Asymptotic Cramér-Rao Bounds and Training Design for MIMO-OFDMA Uplink Transmissions with Frequency Offsets

Serdar Sezginer*, Student Member, IEEE, Pascal Bianchi, Member, IEEE,
Walid Hachem, Member, IEEE.

Abstract

In this paper, we address the data-aided joint estimation of frequency offsets and channel coefficients in the uplink transmission of MIMO-OFDMA systems. A compact and informative expression of the Cramér-Rao Bound (CRB) is derived for large training sequence sizes. It is shown that the asymptotic performance bounds do not depend on the number of users and the values of the frequency offsets. Moreover, we observe that different carrier assignment schemes (CAS) may lead to the same asymptotic CRB. Next, we bring to the fore the training strategies which minimize the asymptotic performance bounds and which are therefore likely to lead to accurate estimates of the parameters. In particular, for a given user, it is shown that accurate frequency offset estimates are likely to be obtained by introducing relevant correlation between training sequences sent at different antennas. On the other hand, accurate channel estimation is achieved when training sequences sent at different antennas are uncorrelated. Simulation results sustain our claims.

Index Terms

Cramér-Rao bound, MIMO, OFDMA, training sequence.

EDICS category: SSP-PERF; SPC-PERF; SPC-MULT

S. Sezginer, P. Bianchi, and W. Hachem are with the Telecommunications Department, Supélec, Plateau de Moulon, F-91192 Gif-sur-Yvette France (e-mail: {serdar.sezginer, pascal.bianchi, walid.hachem}@supelec.fr).
The material in this paper will be presented in part at ICASSP 2006.
* Corresponding Author
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 broadband wireless access at frequencies below 11 GHz. In an OFDMA system, each user modulates a certain group of subcarriers, following a given Carrier Assignment Scheme (CAS). The signal transmitted by a given user is impaired by frequency selective channels and by a frequency offset. Prior estimation of channel coefficients and frequency offsets has a considerable impact on further users detection steps. Indeed, along with its attractive features, the drawbacks associated with Orthogonal Frequency Division Multiplexing (OFDM) are directly inherited by OFDMA. The computation of accurate estimates is thus a crucial issue in OFDMA systems. Estimation of frequency offset and channel coefficients for (single-user) OFDM systems has been investigated in a large number of works (see [1]–[5] and references therein). However, the case of OFDMA uplink usually requires more involved synchronization and channel estimation methods. In [6]–[8], estimators which are designed for a specific CAS are proposed. In case of group orthogonal CAS, 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. A cyclic prefix based method is proposed in [7]. [4] proposes an estimator based on Kurtosis maximization. In case of the so-called interleaved CAS, i.e., when adjacent subcarriers (or adjacent small groups of subcarriers) are modulated by different users, [8] proposes a subspace method based on the periodic structure of the signals transmitted by each user. Recent works [9] and [10] investigate the data-aided frequency offset estimation of frequency offsets in OFDMA uplink for general carrier assignment schemes. For instance, [10] proposes an alternating-projection method to recover sequentially the carrier frequency offsets of all users. Next, the latter estimates of the frequency offsets are used to obtain both the channel parameters and the timing offsets.

Here, we consider an uplink MIMO-OFDMA transmission involving $N_T$ transmit antennas per user and $N_R$ receive antennas at the base station. If $K$ denotes the number of users, the receiver must estimate all $K$ frequency offsets (one for each user) and all $K$ MIMO-channels. In this paper, we investigate the performance of the data-aided estimation of channels and frequency offsets. To that end, we study the Cramér-Rao Bound (CRB) for the joint data-aided estimation of the set of frequency offsets and channel coefficients. Such an analysis provides lower bounds on the Mean Square Error (MSE) associated with estimates of the unknown parameters. Moreover, it emphasizes the parameters which have crucial impact on the performance. In particular, the CRB depends on the training sequences sent by all users. Next,
following the general idea of [11] and [12], our aim is to characterize the training sequences which lead to the lowest CRB. It is indeed reasonable to believe that the use of such “optimal” training sequences is likely to provide accurate estimates of the unknown parameters.

Section II describes the signal model. In Section III, we derive an exact expression of the CRB. Unfortunately, the exact CRB turns out to be complicated. In order to obtain a compact and informative expression of the CRB, we assume that the total number $N$ of modulated subcarriers tends to infinity. In this case, we obtain a simple expression of the CRB. Such an asymptotic analysis of the CRB has been proposed by [12] in case of a single-user single-carrier single-antenna system. Asymptotic analysis of the CRB associated with the frequency offset in case of single carrier MIMO systems has been investigated by [13] considering flat fading channels. Recently, the asymptotic CRB for clock offset and channel estimation has been studied by [14] in case of OFDM. In the present paper, it is shown that the asymptotic CRB associated with parameters of the $k$th user is identical to the asymptotic CRB that one would have obtained in the absence of other users $l \neq k$. We also observe that the asymptotic CRB depends on the choice of the training sequence only via the frequency power profile of the training sequence and via the correlation possibly introduced between transmit antennas (due to the possible use of a beamformer at each transmitter side). It is shown that different CAS can lead to an identical asymptotic CRB. In Section IV, we characterize the training strategies which minimize the asymptotic CRB. In particular, we investigate which power should be allocated to which subcarriers, and what correlation should be introduced between training sequences sent at different antennas so as to lead to accurate estimates of the parameters. Unfortunately, as already noticed by [11] [12] [15], no single training strategy is likely to simultaneously provide the most accurate estimates of the frequency offset and, at the same time, the most accurate estimates of channel coefficients. In practice, one should determine tradeoffs between training strategies providing accurate frequency offset estimates and training strategies providing accurate channel estimates. However, the problem of choosing a relevant tradeoff is difficult and crucially depends on the transmitter and receiver architectures. The proposition of an general procedure for determining such tradeoffs is beyond the scope of this paper. Finally, simulation results of Section V sustain our claims.

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 notational simplicity. In the sequel, we assume that for each $k = 1, \ldots, K$ and for each $t = 1, \ldots, N_T$, sequence $(s_{N,k}^{(t)}(j))_j$ is known by the receiver (training sequence). It is worth noting that in usual OFDMA transmission schemes, only a subset of the $N$ available subcarriers is effectively modulated by a given user $k$, following a given CAS. 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 subcarrier assignment strategy at this point. 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 $(s_{N,k}^{(t)}(j))_j$:

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

(1)

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

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

(2)

For each $k = 1, \ldots, K$, 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 symbol 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 $\sigma^2$ and spectral density $N_0$. Note that equation (2) implicitly assumes that all users are quasi-synchronous in time as in [10]: 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 [16] and is investigated in a recent work [13]. In the present paper, we consider the classical case (see, e.g., [4] and references therein) where $\omega_k$ is constant w.r.t. antenna pairs $(t, r)$. In particular, this assumption is valid in case of an identical angle of arrival at each receive antenna.
In the sequel, it is quite useful to make use of a compact matrix representation of (2). To that end, we introduce the following notations. Define

\[ y_N^{(r)} = [h_k^{(r)}(0), \ldots, h_k^{(r)}(L-1)]^T, \]

where \((\cdot)^T\) represents the transpose operator. Stacking all \(N\) samples \(y_N^{(r)}(n)\) received by antenna \(r\) into one column vector \(y_N^{(r)} = [y_N^{(r)}(0), \ldots, y_N^{(r)}(N-1)]^T\), one obtains:

\[
y_N^{(r)} = \sum_{k=1}^{K} \Gamma_N(\omega_k) \left( \sum_{t=1}^{N_T} A_{N,k}^{(t)} h_k^{(r,t)} \right) + v_N^{(r)},
\]

where \(v_N^{(r)} = [v_N^{(r)}(0), \ldots, v_N^{(r)}(N-1)]^T\) and \(\Gamma_N(\omega_k) = \text{diag}(1, e^{i\omega_k}, \ldots, e^{i\omega_k(N-1)})\). In the above expression, each matrix \(A_{N,k}^{(t)}\) is an \(N \times L\) matrix containing the time-domain training sequence sent at the \(t\)th transmit antenna of user \(k\). More precisely,

\[
A_{N,k}^{(t)} = \left( a_{N,k}^{(t,i,j)}(i-j) \right)_{0 \leq i \leq N-1, \ 0 \leq j \leq L-1}.
\]

In particular, \(A_{N,k}^{(t)}\) is a circulant matrix: this is due to the fact that cyclic prefix is inserted at the transmitter side. We finally stack the samples received by all antennas into a single \(NN_T\times1\) vector

\[
y_N = [y_N^{(1)}^T, \ldots, y_N^{(N_T)}]^T\]

given by

\[
y_N = \sum_{k=1}^{K} \left[ I_{N_R} \otimes (\Gamma_N(\omega_k)A_{N,k}) \right] h_k + v_N,
\]

where \(I_{N_R}\) denotes the \(N_R \times N_R\) identity matrix, \(\otimes\) stands for the Kronecker product, \(A_{N,k} = [A_{N,k}^{(1)}, \ldots, A_{N,k}^{(N_T)}]\), and \(v_N = [v_N^{(1)}^T, \ldots, v_N^{(N_T)}]^T\) is an additive noise vector with independent complex circular Gaussian random entries of variance \(\sigma^2\). Here, vector \(h_k = [h_k^{(1,1)}^T, \ldots, h_k^{(1,N_T)}^T, \ldots, h_k^{(N_T,1)}^T, \ldots, h_k^{(N_T,N_T)}^T]^T\) contains all channel coefficients of a given user \(k\).

In this paper, we address the data-aided estimation of the \(K\) unknown frequency offsets \(\delta f_k\) (or equivalently the \(K\) normalized angular frequency offsets \(\omega_k\)) and the \(KLN_RN_T\) unknown channel coefficients. We denote by \(\theta = [\omega_1, h_1^T, \ldots, \omega_K, h_K^T]^T\) the deterministic parameter vector to be estimated.

In the following section, we provide bounds on the performance of estimates of \(\theta\).

Remark 1: In a number of practical OFDMA systems, the receiver needs to estimate the channel’s response associated with a given user \(k\) only at the subcarriers which are effectively modulated by user \(k\). For example, if a given user is constrained to transmit in a frequency band which is strictly smaller than the total bandwidth of the system [17], then the accuracy of channel estimation outside this frequency band is of few interest. One may thus be interested in estimating a portion of the frequency response of the channel rather than the coefficients of the impulse response, i.e., vector \(h_k\). However, the set of frequency taps which should be estimated in this case can be written as a simple function of \(h_k\). For this
reason, we first investigate the performance of estimates of initial parameter vector $\theta$ and deduce from this the performance of estimates of the frequency taps of interest.

III. CRAMÉR-RAO BOUND

We now study the CRB associated with $\theta$. Such an analysis provides performance bounds for estimates of $\theta$. Moreover, it emphasizes the influence of the choice of the training sequence on the performance.

In this section, we firstly derive the exact CRB for parameter $\theta$. Secondly, we investigate the asymptotic behavior of the CRB as the number of subcarriers tends to infinity. Using these results, we finally derive the asymptotic CRB associated with the channels’ frequency responses.

A. Exact Cramér-Rao bound

Real parameter vector can be written as $\tilde{\theta} = [\tilde{\theta}_1^T, \ldots, \tilde{\theta}_K^T]^T$ where for each $k$, $\tilde{\theta}_k = [\omega_k, h_{R,k}^T, h_{I,k}^T]^T$ denotes the parameter vector corresponding to a given user $k$. In the above definition, $h_{R,k}$ and $h_{I,k}$ respectively represent the real and the imaginary parts of vector $h_k$. The exact CRB for $\tilde{\theta}$ is classically defined as the inverse of the Fisher Information Matrix (FIM) $J_N$. Note that an expression of the FIM for $\tilde{\theta}$ has been recently derived by [10] in the SISO case. Using an approach similar to [12], the FIM for parameter $\tilde{\theta}$ can be obtained as the following $K(1 + 2LN_R N_T) \times K(1 + 2LN_R N_T)$ matrix:

$$J_N = \frac{2}{\sigma^2} \Re \left[ (\nabla_{\theta}^H \eta_N) (\nabla_{\theta}^H \eta_N)^H \right],$$

(4)

where $\eta_N = \sum_{k=1}^{K} [I_{N_R} \otimes (\Gamma_N(\omega_k) A_{N,k})] h_k$. Here, superscript $(\cdot)^H$ denotes the transpose-conjugate, $\Re [x]$ (resp. $\Im [x]$) denotes the real (resp. imaginary) part of $x$. For any $n \times 1$ real column vector $x = [x_1, \ldots, x_n]^T$ and for a given $1 \times m$ row vector $z(x) = [z_1(x), \ldots, z_m(x)]$ function of $x$, matrix $\nabla_x z$ is defined as the $n \times m$ matrix $\left( \frac{\partial z_j(x)}{\partial x_i} \right)_{1 \leq i \leq n, 1 \leq j \leq m}$. After some algebra, we obtain the following “block” representation of matrix $J_N$:

$$J_N = \begin{bmatrix} J_{N,1,1} & \cdots & J_{N,1,K} \\ \vdots & \ddots & \vdots \\ J_{N,K,1} & \cdots & J_{N,K,K} \end{bmatrix}.$$  

(5)

For each $k, l = 1, \ldots, K$, $J_{N,k,l}$ is the $(1 + 2LN_R N_T) \times (1 + 2LN_R N_T)$ matrix given by

$$J_{N,k,l} = \frac{2}{\sigma^2} \begin{bmatrix} \alpha_{N,k,l} & \Re [\beta_{N,l,k}^H] & \Re [\beta_{N,l,k}^H] \\ -\Im [\beta_{N,k,l}] & \Re [U_{N,k,l}] - \Im [U_{N,k,l}] \\ \Re [\beta_{N,k,l}] & \Im [U_{N,k,l}] & \Re [U_{N,k,l}] \end{bmatrix}.$$  

(6)
where

\[
U_{N,k,l} = I_N \otimes (A_{N,k}^H \Gamma_N (\omega_l - \omega_k) A_{N,l}),
\]

(7)

\[
\beta_{N,k,l} = [I_N \otimes (A_{N,k}^H D_N \Gamma_N (\omega_l - \omega_k) A_{N,l})] h_l,
\]

(8)

\[
\alpha_{N,k,l} = \Re [I_N \otimes (A_{N,k}^H D_N^2 \Gamma_N (\omega_l - \omega_k) A_{N,l})] h_l,
\]

(9)

and where \(D_N = \text{diag}(0, 1, \ldots, N - 1)\). Expression (6) can be obtained by straightforward derivation of (4). The proof is omitted due to the lack of space. Unfortunately, the calculation of the exact CRB as the inverse of (5) seems to be a very difficult task. In order to obtain a more compact and informative expression of the CRB, we now investigate the case where the number \(N\) of subcarriers increases.

B. Asymptotic Cramér-Rao Bound

We now study the asymptotic behavior of the CRB for \(\tilde{\theta}\). We assume that \(N\) tends to infinity while i) the number \(K\) of users remains constant and while ii) the number of antennas remains constant. In practice, our results will be valid as long as the number \(N\) of subcarriers is significantly greater than both the number \(K\) of user and the total number of antennas. We also assume that when \(N\) tends to infinity, the overall bandwidth is constant. In other words, sampling rate \(\frac{1}{T}\) remains constant and as a result, the subcarrier spacing \(\frac{1}{NT}\) decreases to zero.

In order to simplify the following analysis, we assume that for a given antenna \(t\) of a given user \(k\), \((s_{N,k}^{(t)}(j))_j\) is a sequence of independent random variables with zero mean. Note that this assumption encompasses usual OFDMA training strategies. However, we do not assume that training symbols are identically distributed. In particular, the variance \(E[|s_{N,k}^{(t)}(j)|^2]\) of the \(j\)th training symbol depends on \(j\). This is motivated by the observation that in practical OFDM systems, different powers may be allocated to different subcarriers. Moreover, for a given CAS, a certain number of subcarriers may not be modulated by user \(k\). If \(j\) is one of these subcarriers, we simply consider that \(E[|s_{N,k}^{(t)}(j)|^2] = 0\). Furthermore, we assume that training sequences \((s_{N,k}^{(t)}(j))_j\) and \((s_{N,k}^{(t')}_{(j)}(j))_j\) 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 \(E[s_{N,k}^{(t)}(j)s_{N,k}^{(t')}_{(j)}(j)^*]\) may be nonzero. We assume that eight-order moments of random variables \((s_{N,k}^{(t)}(j))_j\) are uniformly bounded, i.e.

\[
\sup_N \max_j E \left[ |s_{N,k}^{(t)}(j)|^8 \right] < M
\]

(10)

for each \(t\), where \(M\) is a constant which is independent of \(N\). Finally, we assume that training sequences sent by two different users \(l \neq k\) are independent.
We now study the asymptotic behavior of the CRB matrix defined as \( \text{CRB}_N = J_N^{-1} \). Note that for finite values of \( L \), it is reasonable to assume that the FIM is non-singular. Due to definitions of \( U_{N,k,l} \), \( \beta_{N,k,l} \) and \( \alpha_{N,k,l} \) in (7), (8) and (9), it is clear that the behavior of the CRB for large \( N \) only depends on the asymptotic behavior of matrices \( A^H_{N,k} D_N^\alpha \Gamma_N (\omega_l - \omega_k) A_{N,l} \) for each \( u = 0, 1, 2 \) and for each \( k, l = 1, \ldots, K \). The following lemma provides a simpler expression of the latter matrices as \( N \) increases.

**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 \( k, l = 1, \ldots, K \) and for each \( u = 0, 1, 2 \),

\[
\frac{u+1}{N} A^H_{N,k} D_N^\alpha \Gamma_N (\omega_l - \omega_k) A_{N,l} - \frac{\delta(k-l)}{N} \sum_{j=0}^{N-1} E \left[ s_{N,k}(j) s_{N,k}(j)^H \right]^* \otimes \left[ e\left( \frac{j}{N} \right) \right] e\left( \frac{j}{N} \right)^H \xrightarrow{N \to 0} 0 \quad \text{a.s.}
\]

(11)

where notation \( \xrightarrow{N \to 0} 0 \) stands for the (componentwise) convergence to zero as \( N \) tends to infinity, where “a.s.” stands for “almost surely” and where \( x^* \) stands for the conjugate of \( x \). Here, vector \( s_{N,k}(j) = \left[ s_{N,k}^{(1)}(j), \ldots, s_{N,k}^{(N)}(j) \right]^T \) contains training symbols sent by all antennas of user \( k \) at a given subcarrier \( j \). Coefficient \( \delta(k-l) \) is equal to 1 if \( k = l \) and to zero otherwise.

The proof of Lemma 1 is given in Appendix I. The presence of factor \( \delta(k-l) \) in (11) already gives the insight that elements of non diagonal blocks \( J_{N,k,l} \) for \( k \neq l \) of the FIM \( J_N \) become significantly smaller than the corresponding elements of diagonal blocks \( J_{N,k,k} \) as \( N \) increases. In order to characterize the asymptotic behavior of the desired matrices \( \frac{u+1}{N} A^H_{N,k} D_N^\alpha \Gamma_N (\omega_l - \omega_k) A_{N,l} \), it is convenient to rewrite the second term of (11) using

\[
\frac{1}{N} \sum_{j=0}^{N-1} E \left[ s_{N,k}(j) s_{N,k}(j)^H \right]^* \otimes \left[ e\left( \frac{j}{N} \right) \right] e\left( \frac{j}{N} \right)^H = \int_0^1 \mu_{N,k}(df) \otimes \left[ e(f) e(f)^H \right],
\]

(12)

where \( \mu_{N,k} \) denotes the following matrix-valued measure [19] defined for any Borel set \( A \) of \( [0,1] \) by

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

(13)

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). We denote by \( \mu_{N,k}^{t(t')}(A) \) the coefficient of the \( t \)th row and the \( t' \)th column of (13). In order to have some insights on the meaning of (13), it is interesting to remark that coefficient \( \mu_{N,k}^{(1,1)}(A) \) verifies \( \mu_{N,k}^{(1,1)}(A) = \frac{1}{N} \sum_{j=0}^{N-1} E \left[ |s_{N,k}^{(1)}(j)|^2 \right] \mathcal{I}_A\left( \frac{j}{N} \right) \). Therefore, measure \( \mu_{N,k}^{(1,1)} \) can be interpreted as the power profile of the training sequence sent at the first antenna. In particular, \( \mu_{N,k}^{(1,1)}([0,1]) \) represents the total power transmitted by the first antenna during a whole OFDM block. Generalizing this idea, for any transmit antenna pair \( (t,t') \), \( \mu_{N,k}^{(t,t')} \) can be interpreted as the cross-correlation profile of the training sequences respectively.
sent at antennas \( t \) and \( t' \). Finally, the matrix-valued measure \( \mu_{N,k} \) is in some sense equivalent to the (conjugate) covariance profile of the multi-dimensional training sequence \( s_{N,k}(j) \).

In order to simplify the forthcoming asymptotic study, we now make the following assumption.

**Assumption 1:** For each \( k \), we assume that there is a matrix-valued measure \( \mu_k \) such that \( \mu_{N,k} \) converges weakly to \( \mu_k \) as \( N \) tends to infinity.

In the sequel, we refer to \( \mu_k \) as the asymptotic covariance profile of the training sequence of user \( k \). The introduction of the above covariance profile thoroughly simplifies the asymptotic analysis of the CRB. Furthermore, Assumption 1 encompasses most usual CAS and power allocation strategies for OFDMA systems. Moreover, we shall see below that the asymptotic CRB associated with \( \hat{\theta} \) depends on the training strategy only via \( \mu_k \). Hence, the asymptotic covariance profile is sufficient to characterize the asymptotic CRB. No further assumptions on the particular CAS, the particular power allocation strategy or the particular correlation between antennas are required.

**Remark 2:** In order to have more insights on the meaning of measure \( A \rightarrow \mu_k(A) \), focus for instance on its first component \( A \rightarrow \mu_k^{(1,1)}(A) \) (i.e. the component at the first row and the first column). This component \( \mu_k^{(1,1)} \) is a classical scalar measure. Assume for the sake of illustration that \( \mu_k^{(1,1)} \) has a density \( P_k^{(1,1)}(f) \) w.r.t. the Lebesgue measure on \([0,1] \) (in other words, \( d\mu_k^{(1,1)}(f) = P_k^{(1,1)}(f)df \)). In this case, the density \( P_k^{(1,1)}(f) \) can be interpreted as the amount of power sent at the first antenna of user \( k \) in the neighborhood of frequency \( f \). With language abuse, \( P_k^{(1,1)}(f) \) is in some sense similar to the power density spectrum of the time-domain sequence transmitted at the first antenna of user \( k \). Of course, such a statement is somewhat non rigorous: in OFDMA, the time-domain transmitted sequence is not even stationary and, strictly speaking, its power density spectrum is not well defined. However, understanding \( P_k^{(1,1)}(f) \) as a power density spectrum may be useful in order to interpret the following results. Generalizing this idea, if the \( NT \times NT \) matrix-valued measure \( \mu_k \) has a matrix density \( P_k(f) \), (i.e., \( \mu_k(df) = P_k(f)df \)), then \( P_k(f) \) can be interpreted as the power density spectrum of the \( NT \)-dimensional sequence sent at all \( NT \) antennas of user \( k \).

Assumption 1 implies that for any continuous function \( f \rightarrow F(f) \), the integral \( \int_0^1 F(f)\mu_{N,k}(df) \) converges to \( \int_0^1 F(f)\mu_k(df) \) as \( N \) tends to infinity. Therefore, Lemma 1 along with equality (12) immediately leads to the following Lemma.

**Lemma 2:** For each \( k \), denote by \( R_k \) the following \( LN_T \times LN_T \) matrix

\[
R_k = \int_0^1 \mu_k(df) \otimes [\mathbf{e}(f)\mathbf{e}(f)^H].
\]
Then, for each \( u = 0, 1, 2 \), for each \( k, l = 1, \ldots, K \),
\[
\frac{u + 1}{N^{u+1}} A_{N,k}^H \mathbf{D}_{N}^u \Gamma_N (\omega_l - \omega_k) A_{N,l} \xrightarrow{N \to \infty} \delta(k - l) \mathbf{R}_k \quad \text{a.s.}
\]  

Using Lemma 2, one can now easily characterize the asymptotic behavior of each block (6) of the FIM. For example, due to Lemma 2, matrix \( \frac{1}{N} \mathbf{U}_{N,k,l} \) defined in (7) converges a.s. towards \( \delta(k - l) \mathbf{I}_{N,r} \otimes \mathbf{R}_k \) as \( N \) tends to infinity. Similarly, coefficient \( \frac{1}{N^3} \alpha_{N,k,l} \) given by (9) tends to \( \frac{1}{3} \delta(k - l) \gamma_k \), where
\[
\gamma_k = h_k^H \left[ \mathbf{I}_{N,k} \otimes \mathbf{R}_k \right] h_k = \sum_{r=1}^{N} \int_{0}^{1} h_k^{(r)}(f) H \mu_k(df) h_k^{(r)}(f).
\]

Here, \( h_k^{(r)}(f) = \sum_{l=0}^{L-1} h_k^{(1,r)}(l)\ldots, h_k^{(N_r,r)}(l)\right)^T e^{-2\pi i f l} \) may be interpreted as the overall frequency response of the channel “seen” at receive antenna \( r \). Following this kind of ideas, it is straightforward to characterize the asymptotic behavior of \( \mathbf{J}_N \) and thus of \( \text{CRB}_N \). Remark that Lemma 2 suggests to rather study the asymptotic behavior of the normalized CRB. We define \( \text{CRB}_N = \mathbf{W}_N \text{CRB}_N \mathbf{W}_N \) where \( \mathbf{W}_N \) is the \( K(1+2LN_RN_T) \times K(1+2LN_RN_T) \) diagonal matrix defined by \( \mathbf{W}_N = \text{diag}(\mathbf{w}_N^1, \ldots, \mathbf{w}_N^T) \) where \( \mathbf{w}_N^T \) denotes the \((1+2LN_RN_T)\) row vector
\[
\mathbf{w}_N^T = \left[ N^{3/2}, N^{1/2}, \ldots, N^{1/2} \right].
\]

Using Lemma 2 in the way described above, we obtain the following result.

**Theorem 1:** As \( N \) tends to infinity, the normalized CRB \( \text{CRB}_N \) converges almost surely to the block-diagonal matrix \( \overline{\text{CRB}} \) given by \( \overline{\text{CRB}} = \text{diag}(\mathbf{C}_1, \ldots, \mathbf{C}_K) \). For each \( k = 1, \ldots, K \), \( \mathbf{C}_k \) is the \((1+2LN_RN_T) \times (1+2LN_RN_T) \) matrix equal to
\[
\mathbf{C}_k = \frac{\sigma^2}{2} \begin{bmatrix}
\frac{12}{\gamma_k} & 6 h_k^T / \gamma_k & -6 h_k^T / \gamma_k \\
6 h_k / \gamma_k & 3 h_k h_k^T / \gamma_k & -3 h_k h_k^T / \gamma_k \\
-6 h_k / \gamma_k & 3 h_k h_k^T / \gamma_k & 3 h_k h_k^T / \gamma_k
\end{bmatrix}
\]

**Proof:** Using Lemma 2, it can be shown that the normalized FIM \( \overline{\mathbf{J}}_N = \mathbf{W}_N^{-1} \overline{\mathbf{J}}_N \mathbf{W}_N^{-1} \) converges a.s. to a block-diagonal matrix \( \overline{\mathbf{J}} = \text{diag}(\mathbf{J}_1, \ldots, \mathbf{J}_K) \) as \( N \to \infty \). For finite values of \( L \), it is reasonable to assume that \( \overline{\mathbf{J}} \) is non-singular. As function \( \overline{\mathbf{J}}_N \to \overline{\mathbf{J}}_N^{-1} \) is continuous, the normalized CRB converges a.s. to \( \overline{\text{CRB}} = \overline{\mathbf{J}}^{-1} = \text{diag}(\mathbf{J}_1^{-1}, \ldots, \mathbf{J}_K^{-1}) \). Final expression of the asymptotic CRB can thus be obtained by separate inversion of \( \mathbf{J}_1, \ldots, \mathbf{J}_K \). Using derivations similar to [12], the final result is straightforward.

We now make the following comments.
Comments

- The asymptotic normalized CRB is a block-diagonal matrix. In particular, this implies that for any asymptotically efficient estimator, the (normalized) estimation errors corresponding to parameters of distinct users become non correlated as $N$ tends to infinity.

- Theorem 1 provides asymptotic bounds on the Mean Square Error (MSE) for the parameters of a given user $k$. In the sequel, in order to simplify the notations, $E_N[\cdot]$ denotes the conditional expectation w.r.t. training sequences, i.e., $E_N[X] = E\left[X / (s_{N,1}(j))_j, \ldots, (s_{N,K}(j))_j\right]$ for any random variable $X$. In particular, $E_N[X]$ is a random variable which depends on the training sequences.

**Corollary 1:** For any unbiased estimate $\hat{\theta}_N$ of $\theta$, Proposition 1 implies that the following inequalities hold with probability one (w.p.1).

$$\liminf_{N \to \infty} N^3 E_N \left[ (\hat{\omega}_{N,k} - \omega_k)^2 \right] \geq \frac{6\sigma^2}{\gamma_k}$$

$$\liminf_{N \to \infty} N E_N \left[ \|\hat{\mathbf{h}}_{N,k} - \mathbf{h}_k\|^2 \right] \geq N_0 \sigma^2 \text{tr} \left( \mathbf{R}_k^{-1} \right) + \frac{3\sigma^2}{\gamma_k} \frac{\mathbf{h}_k^H \mathbf{h}_k}{2}$$

where $\hat{\omega}_{N,k}$ and $\hat{\mathbf{h}}_{N,k}$ respectively denote the estimates of the (angular) frequency offset $\omega_k$ and channel coefficients $\mathbf{h}_k$. In the above expression, $\text{tr}(\mathbf{X})$ stands for the trace of $\mathbf{X}$ and $\|\mathbf{x}\|^2 = \mathbf{x}^H \mathbf{x}$ for any column vector $\mathbf{x}$.

- The MSE on channel parameters converges to zero at rate $\frac{1}{N}$ while the MSE on frequency offsets converges to zero at rate $\frac{1}{N^3}$.

- For a given user $k$, the bounds on the MSE for estimates of $\theta_k$ do not depend on the values of parameters $\theta_l$ associated with other users $l \neq k$. It is also worth noting that the bounds do not depend on the number $K$ of users. In other words, the asymptotic CRB associated with parameters of the $k$th user is identical to the asymptotic CRB that one would have obtain in the absence of other users $l \neq k$.

- The asymptotic CRB does not depend on the value of the different frequency offsets.

- Asymptotic bounds for the MSE of unbiased estimates of $\theta$ depend on the training scheme only via the asymptotic covariance profiles $\mu_k$. It is worth noting that two different training schemes may have identical asymptotic covariance profile $\mu_k$ and are thus likely to lead to similar estimation performance. In order to illustrate this claim, consider the following non-orthogonal training strategy $T_1$. Each user $k$ modulates all subcarriers $j = 0, \ldots, N - 1$, with equal power $P_k$. In this case,

---

1A given subcarrier is possibly modulated by several users
the proportion of the bandwidth assigned to user \( k \) is equal to \( N_k/N = 1 \). Now consider a second training strategy \( T_2 \) for which each user \( k \) modulates the set of \( \frac{N}{K} \) subcarriers with the following indices

\[
\left\{ iK + k - 1/i = 0, \ldots, \frac{N}{K} - 1 \right\}
\]

with equal power \( K\mathcal{P}_k \) (for the sake of simplicity, ratio \( \frac{N}{K} \) is assumed to be an integer). One usually refer to this orthogonal CAS as \textit{interleaved} OFDMA [8]. For the sake of illustration, assume that training sequences transmitted by different antennas are uncorrelated in both cases \( T_1 \) and \( T_2 \).

Then, based on definition (13), it is straightforward to show that for both training schemes \( T_1 \) and \( T_2 \), covariance profile \( \mu_{N,k} \) converges weakly to the asymptotic “frequency flat” covariance profile \( \mu_k(A) = \mathcal{P}_k\mathbf{I}_{N_T}\lambda(A) \), where \( \lambda \) is the Lebesgue measure on \([0, 1]\). As a consequence, both training strategies \( T_1 \) and \( T_2 \) lead to the same asymptotic CRB.

This example illustrates the fact that orthogonal and non-orthogonal CAS both have the same asymptotic performance bounds, as long as they have an identical asymptotic covariance profile \( \mu_k \). Intuitively, this means that it is not essential for a given user \( k \) to transmit pilot symbols over all available subcarriers in order to obtain the most accurate estimate of \( \theta_k \).

C. Frequency domain estimation of channel parameters

As mentioned in Section II, one should now investigate the case where the OFDMA receiver aims to estimate the channel’s response associated with a given user \( k \) only at the subcarriers which are effectively modulated by user \( k \), rather than to estimate the time-domain channel coefficients \( h_k \). This case is of particular interest in contexts where a given user is constrained to transmit in a frequency band which is strictly smaller than the total bandwidth of the system [17]. In such a situation, estimation of the channel outside the useful frequency band is of few interest. For this reason, we now study the performance of estimates of the channel only at the frequencies which are effectively used by the user. In the sequel, we consider a given a given user \( k = 1, \ldots, K \). We denote by \( \Xi_{N,k} \) the subset of \( \{0, \ldots, N - 1\} \) corresponding to the subcarriers modulated by user \( k \). We denote by \( j_0 < j_1 < \cdots < j_{N_k - 1} \) the elements of \( \Xi_{N,k} \). For a given transmit antenna \( t \) of user \( k \) and for a given receive antenna \( r \), we denote by

\[
g_{N,k}^{(t,r)}(i) = \sum_{l=0}^{L-1} h_k^{(t,r)}(l)e^{-2\pi ij_i/N}
\]

the frequency tap of the channel at subcarrier \( j_i \). Defining \( g_{N,k}^{(t,r)} = [g_{N,k}^{(t,r)}(0), \ldots, g_{N,k}^{(t,r)}(N_k - 1)]^T \), one obtains the following simple relation between the frequency taps of interest and the initial time-domain
channel coefficients: \( g_{N,k}^{(l,r)} = \Phi_{N,k} h_{N,k}^{(l,r)} \), where \( \Phi_{N,k} \) is the \( N_k \times L \) matrix

\[
\Phi_{N,k} = \left( e^{-2\pi i j/N} \right)_{0 \leq j \leq N_k - 1},
\]

The desired parameter vector for a given user \( k \) is defined by \( g_{N,k} = [g_{N,k}^{(1,1)^T}, \ldots, g_{N,k}^{(N_T,1)^T}, \ldots, g_{N,k}^{(1,N_k)^T}, \ldots, g_{N,k}^{(N_T,N_k)^T}]^T \) and verifies \( g_{N,k} = (I_{N_T N_R} \otimes \Phi_{N,k}) h_k \). Parameter \( g_{N,k} \) is thus a linear function of the initial set of parameters. Recall that the CRB associated with any linear function \( G \theta \) (where \( G \) is any matrix) of the initial parameter vector \( \theta \) is given by \( G(\text{CRB}_N)G^T \) (see [18]). Using this result, it is straightforward to show that for any unbiased estimator \( \hat{g}_{N,k} \) of \( g_{N,k} \), the normalized MSE on channel frequency taps verifies the following inequality:

\[
\frac{1}{N_k} E_N \left[ \| \hat{g}_{N,k} - g_{N,k} \|^2 \right] \geq \frac{1}{N} \text{tr} \left( \begin{bmatrix}
\text{CRB}_{N,k,k} & 0 & 0^{L N_T N_R}^T \\
0 & I_{N_T N_R} \otimes \Re [T_{N,k}] & 0^{L N_T N_R}^T \\
0 & I_{N_T N_R} \otimes \Im [T_{N,k}] & I_{N_T N_R} \otimes \Re [T_{N,k}] 
\end{bmatrix} \right),
\]

where \( \text{CRB}_{N,k,k} \) represents the \( k \)th \( (2 LN_T N_R + 1) \times (2 LN_T N_R + 1) \) diagonal block of the normalized CRB matrix associated with initial parameter vector \( \tilde{\theta} \), i.e., \( \text{CRB}_N = (\text{CRB}_{N,k,l})_{k,l=1,\ldots,K} \), and where \( T_{N,k} \) denotes the \( L \times L \) matrix given by

\[
T_{N,k} = \frac{1}{N_k} \Phi_{N,k}^H \Phi_{N,k}.
\]

Notation \( 0_{L N_T N_R} \) stands for the \( L N_T N_R \times 1 \) null vector. As previously, expression (22) provides few information on the performance of estimates of \( g_{N,k} \). However, previous results allow to provide a compact and informative version of (22) in the case where the number \( N \) of subcarriers increases. Indeed, the right-hand side of (22) only depends on two matrices \( \text{CRB}_{N,k,k} \) and \( T_{N,k} \). Theorem 1 states that the first of these two matrices, i.e., \( \text{CRB}_{N,k,k} \), converges a.s. toward matrix \( C_k \) given by (17) as \( N \) increases. In order to obtain a simple expression of (22), the only task is therefore to study the asymptotic behavior of \( T_{N,k} \). Based on (23), we first remark that for each \( p, q = 0, \ldots, L - 1 \), the coefficient of the \((p + 1)\)th row and the \((q + 1)\)th column of \( T_{N,k} \) can be written as

\[
\frac{1}{N_k} \sum_{j \in \mathbb{N}_N} e^{2\pi i (p-q) j/N} = \int_0^1 e^{2\pi i (p-q) f} \nu_{N,k}(df),
\]

where scalar measure \( \nu_{N,k} \) is defined for any Borel set \( A \) of \([0,1] \) by

\[
\nu_{N,k}(A) = \frac{1}{N_k} \sum_{j \in \mathbb{N}_N} \mathcal{T}_A \left( \frac{j}{N} \right).
\]

Measure \( \nu_{N,k} \) is introduced in order to simplify the asymptotic study of matrix \( T_{N,k} \). For any Borel set \( A \), \( \nu_{N,k}(A) \) is simply equal to the number of subcarriers \( j \) modulated by user \( k \) such that \( \frac{j}{N} \) lies
in $A$, divided by the number $N_k$ of modulated subcarriers. In particular, $\nu_{N,k}([0,1]) = 1$, i.e., $\nu_{N,k}$ is a probability measure. We now make the following assumption.

**Assumption 2:** There is a measure $\nu_k$ such that $\nu_{N,k}$ converges weakly to $\nu_k$ as $N$ tends to infinity. Based on the above assumption, we can directly write that (24) converges to $\int_{0}^{1} e^{2\pi i (p-q)f} \nu_k(df)$ as $N$ tends to infinity. Finally, $\mathbf{T}_{N,k}$ converges a.s. to the following $L \times L$ matrix $\mathbf{T}_k$ defined by

$$
\mathbf{T}_k = \int_{0}^{1} \mathbf{e}(f)\mathbf{e}(f)^H \nu_k(df),
$$

where we recall that $\mathbf{e}(f) = [1, e^{2\pi i f}, \ldots, e^{2\pi i f(L-1)}]^T$. In the sequel, we furthermore assume that limit measure $\nu_k$ has a density w.r.t. the Lebesgue measure on $[0,1]$ and we denote this density by $D_k(f)$ (i.e., $\nu_k(df) = D_k(f)df$). Intuitively, function $D_k(f)$ can be interpreted as the (limit) density of subcarriers modulated by user $k$. This density is normalized in such a way that $\int_{0}^{1} D_k(f)df = 1$. For instance, if all subcarriers are modulated by $k$, $D_k(f) = 1$ for each $f$. On the opposite, if user $k$ does not modulate any subcarrier inside a certain frequency interval $A \subset [0,1]$, then $D_k(f) = 0$ for each $f \in A$. In the sequel, we denote by $\mathcal{D}_k = \{ f \in [0,1]/D_k(f) > 0 \}$ the part of bandwidth used by transmitter $k$. For most practical CAS, it is reasonable to assume that the density $D_k(f)$ of modulated subcarriers is a constant for each $f \in \mathcal{D}_k$. In this case, $D_k(f)$ coincides with the following “frequency mask”:

$$
D_k(f) = \frac{1}{\Delta_k} \mathcal{I}_{\mathcal{D}_k}(f),
$$

where $\Delta_k = \int_{\mathcal{D}_k} df$ is a constant equal to the Lebesgue measure of $\mathcal{D}_k$ and where $\mathcal{I}_{\mathcal{D}_k}(f)$ is the indicator function of set $\mathcal{D}_k$. For instance, assumption (27) encompasses the traditional group-orthogonal CAS [6][7] as well as the interleaved CAS [8]. Although our results can be generalized to more general forms of densities $D_k(f)$, we assume for the sake of simplicity that (27) holds. In this case, $\mathbf{T}_k$ coincides with $\mathbf{T}_k = \frac{1}{\Delta_k} \int_{\mathcal{D}_k} \mathbf{e}(f)e(f)^H df$. We may now put all pieces together in order to express the limit of the righthand side of (22) as $N$ tends to infinity. Based on Theorem 1, it is straightforward to show that (22) reduces to:

$$
\liminf_{N \to \infty} \frac{N}{N_k} E_N \left[ \| \hat{\mathbf{g}}_{N,k} - \mathbf{g}_{N,k} \|^2 \right] \geq \sigma^2 N_R \text{tr} \left( \mathbf{R}_k^{-1} (\mathbf{I}_N \otimes \mathbf{T}_k) \right) + \frac{3\sigma^2}{2} \frac{\beta_k}{\gamma_k},
$$

where $\beta_k = \frac{1}{\Delta_k} \sum_{r=1}^{N} \int_{\mathcal{D}_k} \| h_k^{(r)}(f) \|^2 df$. Before commenting this result, it is worth noting that inequality (28) can be further simplified as long as we assume that the length $L$ of the channels impulse responses is large enough. Indeed, when $L$ increases, the sizes of matrices $\mathbf{R}_k$ and $\mathbf{T}_k$ in the righthand side of (28) also increase. Then, as we shall see below, the first term of the righthand side of (28) becomes predominant, so that the term $\frac{3\sigma^2}{2} \frac{\beta_k}{\gamma_k}$ can be neglected. Moreover, it can be easily seen that $\mathbf{R}_k$ and $\mathbf{T}_k$
are both block-Toeplitz matrices. As their sizes increase with \( L \), results on the behavior of large block-Toeplitz matrices can be used to simplify the first term of the righthand side of \( (28) \). Ultimately, such an analysis will lead us to obtain a lower bound on 

\[
\lim_{L \to \infty} \left( \lim_{N \to \infty} \frac{N}{N_k} E_N \left[ \| \hat{g}_{N,k} - g_{N,k} \|^2 \right] \right).
\]

We infer that this limit will be a relevant performance bound in practical situations where \( N \) and \( L \) are both large, but \( L \ll N \). This claim will be sustained by simulations of Section V.

**Theorem 2:** Consider a given user \( k = 1, \ldots, K \). Assume that \( \mu_k \) has a matrix-valued density \( P_k(f) \) w.r.t. the Lebesgue measure on \( [0,1] \) (i.e., \( \mu_k(df) = P_k(f)df \)). For each \( f \), denote by \( \lambda_k^{(1)}(f) < \cdots < \lambda_k^{(N_T)}(f) \) the (non-negative) eigenvalues of \( P_k(f) \). Assume that \( \lambda_k^{(1)}(f) \geq \epsilon \) for each \( f \in D_k \) and for some \( \epsilon > 0 \). Then,

\[
\lim_{N \to \infty} \frac{N}{N_k} E_N \left[ \| \hat{g}_{N,k} - g_{N,k} \|^2 \right] \geq \sigma^2 L N R \frac{\Delta_k}{\Delta_k} \sum_{t=1}^{N_T} \int_{D_k} \frac{1}{\lambda_k^{(t)}(f)} df + o_L(L) \tag{29}
\]

w.p.1, where \( o(L) \) represents a deterministic term such that \( \frac{o(L)}{L} \) tends to zero as \( L \) tends to infinity.

**Proof:** See Appendix II.

**Comments**

- Theorem 2 provides a bound on the MSE of unbiased estimators of the channel’s frequency taps of each user \( k \). In practice, the bound \( (29) \) is relevant as long as both the number \( N \) of subcarriers and the length \( L \) of the channel impulse response are large enough, but with \( N \gg L \).
- The asymptotic CRB on estimates of \( \hat{g}_{N,k} \) is proportional to the length \( L \) of the channel, for large values of \( L \).
- In order to illustrate the meaning of Theorem 2, it is useful to investigate the case where each user has a single transmit antenna: \( N_T = 1 \). In this case, the density \( P_k(f) \) of the asymptotic covariance profile reduces to a scalar density \( P_k(f) \). As mentioned previously, in some sense, the density \( P_k(f) \) may be interpreted as the amount of power transmitted by user \( k \) in a neighborhood of frequency \( f \). In this case, the bound on the MSE of \( \hat{g}_{N,k} \) in the righthand side of equation \( (29) \) becomes

\[
\frac{\sigma^2 L N R}{\Delta_k} \int_{D_k} \frac{1}{P_k(f)} df + o_L(L).
\]

Therefore, as \( L \) increases, the above asymptotic CRB tends to be proportional to the average of the inverse of the power density.

**IV. Training Sequence Selection**

Theorems 1 and 2 above provide asymptotic bounds on the MSE of channels and frequency offsets estimates. These bounds crucially depend on the training strategies used by each transmitter \( k \) via the
asymptotic covariance profiles $\mu_k$. Thus, following the approach of previous studies [11][12], it is natural to search for the training strategies which minimize these bounds. More precisely, we provide guidelines on the way each user has to design its own training sequence so that the (bound on) the estimation performance is the smallest. For instance, one may wonder which power should be allocated to which subcarriers, and what correlation should be introduced between training sequences sent at different antennas so as to lead to accurate estimates of the parameters.

Based on inequalities (18) and (29), our approach consists in minimizing the following asymptotic bounds

$$\text{asCRB}_{\omega_k} = \frac{6\sigma^2}{\gamma_k}$$

$$\text{asCRB}_{g_{N,k}} = \frac{\sigma^2 LN_R}{\Delta_k} \sum_{t=1}^{Nr} \frac{1}{D_k} \int \lambda_k^{(t)}(f) df$$

for each $k = 1, \ldots, K$, under the constraint that the total power transmitted by each user $k$ does not exceed a certain predefined constant, say $P_k$. Of course, such an approach will be relevant when the number $N$ of subcarriers is large enough. Furthermore, the minimization of $\text{asCRB}_{g_{N,k}}$ in equation (31) is relevant if the length $L$ of the channel impulse response is also large enough (but $N \gg L$), as suggested by Theorem 2. It is worth noticing that for a given $k$, $\text{asCRB}_{\omega_k}$ and $\text{asCRB}_{g_{N,k}}$ only depend on the training sequence of user $k$ and depend neither on the training strategies $\mu_l$ corresponding to other users $l \neq k$ nor on parameters $\theta_l$ associated with other users. This remark is of practical importance. Indeed, it implies that the selection of the training sequence of a given user $k$ can be achieved without any knowledge of other users’ training strategies and parameters.

Of course, one would expect from an ideal training strategy that it simultaneously minimizes both bounds (30) and (31). Unfortunately, results of [11][12][15] tend to show that no single training sequence is likely to jointly minimize both bounds. In order to overcome this problem, [12] proposes to select training sequences so that a given cost function depending on the CRB is minimum. However, the problem of choosing a relevant cost function is difficult and crucially depends on the transmitter and receiver architectures [20]. This issue is out of the scope of the present paper. Here, we focus on separate minimization of $\text{asCRB}_{\omega_k}$ and $\text{asCRB}_{g_{N,k}}$. 
A. Minimization of \( asCRB_{\omega_k} \)

We minimize \( asCRB_{\omega_k} \) under the constraint that the total power transmitted by user \( k \) does not exceed \( P_k \). The latter power constraint is equivalent to the following inequality:

\[
\text{tr}(\mu_k([0,1])) \leq P_k. \tag{32}
\]

As mentioned in Section III, the above expression of the power constraint is a direct consequence of (13). Inequality (32) can be motivated as follows. For finite values of the number \( N \) of subcarriers, the total transmit power for user \( k \) is the sum of the powers transmitted by all \( N_T \) antennas of user \( k \), i.e.,

\[
\sum_{t=1}^{N_T} \frac{1}{N} \sum_{j=0}^{N-1} E\left[|s_{N,k}(j)|^2\right] = \text{tr}(\mu_{N,k}([0,1])).
\]

For large \( N \), the above transmit power simply converges to \( \text{tr}(\mu_k([0,1])) \). This motivates inequality (32). Assuming the above transmit power constraint, training sequences which minimize the bound \( asCRB_{\omega_k} \) on the asymptotic performance of estimates of the frequency offset can be selected according to Proposition 1.

Proposition 1: For each \( f \in [0,1] \), recall that \( h_k^{(r)}(f) = \sum_{l=0}^{L-1}[h_k^{(1,r)}(l), \ldots, h_k^{(N_T,r)}(l)]^T e^{-2\pi fl} \). Denote by \( \lambda_{k,\text{max}}(f) \) the largest eigenvalue of matrix \( \sum_{r=1}^{N_R} h_k^{(r)}(f) h_k^{(r)}(f)^H \). Define \( f_f^{(\text{opt})} = \arg \max_f \lambda_{k,\text{max}}(f) \).

Under power constraint (32), \( asCRB_{\omega_k} \) is minimum if and only if

\[
\mu_k = P_k \left( \nu_k^{(\text{opt})} \nu_k^{(\text{opt})H} \right) \delta_{f_f^{(\text{opt})}}, \tag{33}
\]

where \( \nu_k^{(\text{opt})} \) is the (unit norm) eigenvector associated with \( \lambda_{k,\text{max}}(f_f^{(\text{opt})}) \) and where \( \delta_{f_f^{(\text{opt})}} \) is the Dirac measure at \( f_f^{(\text{opt})} \).

The proof of Proposition 1 is provided in Appendix III. Proposition 1 states that training sequences which are likely to lead to the most accurate estimate of the frequency offset \( \omega_k \) have an asymptotic covariance profile \( \mu_k \) defined by (33). In other words, Proposition 1 describes the best use of the available power \( P_k \) for the aim of frequency offset estimation. We now comment this result.

Comments

- Proposition 1 suggests that an accurate estimate of \( \omega_k \) can be obtained by transmitting all available power at the frequency for which the largest eigenvalue of \( \sum_{r=1}^{N_R} h_k^{(r)}(f) h_k^{(r)}(f)^H \) is maximum. In the single antenna case, i.e., \( N_T = 1, N_R = 1 \), optimal asymptotic covariance profile (33) simply reduces to the scalar measure \( P_k \delta_{f_f^{(\text{opt})}} \) where

\[
\mu_k = P_k \left( \nu_k^{(\text{opt})} \nu_k^{(\text{opt})H} \right) \delta_{f_f^{(\text{opt})}},
\]

\[
f_f^{(\text{opt})} = \arg \max_{f \in [0,1]} \left| \sum_{l=0}^{L-1} h_k(l) e^{-2\pi fl} \right|^2.
\]
In this case, Proposition 1 suggests to transmit all available power at the frequency for which the magnitude of the channel frequency response is the largest. For finite values of the number $N$ of subcarriers, a way of following such a guideline is to allocate most of the power at the subcarriers which are close to the optimal frequency $f_k^{(\text{opt})}$.

- Proposition 1 also states that $asCRB_{\omega_k}$ is minimum when the covariance matrix between training sequences sent at different transmit antennas coincides with $\nu_k^{(\text{opt})} \nu_k^{(\text{opt})H}$ (up to a multiplicative factor). In practice, for finite values of $N$ and for a given subcarrier $j$ whose frequency $j/N$ is close to the optimal frequency $f_k^{(\text{opt})}$, this guideline may be followed by defining the multi-dimensional training sequence $s_{N,k}(j) = [s_{N,k}^{(1)}(j), \ldots, s_{N,k}^{(N_T)}(j)]^T$ as $s_{N,k}(j) = \nu_k^{(\text{opt})} w_{N,k}(j)$, where $w_{N,k}(j)$ is a certain scalar training sequence. Furthermore, the previous comment suggests to select $w_{N,k}(j)$ as a sequence which non zero only at subcarriers $j$ such that $j/N$ is close to $f_k^{(\text{opt})}$.

- The training strategy suggested by Proposition 1 requires some limited channel knowledge consisting in $f_k^{(\text{opt})}$ and $\nu_k^{(\text{opt})}$. As explained and discussed in [13], this limited information can be provided to the transmitter using a downlink control channel.

- The selection of training sequences w.r.t. the guideline provided by Proposition 1 is likely to provide accurate estimate of the frequency offset. Nevertheless, such a selection may be impractical as far as channel estimation is concerned. For instance, if $\mu_k$ is chosen as in (33), then matrix $R_k$ defined by (14) becomes singular. In this case, the asymptotic bound on channel estimates given by the righthand side of (19) becomes infinite. This practically means that channel coefficients cannot be properly estimated if training sequence selection is achieved strictly as dictated by Proposition 1.

This observation confirms that no training sequence allows to jointly provide the most accurate estimates of both the frequency offset and the channel coefficients. In practice, determination of tradeoffs between accurate frequency offset estimation and accurate channel estimation is required.

**B. Minimization of $asCRB_{g_{N,k}}$**

We now study the training sequences which minimize the bound $asCRB_{g_{N,k}}$ given by (31) on the MSE associated with estimates of the (useful) channel frequency taps. Again, the minimization is achieved under power constraint (32). As mentioned previously, we suppose that Assumptions of Theorem 2 are valid. In particular, we assume that $\mu_k$ has a matrix density $P_k(f)$ whose eigenvalues are denoted by $\lambda_k^{(1)}(f) < \cdots < \lambda_k^{(N_T)}(f)$. Training sequences which minimize the bound $asCRB_{g_{N,k}}$ on the asymptotic performance of estimates of the useful channel frequency taps can be selected according to Proposition 2.

**Proposition 2:** Under power constraint (32), $asCRB_{g_{N,k}}$ is minimum if and only if for each $f \in [0, 1]$, $f_k^{(\text{opt})}$.
\[ P_k(f) = \frac{P_k}{N_T} I_{N_T} \mathcal{I}_{D_k}(f), \] where \( I_{N_T} \) is the \( N_T \times N_T \) identity matrix and where \( \mathcal{I}_{D_k}(f) \) is the indicator function of the useful bandwidth \( D_k \).

**Proof:** Based on definition (31), the minimization of \( \text{asCRB}_{g_{N,k}} \) under power constraint (32) is equivalent to the minimization of

\[
\sum_{t=1}^{N_T} \int_{D_k} \frac{1}{\lambda_k^{(t)}(f)} df \int_{D_k} \lambda_k^{(t)}(f) df \geq \Delta_k.
\]  

(34)

Note that equality holds in (34) if and only if \( \lambda_k^{(t)}(f) \) is a constant w.r.t. \( f \) in \( D_k \). Using successively (34) and Jensen’s inequality, we obtain

\[
\text{asCRB}_{g_{N,k}} \geq \sigma^2 L N R \frac{1}{\sum_{t=1}^{N_T} \int_{D_k} \lambda_k^{(t)}(f) df} N_T^2
\]

(35)

Equality holds in (35) if and only if \( i) \lambda_k^{(t)}(f) \) is a constant w.r.t. \( f \in D_k \) and \( ii) \) all elements of \( \left( \int_{D_k} \lambda_k^{(t)}(f) df \right)_{t=1,\ldots,N_T} \) are identical. This is equivalent to: \( \forall f \in D_k, \forall t, \lambda_k^{(t)}(f) = \lambda_k \) where \( \lambda_k \) is a certain constant which is independent from both \( f \) and \( t \). In this case, matrix \( P_k(f) \) is proportional to the identity matrix, i.e., \( P_k(f) = \lambda_k I_{N_T} \) for each \( f \in D_k \). Finally, the denominator of the righthand side of (35) coincides with \( \text{tr}(\int_{D_k} P_k(f) df) \). Thus, it is less than or equal to power constraint \( P_k \) and achieves \( P_k \) if and only if \( \lambda_k = \frac{P_k}{N_T} \) and \( P_k(f) = 0 \) for \( f \notin D_k \). Thus, \( \text{asCRB}_{g_{N,k}} \) is minimum if and only if \( P_k(f) = \frac{P_k}{N_T} I_{N_T} \mathcal{I}_{D_k}(f) \)

**Comments**

- We recall that the non diagonal elements of matrix \( P_k(f) \) can be interpreted as the cross-correlation between training sequences sent at different antennas, in a neighborhood of frequency \( f \). Proposition 2 states that for each \( f \in [0,1] \), the asymptotic covariance density \( P_k(f) \) is a diagonal matrix. This suggests in particular that accurate estimates of channel coefficients can be obtained as long as training sequences sent at different antennas are uncorrelated.

- Proposition 2 indicates that matrix \( P_k(f) \) should be constant w.r.t. \( f \) in the bandwidth of interest \( D_k \) so as to minimize \( \text{asCRB}_{g_{N,k}} \). In other words, if the length \( L \) of the channel is large enough, the uniform power allocation in the modulated part of the bandwidth is likely to provide the most accurate estimates of the desired channel coefficients.
V. SIMULATION RESULTS

In order to illustrate and sustain our claims, we present simulation results considering MIMO-OFDMA transmission with QPSK signaling. The number $N_T$ of transmit antennas is set to 1 or 2 depending on the context. The number of receive antennas is set to $N_R = 2$. The shaping pulse is a raised cosine filter with roll-off 0.25. For each user and for each transmit-receive antenna pair, we consider a multipath fading channel with $Q$ independent paths. For each channel realization, the number of paths $Q$ is chosen uniformly between 2 and 4. Complex gains associated with each path are assumed to be circular complex Gaussian random variables with zero mean and unit variance. Delays of all paths are chosen from the uniform distribution on interval $[0, 5T]$. The length $L$ of the channel parameter vector to be estimated is set to $L = 8$. For each user $k$, the value of $\omega_k$ is randomly chosen in the interval $[-0.01, 0.01]$. Average transmitted powers $P_k$ are supposed to be equal for all users. In the sequel, without loss of generality, we focus on the results corresponding to the first user $k = 1$. All results are averaged over 1000 realizations of the training sequences and the channel parameters.

A. Comparison of exact and asymptotic CRB

In this paragraph, we study the values of $N$ for which our asymptotic results provide an accurate approximation of the exact CRB. Figure 1a compares exact and asymptotic CRB associated to channel parameters $g_{N,1}$, for different values of the number $N$ of subcarriers. The ratio $P_1/N_0$ between the transmitted energy per symbol and the noise power density is set to 20dB. The number of users is set to either 2 or 4. Two different training strategies are studied. In the first case, each user is assumed to use the training strategy $T_1$ depicted in Section III-B (i.e., all subcarriers are modulated by all users with equal power, training sequences sent at different antennas are uncorrelated). In the second case, training strategy $T_2$ is used (i.e., subcarriers are assigned following an interleaved CAS, assigned subcarriers are modulated with equal power, training sequences sent at different antennas are uncorrelated). The solid line represents the value $\frac{1}{N} asCRB_{L,g_{N,1}}$, where $asCRB_{L,g_{N,1}}$ denotes the asymptotic CRB on channel frequency taps equal to the righthand side of (28). Note that the value of $\frac{1}{N} asCRB_{L,g_{N,1}}$ is identical for $T_1$ and $T_2$. Indeed, the asymptotic CRB only depends on the training sequence via the asymptotic covariance profile $\mu_1$ and that both training strategies $T_1$ and $T_2$ lead to the same asymptotic covariance profile. The exact CRB on channel frequency taps is defined as the righthand side of (22). As shown by Figure 1a, the exact CRB depends on the particular training strategy $T_1$, $T_2$. Figure 1a shows however that, as long as the number $N$ of subcarriers is large enough, exact CRB corresponding to $T_1$ and $T_2$ respectively both fit to the asymptotic CRB. This sustains the claim that both training strategies have a similar performance.
for large $N$. It is also worth noting that the exact CRB is close to the asymptotic CRB even for moderate values of $N$. The dotted line represents the value of $\frac{1}{N} \text{asCRB}_{g_{N,1}}$, where $\text{asCRB}_{g_{N,1}}$ is given by (31). Due to Theorem 2, $\text{asCRB}_{g_{N,1}}$ approximates the asymptotic CRB $\text{asCRB}_{L,g_{N,1}}$ of large values of $L$. In Figure 1, we observe however that solid and dotted lines (i.e., respectively asymptotic CRB for finite $L$ and its dominant term for large $L$) are very close. In the present case, $L = 8$. This tends to show that the lower bound given by Theorem 2 is relevant even for moderate values of the length $L$ of the channel impulse response.

Figure 1b compares exact and asymptotic CRB for frequency offset estimation as a function of $N$. Again, the solid line represents $\frac{1}{N} \text{asCRB}_{\omega_1}$, where $\text{asCRB}_{\omega_1}$ is defined by (30). Other curves represent the exact CRB for $\omega_1$, defined as the coefficient at the first row and the first column of $\text{CRB}_N$. As expected, exact CRB corresponding respectively to training strategies $T_1$ and $T_2$ tend to be identical when $N$ increases. Both of them fit to the asymptotic bound.

### B. Estimation Performance

We now study the estimation performance associated to the frequency offset with $K = 2$ users and $N = 256$ subcarriers. The Maximum Likelihood (ML) estimator is used on the received signal to compute estimates of the unknown parameters. Figure 2 represents the corresponding MSE $E_N[(\hat{\omega}_{N,1} - \omega_1)^2]$ as a function of ratio $\frac{P_1}{N_0}$, for different training strategies. The MSE is compared with the asymptotic CRB $\frac{1}{N} \text{asCRB}_{\omega_k}$ in solid lines. Figures 2a and 2b illustrate the performance when the number of transmit antennas is set to $N_T = 1$ or $N_T = 2$ respectively. Results are given for two different training strategies. In the first case, the above interleaved training strategy $T_2$ is used. In the second case, the optimal training strategy for frequency offset estimation, say $T_3$, is used. This optimal strategy is depicted in Section IV. In the single antenna case, it consists in sending one training symbol at the frequency $j$ such that $\frac{j}{N}$ is close to $j^{(opt)}_1$. In the multiple transmit antenna case, relevant correlation between antennas given by vector $\nu^{(opt)}_1$ is furthermore introduced. For both training strategies, Figure 2 illustrates the fact that the MSE corresponding to the ML estimate of $\omega_1$ is close to the asymptotic CRB. This motivates the fact that the asymptotic CRB can be interpreted as a relevant indicator of the estimation performance of practical estimators. Figure 2a allows to compare the performance associated with both training strategies $T_2$ and $T_3$ and for a single transmit antenna. A gain of about 2.6dB in terms of $\frac{P_1}{N_0}$ is observed between training strategy $T_2$ and $T_3$. Of course, recall that $T_3$ can hardly be used in practice as it does not allow to accurately estimate channel coefficients. However, the CRB obtained when strategy $T_3$ is used can be interpreted as the best lower bound on estimates of the frequency offset. The best gain which can be
expected from the use of a “non-uniform” power allocation strategy instead of \( T_2 \) is thus of 2.6dB when \( N_T = 1 \). Figure 2b shows that this gain increases with the number of transmit antennas. When \( N_T = 2 \), the use of spatially correlated training sequences leads to a gain of more than 5dB compared to the case of uncorrelated training sequences.

VI. CONCLUSION

The performance of joint data-aided estimators of frequency offsets and channel parameters for MIMO-OFDMA uplink has been addressed in this paper. An exact expression of the Cramér-Rao bound has been derived. When all training sequences sent by all users are modeled as sequences of random variables, the above CRB can be shown to converge almost surely to a deterministic matrix as the number \( N \) of subcarriers tends to infinity. The analysis of this asymptotic CRB matrix allowed to conclude that for any asymptotically efficient estimator, the estimation performance associated with parameters of a given user \( k \) becomes identical to the performance which one would have obtain if all parameters of all other users were perfectly known. The MSE on channel parameters has been shown to converge to zero at rate \( \frac{1}{N} \) while the MSE on frequency offsets converges to zero at rate \( \frac{1}{N^2} \). Asymptotic performance bounds have been shown to depend on the training sequence only via its asymptotic covariance profile. The asymptotic covariance profiles which minimize the asymptotic bounds have been provided. It has been shown that accurate estimates of channel parameters are obtained by transmission of spatially uncorrelated training sequences with uniform frequency power profile. Accurate estimates of the frequency offsets are obtained by allocating most of the power of the training sequence at the appropriate frequency and by introducing a relevant correlation between transmit antennas.

APPENDIX

PROOF OF LEMMA 1

Given \( k, l = 1, \ldots, K \), transmit antenna pair \((t,t')\), and \( u = 0, 1, 2 \), we denote by \( \zeta^{(t,t')}_N,k,l(u,p,q) \) the \((p+1,q+1)\)th element of matrix \( \frac{u+1}{N^{u+1}} \mathbf{A}^{(t)}_N \mathbf{D}^{(t)}_N \mathbf{G}_N(\omega_l - \omega_k) \mathbf{A}^{(t')}_N \). Defining \( \Delta f_{lk} = \frac{\omega_l - \omega_k}{2\pi} \) and using (1), the coefficient \( \zeta^{(t,t')}_N,k,l(u,p,q) \) can be expressed as

\[
\zeta^{(t,t')}_N,k,l(u,p,q) = \sum_{n=0}^{N-1} a^{(t)}_{N,k}(n-p) * a^{(t')}_{N,l}(n-q) e^{2\pi\nu \Delta f_{lk}}
\]

\[
= \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} s^{(t)}_{N,k}(i) * s^{(t')}_{N,l}(j) e^{2\pi\nu \frac{(i-j)(i+j)}{N}} \psi_{N,u}(j-i + N \Delta f_{lk}). \tag{36}
\]
where for each \( x \), \( \psi_{N,u}(x) = \frac{n+1}{N+1} \sum_{n=0}^{N-1} n^u e^{2\pi \frac{n^u x}{N}} \). The following proof relies on the observation that for each \( x \) such that \( 0 < |x| \leq \frac{N}{2} \),

\[
|\psi_{N,u}(x)| \leq \frac{C_u}{|x|},
\]

(37)

where \( C_u \) is a constant. This inequality was recently used by [14]. For the sake of completeness, we provide a sketch of the proof. The claim can be easily shown for \( \psi_{N,0}(x) = e^{\pi \frac{(N-1)\pi x}{N} \sin x} \). Using the fact that \( |\sin \frac{x}{2}| \geq 2 |\frac{x}{N}| \) for \( |x| \leq \frac{N}{2} \). When \( u = 1 \), it can be shown that

\[
\psi_{N,1}(x) = \frac{1}{N \sin \left( \frac{\pi x}{N} \right)} \left( e^{2\pi \frac{(N-1)\pi x}{2N}} - e^{\pi \sin \left( \frac{\pi x}{N} \right)} \right).
\]

(38)

To obtain the desired bound (37), we first notice that the term enclosed with parenthesis in (38) is bounded. Indeed, using the triangular inequality, this term is less than \( 1 + e^{\pi \sin \left( \frac{\pi x}{N} \right)} \) \( \leq 2 \). The case \( u = 2 \) can be treated using similar arguments. The (tedious) proof is omitted due to the lack of space.

Now, using equation (37), Lemma 1 can be proved as follows. First, consider the case \( k = l \). In this case, our aim is to prove that for each \( p,q \),

\[
\chi_{N,k,k} = \frac{c(t,t,u)}{S_{N,k,k}} - \frac{1}{N} \sum_{i=0}^{N-1} E \left[ s_N(t,i)^* s_N(t,i) \right] e^{2\pi \frac{(N-u)i}{N}} - N_0 \quad \text{a.s.}
\]

(39)

Note that we omitted subscripts \( (t, t') \), \( u \) and indices \( (p,q) \) in the above definition for the sake of readability. In the case \( k = l \), note that \( \Delta f_{k,k} = 0 \) in (36). In order to show (39), we write \( \chi_{N,k,k} \) as a sum of two terms \( \chi_{N,k,k} = \chi_{N,k,k}^a + \chi_{N,k,k}^b \) where

\[
\chi_{N,k,k}^a = \frac{1}{N} \sum_{i=0}^{N-1} s_N(t,i)^* s_N(t,i) - E \left[ s_N(t,i)^* s_N(t,i) \right] e^{2\pi \frac{(N-u)i}{N}}
\]

(40)

and we prove the almost sure convergence to zero of both terms. We first study \( \chi_{N,k,k}^b \). In this case, the proof is quite similar to the proof of [14]. Again, we split the sum in (41) as \( \chi_{N,k,k}^b = \chi_{N,k,k}^{b,1} + \chi_{N,k,k}^{b,2} + \chi_{N,k,k}^{b,3} \) and prove that each of these terms tends a.s. to zero. Here, \( \chi_{N,k,k}^{b,1} \) coincides with the right-hand side of (41) except that the inner sum w.r.t. \( j \) is restricted to the set \( E_1^1 = \{ j = 0, \ldots, N - 1 / j \neq i, |j - i| \leq \frac{N}{2} \} \). Similarly, terms \( \chi_{N,k,k}^{b,2} \) and \( \chi_{N,k,k}^{b,3} \) correspond to the restriction of the inner sum of (41) to the sets \( E_1^2 = \{ j = 0, \ldots, N - 1 / j \neq i, \frac{N}{2} < j - i \leq N - 1 \} \) and \( E_1^3 = \{ j = 0, \ldots, N - 1 / j \neq i, N + 1 \leq j - i < -\frac{N}{2} \} \) respectively. We now prove that \( \chi_{N,k,k}^{b,1} \overset{\text{a.s.}}{\to} 0 \). To that end, we show that
\[ E[|X_{N,k,k}^{b,1}|^4] \leq \frac{C}{N^2}, \text{ where } C \text{ is a constant.} \]

We first expand \( E[|X_{N,k,k}^{b,1}|^4] \) as follows.

\[
E[|X_{N,k,k}^{b,1}|^4] = \frac{1}{N^4} \sum_{(i_1,i_2,i_3,i_4,j_2,j_3,j_4) \in V} E \left[ \left( \prod_{n=1}^{4} s_{N,k}^{(1)}(i_n)^* s_{N,k}^{(2)}(j_n) \right) \prod_{n=1}^{4} \frac{e^{2\pi i (p_n - q_n)}}{\lambda} \psi_{N,u}(j_n - i_n) \right] \tag{42}
\]

where \( \pi^{(n)} \) is equal to \( n \) if \( x \) is odd, and to \( x^n \) if \( n \) is even. Inequality (43) comes from the triangular inequality together with (37) and with the fact that the expectation in the righthand side of (42) is bounded by a constant \( C \) (just use Jensen and Cauchy-Schwarz inequalities along with assumption (10) in order to show this last claim). The sum in (42) is consider w.r.t. to all 8-uplet \((i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4) \in V\) where \( V \) denotes the set of values of \( i_1, \ldots, i_4, j_1, j_2, j_3, j_4 \) such that \( i \) for each \( n = 1, \ldots, 4 \), \( j_n \in E_{i_n}^1 \) and such that \( ii) \) each value in the 8-uplet appears at least twice. This restriction of the sum is due to the fact that the expectation in (42) is zero as soon as there exist one value in the 8-uplet \((i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4) \) which appears only once. For instance, some terms of the sum (43) correspond to the situation where \( i_1 = i_2, i_3 = i_4, j_1 = j_2, j_3 = j_4 \). The modulus of the corresponding term is given by

\[
\frac{C}{N^4} \sum_{1 \leq i_1 < i_2 < i_3 < i_4} \sum_{1 \leq j_1 < j_2 < j_3 < j_4} \frac{1}{|j_1 - i_1|^2 |j_2 - i_2|^2 |j_3 - i_3|^2 |j_4 - i_4|^2} = \frac{C}{N^4} \left( \sum_{i=0}^{N-1} \sum_{j \in E_{i_1}^1} \frac{1}{|j - i|^2} \right)^2 \leq \frac{C'}{N^2}
\]

where \( C' \) is a constant. Other terms can be treated similarly. After some algebra, we obtain that \( E[|X_{N,k,k}^{b,1}|^4] \) can be written as \( \frac{1}{N^4} \) times (less than) \( 3N^2 \) bounded terms. Thus \( E[|X_{N,k,k}^{b,2}|^4] \) is bounded for each \( k \). This completes the proof of (39).

The case \( k \neq l \) can be treated using similar arguments. Due to the fact that \( \psi_{N,u} \) is an \( N \)-periodic function, we may assume without restriction that \( N\Delta f_{lk} \) in (36) verifies \( -\frac{N}{2} \leq N\Delta f_{lk} \leq \frac{N}{2} \). We now put \( \xi_{N,k,l}^{(1)}(p,q) = \xi_{N,k,l}^{(1)} + \xi_{N,k,l}^{(2)} + \xi_{N,k,l}^{(3)} \), where \( \xi_{N,k,l}^{(1)} \) correspond to the righthand side of (36) except that the inner sum w.r.t. \( j \) is restricted to \( j \in F_j \), where for each \( i = 0, \ldots, N-1, F_i = \{ j = 0, \ldots, N-1/|j - i + N\Delta f_{lk}| \leq \frac{N}{2} \} \). Similarly, \( \xi_{N,k,l}^{(2)} \) and \( \xi_{N,k,l}^{(3)} \), respectively, correspond to an inner sum w.r.t. indices \( j \) verifying \( \frac{N}{2} < |j - i + N\Delta f_{lk}| \leq N - 1 + \frac{N}{2} \) and \( -N + 1 - \frac{N}{2} < |j - i + N\Delta f_{lk}| < -\frac{N}{2} \). In order to prove that \( \xi_{N,k,l}^{(1)} \rightarrow 0 \), it can be shown as in (43), by expanding \( E[|\zeta_{N,k,l}^{(1)}|^4] \) and using (37), that

\[
E[|\zeta_{N,k,l}^{(1)}|^4] \leq \frac{C}{N^4} \sum_{i_1,i_2,i_3,i_4} \sum_{j_1,\ldots,j_4} \prod_{n=1}^{4} \mathcal{M}(|j_n - i_n + N\Delta f_{lk}|) \tag{44}
\]
where the outer sum is restricted to indices \( i_1, \ldots, i_4 \) such that each value in \((i_1 \ldots i_4)\) appears at least twice (typically, \( i_1 = i_2, i_3 = i_4 \)), and where the inner sum is the restriction of \( F_{i_1} \times \ldots \times F_{i_4} \) such that, again, each value of \((j_1, \ldots, j_4)\) appears at least twice (typically, \( j_1 = j_2, j_3 = j_4 \)). Function \( \mathcal{M}(|x|) \) is defined as \( \frac{1}{|x|} \) for \( x \neq 0 \) and \( \mathcal{M}(0) = 1 \). By studying each combinations of such \( i_1, \ldots, i_4, j_1, \ldots, j_4 \), it can be shown as above that \( E[|\zeta_{N,k,l}^{(1)}|^4] \leq \frac{C}{N^6} \). This proves that \( \zeta_{N,k,l}^{(1)} \xrightarrow{a.s.} 0 \). Terms \( \zeta_{N,k,l}^{(2)} \) and \( \zeta_{N,k,l}^{(3)} \) can be treated using the same approach.

**APPENDIX II**

**PROOF OF THEOREM 2**

We study the behavior of the righthand side of (28) as the length \( L \) of the channel increases. It is worth keeping in mind that (28) is the limit of the normalized CRB associated with the desired channel coefficients when the number \( N \) of subcarriers tends to infinity. Here, we further study the case where \( L \) (in addition to \( N \)) tends to infinity. In this paragraph, the word asymptotic thus refers to the case when \( L \) tends to infinity. The main task is the study of the asymptotic behavior of the trace of matrix \( R_k^{-1}(I_{N_T} \otimes T_k) \). For this, we make use of classical results on the behavior of large Toeplitz matrices [21] [23] [24]. More precisely, the proof requires the use of results on large block-Toeplitz matrices, which are direct generalizations of [21] (see [25] and references therein). Note however that the proof of Theorem 2 requires slightly more general results than those of [25].

Firstly, we study separately the asymptotic behaviors of \( R_k \) and \( T_k \). Secondly, we deduce from this the asymptotic behavior of \( \text{tr}(R_k^{-1}(I_{N_T} \otimes T_k)) \). Denote by \( P_k(f) \) the density of complex matrix-valued measure \( \mu_k \) w.r.t. the Lebesgue measure. We denote by \( P_k^{(t,u)}(f) \) the element of the \( t \)th row and the \( u \)th column of \( P_k(f) \). We assume that each component \( P_k^{(t,u)}(f) \) is a bounded function on \([0,1]\). For each \( t, u = 1, \ldots, N_T \), we denote by \( R_k^{(t,u)} \) the \( L \times L \) block of coordinates \((t,u)\) of matrix \( R_k \), i.e.,

\[
R_k = \begin{pmatrix} R_k^{(t,u)} \end{pmatrix}_{t,u=1,\ldots,N_T}.
\]

We denote by \( F \) the \( L \times L \) Fourier matrix \( F = \left( \frac{1}{\sqrt{L}} e^{-2i\pi \frac{t \cdot j}{L}} \right)_{i,j=0,\ldots,L-1} \). Classical results on the asymptotic behavior of large Toeplitz matrices [21] imply that matrix \( R_k^{(t,u)} \) is asymptotically equivalent to matrix \( \Theta_k^{(t,u)} \) defined by

\[
\Theta_k^{(t,u)} = F^H \Pi_k^{(t,u)} F,
\]  

(45)

where \( \Pi_k^{(t,u)} = \text{diag} \left( P_k^{(t,u)}(0), P_k^{(t,u)}(1), \ldots, P_k^{(t,u)}(L-1) \right) \). By asymptotically equivalent, we mean that spectral norms of \( R_k^{(t,u)} \) and \( \Theta_k^{(t,u)} \) are both bounded as \( L \rightarrow \infty \) and that the normalized Frobenius norm \( |R_k^{(t,u)} - \Theta_k^{(t,u)}| \) of \( R_k^{(t,u)} - \Theta_k^{(t,u)} \) tends to zero as \( L \) tends to infinity [24]. We recall that the normalized Frobenius norm \(|X|\) of a given \( n \times n \) matrix \( X = (x_{ij})_{i,j=1,\ldots,n} \) verifies \(|X|^2 = \frac{1}{n} \sum_i \sum_j |x_{ij}|^2 \).
From the above claim, it is straightforward to show that \( \mathbf{R}_k \) is asymptotically equivalent to \( \Theta_k = \left( \Theta_k^{(t,u)} \right)_{t,u=1,\ldots,N_T} \) as \( L \to \infty \). Based on (45), it is useful to note that \( \Theta_k \) can be written as

\[
\Theta_k = (\mathbf{I}_{N_T} \otimes \mathbf{F}^H) \Pi_k (\mathbf{I}_{N_T} \otimes \mathbf{F}),
\]

where \( \Pi_k = \left( \Pi_k^{(t,u)} \right)_{t,u=1,\ldots,N_T} \). Each block \( \Pi_k^{(t,u)} \) of \( \Pi_k \) is a diagonal matrix. “Renumbering” the elements of \( \Pi_k \) as in [25], it is straightforward to show that \( \Pi_k \) is equivalent to a block diagonal matrix up to a permutation of its rows and columns:

\[
\Pi_k = \mathbf{Q}^T \text{diag} \left( \mathbf{P}_k(0), \ldots, \mathbf{P}_k\left(\frac{L-1}{L}\right) \right) \mathbf{Q},
\]

where \( \mathbf{Q} \) is a permutation matrix. Since our aim is to study the asymptotic behavior of \( \mathbf{R}_k^{-1}(\mathbf{I}_{N_T} \otimes \mathbf{T}_k) \), the most natural approach would consist in studying the asymptotic behavior of \( \mathbf{R}_k^{-1}, \mathbf{T}_k \) and then to deduce the asymptotic behavior of the whole desired matrix. Unfortunately, in certain contexts, matrix \( \mathbf{R}_k \) may be ill-conditioned as \( L \) tends to infinity. A typical example is the case where user \( k \) does not modulate any subcarrier inside a whole frequency interval. In this case, function \( f \to \mathbf{P}_k(f) \) is zero inside this interval. In such cases, it can be shown using the same kind of arguments than in [23] that matrix \( \mathbf{R}_k^{-1} \) becomes singular as \( L \to \infty \). In order to overcome this problem, we consider an arbitrarily small real number \( \epsilon > 0 \). Following the approach of [14], we define the “pseudo-inverse” function as the function \( F_\epsilon \) defined for any non-negative real number \( x \) by \( F_\epsilon(x) = \frac{1}{x} \) if \( x > \epsilon \), \( F_\epsilon(x) = 0 \) otherwise.

For any non-negative Hermitian matrix \( \mathbf{X} = \mathbf{V}^H \text{diag}(\lambda_1, \ldots, \lambda_n) \mathbf{V} \) where \( \mathbf{V} \) is unitary, we generalize the above pseudo-inverse function by

\[
F_\epsilon(\mathbf{X}) = \mathbf{V}^H \text{diag}(F_\epsilon(\lambda_1), \ldots, F_\epsilon(\lambda_n)) \mathbf{V}.
\]

In the sequel, instead of studying \( \mathbf{R}_k^{-1}(\mathbf{I}_{N_T} \otimes \mathbf{T}_k) \) for large \( L \), we rather study matrix \( F_\epsilon(\mathbf{R}_k)(\mathbf{I}_{N_T} \otimes \mathbf{T}_k) \). Following [23], it is straightforward to show that \( F_\epsilon(\mathbf{R}_k) \) is asymptotically equivalent to \( F_\epsilon(\Theta_k) \) as \( L \to \infty \). Using (46) and (47), we obtain

\[
F_\epsilon(\Theta_k) = (\mathbf{I}_{N_T} \otimes \mathbf{F}^H) \mathbf{Q}^T \text{diag} \left( F_\epsilon(\mathbf{P}_k(0)), \ldots, F_\epsilon(\mathbf{P}_k(\frac{L-1}{L})) \right) \mathbf{Q} (\mathbf{I}_{N_T} \otimes \mathbf{F}).
\]

We now study the asymptotic behavior of matrix \( \mathbf{T}_k = \frac{1}{\Delta_k} \int_{D_k} e(f) e(f)^H df \). Matrix \( \mathbf{T}_k \) is an \( L \times L \) Toeplitz matrix. Using again [21] along with the fact that function \( \mathcal{I}_{D_k}(f) \) is bounded, we immediately obtain that \( \mathbf{T}_k \) is asymptotically equivalent to matrix \( \mathbf{F}^H \mathbf{Y} \mathbf{F} \) where \( \mathbf{Y} = \frac{1}{\Delta_k} \text{diag} (\mathcal{I}_{D_k}(0), \ldots, \mathcal{I}_{D_k}(\frac{L-1}{L})) \). As a consequence, \( \mathbf{I}_{N_T} \otimes \mathbf{T}_k \) is asymptotically equivalent to

\[
\text{diag} (\mathbf{F}^H \mathbf{Y} \mathbf{F}, \ldots, \mathbf{F}^H \mathbf{Y} \mathbf{F}) = (\mathbf{I}_{N_T} \otimes \mathbf{F}^H) (\mathbf{I}_{N_T} \otimes \mathbf{Y}) (\mathbf{I}_{N_T} \otimes \mathbf{F}).
\]

Permuting rows and columns of \( \mathbf{Y} \) in the same way than for \( \Pi_k \), we obtain \( \mathbf{I}_{N_T} \otimes \mathbf{Y} = \mathbf{Q}^T (\mathbf{Y} \otimes \mathbf{I}_{N_T}) \mathbf{Q} \), where \( \mathbf{Q} \) is the same permutation matrix than in (48). Plugging this equality into (49) and putting (48)
and (49) together, we conclude that matrix $F_r(\Theta_k) \left( I_{N_r} \otimes (F^H \Psi F) \right)$ coincides with the following matrix

$$
\Psi = \frac{1}{\Delta_k} \left( I_{N_r} \otimes F^H \right) Q^T \text{diag} \left( J_{\Delta_k}(0) F_r(\Psi_{k}(0)), \ldots, J_{\Delta_k}(L-1) F_r(\Psi_{k}(L-1)) \right) Q \left( I_{N_r} \otimes F \right).
$$

Consequently, $F_r(R_k)(I_{N_r} \otimes T_k)$ is asymptotically equivalent to the above matrix. Classical results \[21\] finally yield

\[
\lim_{L \to \infty} \frac{1}{L} \text{tr} \left( F_r(R_k)(I_{N_r} \otimes T_k) \right) = \lim_{L \to \infty} \frac{1}{L} \text{tr} \left( \Psi \right) = \frac{1}{\Delta_k L} \sum_{j=0}^{L-1} J_{\Delta_k}(\frac{j}{L}) \text{tr} \left( F_r(\Psi_{k}(\frac{j}{L})) \right) = \frac{1}{\Delta_k} \int_{D_k} \text{tr} \left( F_r(\Psi_{k}(f)) \right) df = \frac{1}{\Delta_k} \sum_{t=1}^{N_r} \int_{D_k} F_r(\lambda_{k}^{(t)}(f)) df,
\]

where $\lambda_{k}^{(1)}(f) < \cdots < \lambda_{k}^{(N_r)}(f)$ are the eigenvalues of $\Psi_{k}(f)$. We now use the fact that for a given (finite) value of $L$, $R_k$ and $T_k$ are positive definite Hermitian matrices. This implies that $\text{tr} \left( R_k^{-1}(I_{N_r} \otimes T_k) \right) \geq \text{tr} \left( F_r(R_k)(I_{N_r} \otimes T_k) \right)$. Using (28) and (50), we finally obtain

\[
\liminf_{L \to \infty} \frac{1}{L} \left( \liminf_{N \to \infty} \frac{N}{N_k} E \left[ \| g_{N,k} - g_{N,k} \|^2 \right] \right) \geq \frac{\sigma^2 N R}{\Delta_k} \sum_{t=1}^{N_r} \int_{D_k} F_r(\lambda_{k}^{(t)}(f)) df.
\]

The above equation holds for arbitrarily small values of $\epsilon$. Due to the assumptions of Theorem 2, $\int_{D_k} \frac{1}{\lambda^{(1)}(f)} df$ exists. Therefore, the righthand side of (50) has a limit when $\epsilon \to 0$. This limit coincides with $\frac{\sigma^2 N R}{\Delta_k} \sum_{t=1}^{N_r} \int_{D_k} \frac{1}{\lambda_{k}^{(t)}(f)} df$. This proves Theorem 2.

**APPENDIX III**

**PROOF OF PROPOSITION 1**

As $as\text{CRB}_{\omega_k} = \frac{6\sigma^2}{\gamma_k}$, the minimization of $as\text{CRB}_{\omega_k}$ is equivalent to the maximization of $\gamma_k$. Using the definition of the Lebesgue integral \[19\], coefficient $\gamma_k$ defined by (16) also coincides with

\[
\gamma_k = \sup_{(A_i)} \sum_i \inf_{f \in A_i} \left[ \sum_{r=1}^{N_k} h_k^{(r)}(f) H \mu_k(A_i) h_k^{(r)}(f) \right],
\]

where the supremum is taken w.r.t. all decompositions $(A_i)_i$ of interval $[0,1]$. Let $(A_i)_i$ be such a decomposition. We first note that $\mu_k(A_i)$ is a non-negative Hermitian matrix. This is due to the fact that $\mu_k(A_i)$ is the limit of a sequence $\mu_{N,k}(A_i)$ of non-negative Hermitian matrices. Based on this remark, we now make use of the following lemma.
Lemma 3: Denote by \((x_r)_{r=1,...,N_r}\) a sequence of \(N_R\) complex column vectors of size \(N_T \times 1\). Denote by \(\lambda_{\text{max}}\) the largest eigenvalue of matrix \(\sum_{r=1}^{N_R} x_r x_r^H\) and by \(\nu\) the corresponding eigenvector. For each \(N_T \times N_T\) non-negative Hermitian matrix \(M\),

\[
\sum_{r=1}^{N_R} x_r^H M x_r \leq \lambda_{\text{max}} \text{tr}(M),
\]

with equality if and only if (iff) \(M\) has the form \(\beta \nu \nu^H\) where \(\beta\) is any non-negative real number. The proof of the above lemma is omitted due to the lack of space. For each integer \(i\), we now use Lemma 3 with \(M = \mu_k(A_i)\). For each \(f \in [0, 1]\), define \(\lambda_{k,\text{max}}(f)\) as the maximum eigenvalue of \[
\left[ \sum_{r=1}^{N_R} h_k^{(r)}(f) h_k^{(r)}(f) \right] = \max_{\mu_k(A_i)} \left( \text{tr}(\mu_k(A_i)) \right)
\]

Then, using Lemma 3, we have

\[
\sum_i \inf_{\{f \in A_i\}} \left[ \sum_{r=1}^{N_R} h_k^{(r)}(f) \mu_k(A_i) h_k^{(r)}(f) \right] \leq \sum_i \inf_{\{f \in A_i\}} \left[ \text{tr}(\mu_k(A_i)) \lambda_{k,\text{max}}(f) \right]
\]

Equation (54) holds with equality iff for each \(A_i\) and for each \(f\), \(\mu_k(A_i)\) has the form \(\mu_k(A_i) = \beta(A_i) \nu_k(f) \nu_k(f)^H\), where \(\nu_k(f)\) represents the eigenvector associated with \(\lambda_{k,\text{max}}(f)\). As a consequence, equation (55) holds with equality iff i) decomposition \(A_i\) contains the singleton \(\{f_k^{(opt)}\}\), ii) \(\mu_k(f_k^{(opt)}) = \beta(f_k^{(opt)}) \nu_k^{(opt)} \nu_k^{(opt)^H}\) where \(\nu_k^{(opt)} = \nu_k(f_k^{(opt)})\), and iii) \(\mu_k(A_i) = 0\) if \(f_k^{(opt)} \notin A_i\).

Considering the supremum of the lefthand side of (54), we obtain that \(\gamma_k \leq \text{tr}(\mu_k([0, 1])) \lambda_{k,\text{max}}(f_k^{(opt)})\) with equality iff \(\mu_k\) has the form \(\mu_k = \beta(f_k^{(opt)}) \nu_k^{(opt)} \nu_k^{(opt)^H} \delta_{f_k^{(opt)}}\). Finally, introducing the power constraint (32), \(\gamma_k \leq \mathcal{P}_k \lambda_{k,\text{max}}(f_k^{(opt)})\) with equality iff \(\mu_k = \mathcal{P}_k \nu_k^{(opt)} \nu_k^{(opt)^H} \delta_{f_k^{(opt)}}\).

References


Fig. 1. Exact and asymptotic CRB as a function of $N - \frac{P_1}{N_0} = 20\text{dB}$, $N_T = N_R = 2$. 

Fig. 2. MSE of the ML estimator of $\omega_1$ and CRB as a function of $\frac{P_1}{N_0}$, $N = 256$, $K = 2$. 

(a) $N_T = 1$, $N_R = 2$  
(b) $N_T = 2$, $N_R = 2$