}(\delta_{\varphi(N),k})$ tends to $e^{i \frac{\delta \sin(\beta/2)}{\beta/2}}$. Using again condition $q_{\varphi(N)}(\delta_{\varphi(N),k}) \to 1$, we conclude that $\beta = 0$. Since every convergent subsequence extracted from sequence $N \delta_{N,k}$ converges to zero, it follows that $N \delta_{N,k}$ converges to zero. We conclude that $\lim_{N \to \infty} N \delta_{N,k} = 0$ with probability one.

**APPENDIX II**

**PROOF OF LEMMA 3**

Consider a given user $k \in \{1, \ldots, K\}$. We derive the $k$th component of the gradient vector $\frac{\partial \mathcal{J}_N}{\partial \omega_k}$. Due to (20), cost function $\mathcal{J}_N$ can be written as $\mathcal{J}_N(\bar{\omega}) = z_N(\bar{\omega})^H z_N(\bar{\omega})$ where

$$z_N(\bar{\omega}) = \left( \frac{1}{N} Q_N(\bar{\omega}) R^{-1} Q_N(\bar{\omega})^H - I_{NN,H} \right) y_N. \quad (39)$$

Therefore, $\frac{\partial \mathcal{J}_N(\bar{\omega})}{\partial \omega_k} = 2 \text{Re} \left[ z_N(\bar{\omega})^H \frac{\partial z_N(\bar{\omega})}{\partial \omega_k} \right]$. Now, the derivative of vector $z_N(\bar{\omega})$ can be written as follows:

$$\frac{\partial z_N(\bar{\omega})}{\partial \omega_k} = -i I_{Nh} \otimes (T_{N,k}(\bar{\omega}_k) D_N - D_N T_{N,k}(\bar{\omega}_k)) y_N, \quad (40)$$
where $D_N = \text{diag}(0, 1, \ldots, N-1)$ and where $T_{N,k}(\tilde{\omega}_k) = \frac{1}{N} \Gamma_N(\tilde{\omega}_k) A_{N,k} R_k^{-1} A_{N,k}^H \Gamma_N(\tilde{\omega}_k)$. Using (40), the derivative of $J_N$ can be written as the sum $\frac{1}{N} \frac{\partial J_N(\omega)}{\partial \omega_k} = A_{N,k} + B_{N,k} + C_{N,k}$ where

$$A_{N,k} = \frac{2}{N^{3/2}} \sum_{r=1}^{N_u} \sum_{l=1}^{K} \text{Im} \left[ y_N^{(r)H} T_{N,l}(\tilde{\omega}_l) T_{N,k}(\tilde{\omega}_k) D_N y_N^{(r)} \right]$$

$$B_{N,k} = -\frac{2}{N^{3/2}} \sum_{r=1}^{N_u} \sum_{l=1}^{K} \text{Im} \left[ y_N^{(r)H} T_{N,l}(\tilde{\omega}_l) D_N T_{N,k}(\tilde{\omega}_k) y_N^{(r)} \right]$$

$$C_{N,k} = -\frac{4}{N^{3/2}} \sum_{r=1}^{N_u} \text{Im} \left[ y_N^{(r)H} T_{N,k}(\tilde{\omega}_k) D_N y_N^{(r)} \right].$$

We now study the asymptotic behaviors of the above three variables at point $\tilde{\omega} = \omega$. The first step of the proof consists in re-expressing the inner terms of the above equations as a function of more convenient quantities, whose asymptotic behavior can be easily characterized. Due to the definition of matrices $T_{N,k}$, it turns out that $A_{N,k}$, $B_{N,k}$ and $C_{N,k}$ mainly depend on the two quantities

$$R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) y_N^{(r)} \frac{1}{N} \quad \text{and} \quad 2R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N y_N^{(r)} \frac{1}{N^2}.$$  \hspace{1cm} (44)

It is thus convenient to express the above quantities in a more compact way. For each $u = 0, 1$,

$$(u + 1) R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N y_N^{(r)} \frac{1}{N^{u+1}} = (u + 1) R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N y_N^{(r)} \left( \sum_{l=1}^{K} T_N(\omega_l) A_{N,l} h_l^{(r)} + v_N^{(r)} \right) ,$$

$$= h_k^{(r)} + \frac{x_{N,k}^{(r),u}}{\sqrt{N}}.$$  \hspace{1cm} (45)

where

$$(u + 1) R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N y_N^{(r)} \frac{1}{N^{u+1}} = (u + 1) R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N y_N^{(r)} \left( \sum_{l=1}^{K} T_N(\omega_l) A_{N,l} h_l^{(r)} + v_N^{(r)} \right) ,$$

$$= h_k^{(r)} + \frac{x_{N,k}^{(r),u}}{\sqrt{N}}.$$  \hspace{1cm} (45)

where

$$x_{N,k}^{(r),u} = R_k^{-1} \sum_{l=1}^{K} P_{N,k,l}^{u} h_l^{(r)} + w_{N,k}^{(r),u}$$

$$P_{N,k,l}^{u} = \sqrt{N} \left( \frac{u + 1}{N^{u+1}} A_{N,k}^H \Gamma_N(\omega_l - \omega_k) D_N A_{N,l} - \delta(k - l) R_k \right)$$

$$w_{N,k}^{(r),u} = \frac{u + 1}{N^{u+1}} R_k^{-1} A_{N,k}^H \Gamma_N^H(\omega_k) D_N v_N^{(r)}.$$  \hspace{1cm} (47)

Decomposition (45) is particularly useful because it stresses the fact that the key quantities (44) simply coincide with $h_k^{(r)}$ up to a term which (as we shall prove in the sequel) tends to zero as $N$ tends to infinity. Using these notations, we may rewrite $A_{N,k}$, $B_{N,k}$ and $C_{N,k}$ as simple functions of $h_k^{(r)}$ and $x_{N,k}^{(r),u}$, $u = 0, 1$. As an example, it is straightforward to show that

$$C_{N,k} = -2\sqrt{N} \sum_{r=1}^{N_u} \text{Im} \left[ \left( h_k^{(r)} + \frac{x_{N,k}^{(r),0}}{\sqrt{N}} \right)^H R_k \left( h_k^{(r)} + \frac{x_{N,k}^{(r),1}}{\sqrt{N}} \right) \right].$$
Using (47) and summing the three terms $A_{N,k} + B_{N,k} + C_{N,k}$, we obtain after some algebra
\[
\frac{1}{N\sqrt{N}} \frac{\partial J_N(\omega)}{\partial \omega_k} = -\sqrt{N} \sum_{r=1}^{N_N} \sum_{t \neq k} \Im \left[ \left( h_t^{(r)} + \frac{x_{N,l}^{(r),0}}{\sqrt{N}} \right)^H P_{N,l,k}^1 \left( h_k^{(r)} + \frac{x_{N,k}^{(r),0}}{\sqrt{N}} \right) \right] \\
+ \sqrt{N} \sum_{r=1}^{N_N} \sum_{l=1}^{K} \Im \left[ \left( h_t^{(r)} + \frac{x_{N,l}^{(r),0}}{\sqrt{N}} \right)^H \left( \delta(k-l) R_k + \frac{P_{0,l,k}^{(r)}}{\sqrt{N}} \right) \left( h_k^{(r)} + \frac{x_{N,k}^{(r),1}}{\sqrt{N}} \right) \right].
\] (49)

The rest of the proof simply consists in expanding the above expression and in only keeping the dominant terms, i.e., the terms which do not contain any factor $1/\sqrt{N}$ or $1/N$. In order to motivate this, we make use of the following lemma.

**Lemma 9:** For each $u, v = 0, 1$, for each $k, l, l' = 1, \ldots, K$, the three following random matrices
\[
\frac{P_{N,l,k}^u}{\sqrt{N}}, \quad \frac{P_{u,l,k}^{r} R_k^{-1} P_{N,k,l'}^u}{\sqrt{N}}, \quad \frac{P_{u,N,l,k}^{r} w_{N,k}^{(r),u}}{\sqrt{N}}
\]
converge to zero as $N$ tends to infinity with probability one.

Before providing the proof of the above lemma, we first show how Lemma 9 allows to complete the proof of the final result. Using Lemma 9, it becomes clear that terms such as (typically) $P_{N,l,k}^0 x_{N,k}^{(r),1} / \sqrt{N}$ or $x_{N,k}^{(r),H} x_{N,k}^{(r),1} / \sqrt{N}$ tend to zero as $N$ tends to infinity. Consequently, if one expand all factors of the righthand side of equation (49), only a few of them do not converge to zero. Remarking also that
\[
\sqrt{N} \Im \left[ h_k^{(r)H} R_k h_k^{(r)} \right] = 0,
\]
we obtain:
\[
\frac{1}{N\sqrt{N}} \frac{\partial J_N(\omega)}{\partial \omega_k} = \sum_{r=1}^{N_N} \Im \left[ h_k^{(r)H} R_k \left( x_{N,k}^{(r),0} - x_{N,k}^{(r),1} \right) - \sum_{l \neq k} h_k^{(r)H} \left( P_{N,k,l}^0 - P_{N,l,k}^1 \right) h_l^{(r)} \right] + o_{\text{as}}(1)
\]
\[
= \sum_{r=1}^{N_N} \Im \left[ h_k^{(r)H} R_k \left( w_{N,k}^{(r),0} - w_{N,k}^{(r),1} \right) \right] + o_{\text{as}}(1),
\]
where we used the fact that for $k = l$, the imaginary part of $h_k^{(r)H} \left( P_{N,k,k}^0 - P_{N,k,k}^1 \right) h_k^{(r)}$ is equal to zero. Replacing $w_{N,k}^{(r),0}$ and $w_{N,k}^{(r),1}$ with the corresponding expressions, we obtain the final result. We now prove Lemma 9.

**Proof:** The almost sure convergence to zero of the first random matrix $P_{N,l,k}^u / \sqrt{N}$ has been proved in [11]. We refer to [11] for a proof of this result. Here, we only address the most difficult part, which is the almost sure convergence to zero of the second random matrix $P_{u,l,k}^{r} R_k^{-1} P_{N,k,l'}^u / \sqrt{N}$. We focus, without restriction, on the case $u = v = 0$ in order to keep the notations simple. The general case can be treated similarly, without more difficulties. Moreover, we focus on the single transmit antenna case $N_T = 1$ in order to avoid the heavy indices $t, t'$ representing the antenna numbers. The generalization to
Matrix $P_{N,l,k}$ is equal to zero for $k \neq l$. For $k = l$, the component $(p, q)$ of matrix $P_{N,k,k}$ can be written as

$$P_{N,k,k}(p, q) = \sqrt{N} \left( \frac{1}{N} \sum_{n=0}^{N-1} E[a_{N,k}(n-p)*a_{N,k}(n-q)] - R_k(p, q) \right)$$

$$= \sqrt{N} \left( \frac{1}{N} \sum_{j=0}^{N-1} E[|s_{N,k}(j)|^2 e^{2\pi i (p-q)}] - R_k(p, q) \right)$$

$$= \sqrt{N} \left( \int_0^1 e^{2\pi i (p-q)f} \mu_{N,k}(df) - \int_0^1 e^{2\pi i (p-q)f} \mu_k(df) \right).$$

According to Assumption 5, the above term is bounded with a constant which does not depend on $N$. Using the almost sure convergence to zero of the first random matrix $P_{N,l,k}/\sqrt{N}$, $P_{N,l,k} R_k^{-1} P_{N,k,l}/\sqrt{N}$ becomes the product between a deterministic matrix which bounded componentwise and a matrix which tends almost surely to zero. The product thus converges almost surely to zero. With such observations, it turns out that the proof of the almost sure convergence to zero of $P_{N,l,k} R_k^{-1} P_{N,k,l}/\sqrt{N}$ simply reduces to the proof that $P_{N,l,k} R_k^{-1} P_{N,k,l}/\sqrt{N} \to 0$ a.s. To that end, it is sufficient to prove that for each $(p, q), (p', q')$, $\xi_N = \frac{1}{N} P_{N,l,k}(p, q) P_{N,k,l}(p', q')$ converges almost surely to zero. After some algebra, we obtain

$$P_{N,l,k}(p, q) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sigma_{N,l,k}(i, j) \psi_N^0(j - i + N\delta f_{lk})$$

where $\delta f_{lk} = \frac{\omega_{l,k}}{2\pi}$. $\psi_N^0(x) = \frac{1}{N} \sum_{n=0}^{N-1} e^{inx/N}$ and where $\sigma_{N,l,k}(i, j) = s_{N,k}(i)^* s_{N,l}(j) - E[s_{N,k}(i)^* s_{N,l}(j)]$.

We now derive the fourth-order moment of $\xi_N$. By simply expanding the product $|\xi_N|^4$, taking the expectation and finally using the triangular inequality, we obtain

$$E[|\xi_N|^4] \leq \frac{1}{N^6} \sum_{i_1, \ldots, i_5} \sum_{j_1, \ldots, j_5} \left| E \left[ \prod_{n=1}^{4} \sigma_{N,l,k}(i_n, j_n)^* \prod_{n=5}^{8} \sigma_{N,k,l'}(i_n, j_n)^* \right] \right|$$

$$\times \left| \prod_{n=1}^{4} \psi_N^0(j_n - i_n + N\delta f_{lk}) \prod_{n=5}^{8} \psi_N^0(j_n - i_n + N\delta f_{lk}) \right|,$$
where \( x^{+n} = x \) if \( n \) is even and \( x^{-n} = x^* \) if \( n \) is odd. Due to the fact that, \( s_{N,k}(i) \) is a sequence of independent random variables, it can be seen that the expectation inside the sum in the above equation is zero for most 16-uplets \( i_1, i_2, \ldots, j_8 \). Only a limited number of terms are non zero. For instance, one of these terms is the one obtained for \( i_1 = i_2, j_1 = j_2, i_3 = i_4, j_3 = j_4, i_5 = i_6, j_5 = j_6, i_7 = i_8, j_7 = j_8 \). We now study the latter term in more details. This term can be written as

\[
\sum_{i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4} \alpha_N(i_1, \ldots, j_4) \prod_{n=0}^3 |\psi_N^j(j_2n+1 - i_2n+1 + N\delta f_{lk})|^2
\]

Finally, we use the fact that for each \( \delta f_{lk} \) which does not depend on \( N \), we finally obtain that

\[
\lim_{N \to \infty} \frac{1}{N^6} \sum_{i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4} \alpha_N(i_1, \ldots, j_4) \prod_{n=0}^3 |\psi_N^j(j_2n+1 - i_2n+1 + N\delta f_{lk})|^2 = C'
\]

where \( \alpha_N(i_1, i_3, \ldots, j_7) = E \left[ |\sigma_{N,l,k}(i_1, j_1)|^2 |\sigma_{N,l,k}(i_3, j_3)|^2 |\sigma_{N,k,l}(i_5, j_5)|^2 |\sigma_{N,k,l}(j_7, j_7)|^2 \right] \) is bounded by \( \left( E \left[ |\sigma_{N,l,k}(i_1, j_1)|^6 \right] \right) \left( E \left[ |\sigma_{N,k,l}(i_5, j_5)|^6 \right] \right) \left( E \left[ |\sigma_{N,k,l}(j_7, j_7)|^6 \right] \right) \) by the Cauchy-Schwarz inequality. By expanding the fourth order moment of \( \sigma_{N,l,k} \) and by using Cauchy-Schwarz and Jensen inequalities, it is straightforward to show that \( \alpha_N(i_1, \ldots, j_7) \) is bounded with a constant, say \( C' \), which does not depend on \( N \). This implies that the term (51) is less than

\[
\frac{C'}{N^2} \sum_{m=-N+1}^{N-1} \left| \psi_N^0(m + N\delta f_{lk}) \right|^2 \left( \sum_{m=-N+1}^{N-1} \left| \psi_N^0(m + N\delta f_{lk}) \right|^2 \right)^2
\]

Finally, we use the fact that for each \( k, l \), the sum \( \sum_{m=-N+1}^{N-1} \left| \psi_N^0(m + N\delta f_{lk}) \right|^2 \) is bounded with a constant which does not depend on \( N \) [11]. Finally, (51) is less than \( \frac{C''}{N^2} \), where \( C'' \) is a certain constant. Using the same kind of study for all other non zero terms of the sum (50), we conclude that \( E[|\xi_N|^4] \leq \frac{C}{N^2} \), where \( C \) is a constant which does not depend on \( N \). Using Chebyshev inequality followed by the Borel-Cantelli lemma, we finally obtain that \( \xi_N \) converges almost surely to zero. This completes the proof.

**APPENDIX III**

**PROOF OF LEMMA 4**

We first consider the case \( k = l \) and we derive \( \frac{\partial^2 J_N}{\partial \omega_k^2} \) at point \( \tilde{\omega} = \tilde{\omega}_N \). After some algebra, it can be shown that \( \frac{1}{N^2} \frac{\partial^2 J_N}{\partial \omega_k^2} (\tilde{\omega}_N) \) can be written as the sum of a certain number of terms. For instance, the first of these terms coincides with

\[
\hat{A}_{N,k} = -2 \frac{2}{N^3} \sum_{r=1}^N \sum_{l=1}^K \Re \left[ y_N^{(r)} H_T N (l) N (l) N (k) N (k) D_N y_N^{(r)} \right]
\]

Other terms can be written under a similar form. Their expressions are not provided here due to the lack of space, but a more detailed proof is available in [15]. Clearly, the asymptotic study of \( \frac{1}{N^2} \frac{\partial^2 J_N}{\partial \omega_k^2} (\tilde{\omega}_N) \)
reduces to the study of terms such as $\tilde{A}_{N,k}$. Here, we only focus on the study of $\tilde{A}_{N,k}$ for the sake of conciseness. Replacing the above matrices $T_{N,k}(\tilde{\omega}_{N,k})$ with definition of matrices $T_{N,k}$, we obtain

$$\tilde{A}_{N,k} = -\frac{2}{3} \sum_{r=1}^{N_u} \sum_{l=1}^{K} \Re \left[ \frac{y_n^{(r)}}{N} \Gamma_N(\tilde{\omega}_{N,l}) A_{N,l} R_l^{-1} A_{N,k}^T \Gamma_N(\tilde{\omega}_{N,k} - \tilde{\omega}_{N,l}) A_{N,k} R_k^{-1} 3A_{N,k}^T \Gamma_N(\tilde{\omega}_{N,k}) D_N^2 y_n^{(r)} R_k^{-1} N^3 \right]$$

(52)

The limit of $\tilde{A}_{N,k}$ can be directly obtained by making use of the following lemma.

**Lemma 10:** For each $u = 0, 1$, for each $k, l$,

$$\lim_{N \to \infty} \left( u + 1 \right) \frac{A_{N,k}^H \Gamma_N(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k}) D_N^2 A_{N,l}}{N^{u+1}} = \delta(k - l) R_k \text{ a.s.}$$

(53)

$$\lim_{N \to \infty} \left( u + 1 \right) \frac{A_{N,k}^H \Gamma_N(\tilde{\omega}_{N,k})^H D_N^2 y_n^{(r)}}{N^{u+1}} = R_k h_k^{(r)} \text{ a.s.}$$

(54)

By straightforward application of the above lemma, $\tilde{A}_{N,k}$ converges to $-\frac{2}{3} \sum_{r=1}^{N_u} h_k^{(r)} R_k h_k^{(r)} = -\frac{2}{3} \gamma_k$ as $N$ tends to infinity. Using similar arguments for all components such as $\tilde{A}_{N,k}$, we obtain that $\frac{\partial^2 J_N}{\partial x_k^2}$ converges to $\frac{2}{N}$ as $N$ tends to infinity. The rest of the proof consists in deriving the non diagonal terms of the Hessian matrix at point $\tilde{\omega}_N$ and in proving that for each $k \neq l$, $\frac{1}{N} \frac{\partial^2 J_N(\omega)}{\partial x_k \partial x_l}(\tilde{\omega}_N)$ converges a.s. to zero as $N$ tends to infinity. The proof follows exactly the same approach. This completes the proof of Lemma 4. We now provide a sketch of the proof of Lemma 10.

**Proof:** Again, in order to simplify the proof and to avoid the overabundance of indices, we provide the proof in the single antenna case $N_T = 1$ without loss of generality. We focus on the proof of (53). A more detailed proof is available in [15]. The element $(p, q)$ of the matrix in the righthand side of (53) coincides with

$$\frac{u + 1}{N^{u+1}} \sum_{n=0}^{N-1} a_{N,k}(n-p) a_{N,l}(n-q) n^u e^{iN(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k})}$$

$$= \frac{u + 1}{N^{u+1}} \sum_{n=0}^{N-1} E \left[ a_{N,k}(n-p) a_{N,l}(n-q) e^{iN(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k})} + S_{N,k,l}(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k}) \right]$$

$$= \frac{\delta(k - l) q_k^N(0)}{N} \sum_{j=0}^{N-1} E \left[ |s_{N,k,j}|^2 e^{2\pi j (q-p)} \right] + S_{N,k,l}(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k}),$$

where $S_{N,k,l}^{p,q,u}$ is defined by (37) and where $q_k^N(x) = \frac{u+1}{N^{u+1}} \sum_{n=0}^{N-1} n^u e^{inx}$. Due to Lemma 6, $S_{N,k,l}(\tilde{\omega}_{N,l} - \tilde{\omega}_{N,k})$ converges a.s. to zero as $N$ tends to infinity. On the other hand, the first term of the above equation converges to $\delta(k - l) R_k(p, q)$. This completes the proof. □
APPENDIX IV

DERIVATION OF THE LIMIT COVARIANCE MATRIX OF $\Im [Z_N^{}v_N^{}]$

We first remark that

$$E_N^{} \left[ \Im [Z_N^{}v_N^{}] \Im [Z_N^{}v_N^{}]^T \right] = \frac{\sigma_2}{2} \Re \left[ Z_N^{}Z_N^H \right].$$

Matrix $Z_N^{}Z_N^H$ is composed of $K \times K$ blocks, i.e., $Z_N^{}Z_N^H = \left[ Z_{N,k}^{}Z_{N,l}^H \right]_{k,l=1,\ldots,K}$. We study one of these blocks. Due to the definition (36) of $Z_{N,k}$, the product $Z_{N,k}^{}Z_{N,l}^H$ can be written as a simple function of the following four matrices: $\Phi_{N,k}^{}\Psi_{N,l}^H$, $\Psi_{N,k}^{}\Phi_{N,l}^H$, $\Phi_{N,k}^{}\Psi_{N,l}^H$, and $\Psi_{N,k}^{}\Phi_{N,l}^H$. It is thus sufficient to study the limit of each of these matrices as $N$ tends to infinity. Clearly,

$$\Psi_{N,k}^{}\Psi_{N,l}^H = I_{N,k} \otimes R_{k}^{-1} A_{N,k}^H\Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H R_{k}^{-1}$$

converges almost surely to $I_{N,k} \otimes R_{k}^{-1} (\delta(k-l)R_{k}^{} ) R_{k}^{-1} = \delta(k-l)I_{N,k} \otimes R_{k}^{-1}$. Similarly,

$$\Phi_{N,k}^{}\Psi_{N,l}^H = I_{N,k} \otimes A_{N,k}^H (I_{N}^{} - \frac{2}{N}D_{N}^{} ) \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H R_{k}^{-1}$$

$$= I_{N,k} \otimes A_{N,k}^H R_{k}^{-1} \left( \frac{2A_{N,k}^H D_{N}^{} \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H}{N} \right) R_{k}^{-1}$$

and thus converges almost surely to zero as both terms enclosed in the parenthesis of the above equation converge almost surely to $\delta(k-l)R_{k}^{}$. Finally,

$$\Phi_{N,k}^{}\Phi_{N,l}^H = I_{N,k} \otimes A_{N,k}^H (I_{N}^{} - \frac{2}{N}D_{N}^{} )^2 \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H$$

$$= I_{N,k} \otimes A_{N,k}^H R_{k}^{-1} \left( \frac{A_{N,k}^H D_{N}^{} \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H}{N} \right)$$

$$- \frac{4A_{N,k}^H D_{N}^{} \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H}{N^2} + \frac{4A_{N,k}^H D_{N}^4 \Gamma_{N}^H(\omega_l^{}-\omega_k^{} ) A_{N,l}^H}{N^3}$$

converges a.s. to $I_{N,k} \otimes (1-2+\frac{4}{3})\delta(k-l)R_{k} = \frac{1}{3}\delta(k-l)I_{N,k} \otimes R_{k}$. These results imply that $Z_{N,k}^{}Z_{N,l}^H$ converges to zero if $k \neq l$. For $k = l$, the use of the above results along with (36), it is straightforward to show that $\frac{\sigma_2}{2} \Re \left[ Z_N^{}Z_N^H \right]$ converges almost surely to $\Sigma_k$. Finally, $\frac{\sigma_2}{2} \Re \left[ Z_N^{}Z_N^H \right]$ converges almost surely to $\Sigma_k$.

REFERENCES


Fig. 1. Convergence behavior of the proposed estimator: \( N_R = N_T = 1, \frac{P_1}{\sigma^2} = 10 \text{dB} \).

Fig. 2. Convergence behavior of the proposed estimator: \( N_R = N_T = 2, \frac{P_1}{\sigma^2} = 10 \text{dB} \).
Fig. 3. Performance of the proposed estimator for frequency offset estimation with $N = 1024$.

Fig. 4. Comparison of the proposed estimator with APFE : $N_R = N_T = 1$, $K = 4$, $N = 256$. 