# Asymptotic Cramér-Rao Bounds and Training Design for Uplink MIMO-OFDMA Systems 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 proved that the asymptotic performance bounds depend on the choice of the training sequence only via the asymptotic covariance profiles. Moreover, it is shown that the asymptotic performance bounds do not depend on the number of users and the values of the frequency offsets. 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, tel: +33 (0)1 69 85 14 55, fax: +33 (0)1 69 85 14 69).

The material in this paper was 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 subcarrier assignment scheme (SAS). The signal transmitted by a given user is impaired by a frequency selective channel and by a frequency offset. Prior estimation of channel coefficients and frequency offsets has a considerable impact on further users detection steps. The computation of accurate estimates is thus a crucial issue in OFDMA systems. Estimation of frequency offsets and channel coefficients for (single-user) OFDM systems has been investigated in a large number of works (see, e.g., [1]–[2] and references therein). However, the case of OFDMA uplink usually requires more involved synchronization and channel estimation methods. In [3]–[5], estimators which are designed for a specific SAS are proposed. Recent works [6] and [7] investigate the data-aided estimation of frequency offsets in OFDMA uplink for general SASs.

In this paper, 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. We investigate the performance of the data-aided estimation of channels and frequency offsets at the base station with perfect knowledge of the training symbols. 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. Once the expression of the CRB has been obtained under a compact form, it is natural to exploit this result following the general idea of [8], [9] and [10]. These authors proposed 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.

We first derive the exact expression of the CRB (Section III). 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. Note that the idea of studying the CRB in the asymptotic regime in order to simplify its expression has been used previously by [8] and [9] in the case of a single-user single-antenna single-carrier transmission and later by [11] for the issue of channel and clock-offset estimation in OFDM systems. However, a performance study of channel and frequency offsets estimation in the OFDMA case has not been previously undertaken. One of the main difficulty is to provide a general expression of the asymptotic CRB which is valid *for any* OFDMA system: our results should not be specific to a particular resource allocation strategy. Hence, the asymptotic analysis of the CRB in the multi-user OFDMA scenario is more involved than the classical single-user case investigated in [8]–[10], and requires the introduction of novel tools. We emphasize the characteristics of the resource allocation strategy which influence the asymptotic performance. 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). We also remark that orthogonal and non-orthogonal SASs both lead to the same performance: it is thus not essential for a given user $k$ to transmit pilot symbols over all available subcarriers in order to obtain the most accurate estimates. In the present paper, we furthermore investigate the case of a MIMO transmission. Unlike [9], we show that the estimation performance crucially depends on the particular beamformer used by each transmitter. We also prove that the asymptotic CRB associated with the parameters of the $k$th user is identical to the asymptotic CRB that one would have obtained in the absence of other users $l \neq k$. Intuitively, there are therefore some close links between the performance in the OFDMA context and the traditional single-user single-carrier case of [9].

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 [8][10][12], 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. Although the point is briefly discussed for the sake of completeness, the proposition of a general procedure for determining such tradeoffs is beyond the scope of this paper. Finally, simulation results of Section V sustain all 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 $s^{(t)}_{N,k}(0), \ldots, s^{(t)}_{N,k}(N-1)$. 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 $(s^{(t)}_{N,k}(j))_j$ 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 $s^{(t)}_{N,k}(j) = 0$ 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 $(s^{(t)}_{N,k}(j))_j$ and $(s^{(t')}_{N,k}(j))_j$ sent at different antennas $t$ and $t'$ are possibly different. For a given user $k$ and a given antenna $t$, we denote by $(a^{(t)}_{N,k}(n))_n$ the inverse discrete Fourier transform of sequence $(s^{(t)}_{N,k}(j))_j$:

$$a^{(t)}_{N,k}(n) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} s^{(t)}_{N,k}(j) e^{2\imath\pi \frac{nj}{N}} \tag{1}$$

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

$$y^{(r)}_N(n) = \sum_{k=1}^{K} e^{\imath\omega_k n} \sum_{t=1}^{N_T} \sum_{l=0}^{L-1} h^{(t,r)}_k(l) a^{(t)}_{N,k}(n-l) + v^{(r)}(n). \tag{2}$$

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 sampling period. Parameter $h^{(t,r)}_k(l)$ 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 as in [7]: 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 [13]. In the present paper, we consider the classical case (see, e.g., [1] and references therein) where $\omega_k$ is constant w.r.t. antenna pairs $(t, r)$. 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 $\mathbf{h}_k^{(t,r)} = [h_k^{(t,r)}(0), \ldots, h_k^{(t,r)}(L-1)]^T$, where $(\cdot)^T$ represents the transpose operator. Stacking all $N$ samples $y_N^{(r)}(n)$ received by antenna $r$ into one column vector $\mathbf{y}_N^{(r)} = [y_N^{(r)}(0), \ldots, y_N^{(r)}(N-1)]^T$, one obtains:

$$\mathbf{y}_N^{(r)} = \sum_{k=1}^{K} \mathbf{\Gamma}_N(\omega_k) \left( \sum_{t=1}^{N_T} \mathbf{A}_{N,k}^{(t)} \mathbf{h}_k^{(t,r)} \right) + \mathbf{v}_N^{(r)},$$

where $\mathbf{v}_N^{(r)} = [v^{(r)}(0), \ldots, v^{(r)}(N-1)]^T$ and $\mathbf{\Gamma}_N(\omega_k) = \mathrm{diag}(1, e^{\imath\omega_k}, \ldots, e^{\imath\omega_k(N-1)})$. In the above expression, each matrix $\mathbf{A}_{N,k}^{(t)}$ is an $N \times L$ matrix containing the time-domain training sequence sent at the $t$th transmit antenna of user $k$. More precisely,

$$\mathbf{A}_{N,k}^{(t)} = \left( a_{N,k}^{(t)}(i-j) \right)_{\substack{0 \le i \le N-1 \\ 0 \le j \le L-1}}.$$

In particular, $\mathbf{A}_{N,k}^{(t)}$ 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 $NN_R \times 1$ vector $\mathbf{y}_N = [\mathbf{y}_N^{(1)T}, \ldots, \mathbf{y}_N^{(N_R)T}]^T$ given by

$$\mathbf{y}_N = \sum_{k=1}^{K} \left[ \mathbf{I}_{N_R} \otimes (\mathbf{\Gamma}_N(\omega_k) \mathbf{A}_{N,k}) \right] \mathbf{h}_k + \mathbf{v}_N, \qquad (3)$$

where $\mathbf{I}_{N_R}$ denotes the $N_R \times N_R$ identity matrix, $\otimes$ stands for the Kronecker product, $\mathbf{A}_{N,k} = [\mathbf{A}_{N,k}^{(1)}, \ldots, \mathbf{A}_{N,k}^{(N_T)}]$, and $\mathbf{v}_N = [\mathbf{v}_N^{(1)T}, \ldots, \mathbf{v}_N^{(N_R)T}]^T$ is an additive noise vector with independent complex circular Gaussian random entries of variance $\sigma^2$. Here, vector $\mathbf{h}_k = [\mathbf{h}_k^{(1,1)T}, \ldots, \mathbf{h}_k^{(N_T,1)T}, \ldots, \mathbf{h}_k^{(1,N_R)T}, \ldots, \mathbf{h}_k^{(N_T,N_R)T}]^T$ 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 $\boldsymbol{\theta} = [\omega_1, \mathbf{h}_1^T, \ldots, \omega_K, \mathbf{h}_K^T]^T$ the deterministic parameter vector to be estimated. In the following section, we provide bounds on the performance of estimates of $\boldsymbol{\theta}$.

*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 $k$. 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 $\mathbf{h}_k$. However, the set of frequency taps which should be estimated in this case can be written as a simple function of $\mathbf{h}_k$. For this reason, we first investigate the performance of estimates of initial parameter vector $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$. Such an analysis provides performance bounds for estimates of $\boldsymbol{\theta}$. 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 $\boldsymbol{\theta}$. 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 $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}_1^T, \ldots, \tilde{\boldsymbol{\theta}}_K^T]^T$ where for each $k$, $\tilde{\boldsymbol{\theta}}_k = [\omega_k, \mathbf{h}_{R,k}^T, \mathbf{h}_{I,k}^T]^T$ denotes the parameter vector corresponding to a given user $k$. In the above definition, $\mathbf{h}_{R,k}$ and $\mathbf{h}_{I,k}$ respectively represent the real and the imaginary parts of vector $\mathbf{h}_k$. The exact CRB for $\tilde{\boldsymbol{\theta}}$ is classically defined as the inverse of the Fisher Information Matrix (FIM) $\mathbf{J}_N$. Note that an expression of the FIM for $\tilde{\boldsymbol{\theta}}$ has been recently derived by [7] in the SISO case. Using an approach similar to [10], the FIM for parameter $\tilde{\boldsymbol{\theta}}$ can be obtained as the following $K(1+2LN_RN_T) \times K(1+2LN_RN_T)$ matrix:

$$\mathbf{J}_N = \frac{2}{\sigma^2}\Re\left[\left(\boldsymbol{\nabla}_{\tilde{\boldsymbol{\theta}}}\boldsymbol{\eta}_N^H\right)\left(\boldsymbol{\nabla}_{\tilde{\boldsymbol{\theta}}}\boldsymbol{\eta}_N^H\right)^H\right], \tag{4}$$

where $\boldsymbol{\eta}_N = \sum_{k=1}^K \left[\mathbf{I}_{N_R} \otimes (\boldsymbol{\Gamma}_N(\omega_k)\mathbf{A}_{N,k})\right]\mathbf{h}_k$. 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 $\mathbf{x} = [x_1, \ldots, x_n]^T$ and for a given $1 \times m$ row vector $\mathbf{z}(\mathbf{x}) = [z_1(\mathbf{x}), \ldots, z_m(\mathbf{x})]$ function of $\mathbf{x}$, matrix $\boldsymbol{\nabla}_{\mathbf{x}}\mathbf{z}$ is defined as the $n \times m$ matrix $\left(\frac{\partial z_j(\mathbf{x})}{\partial x_i}\right)_{\substack{1\le i\le n\\1\le j\le m}}$. After some algebra, we obtain the following "block" representation of matrix $\mathbf{J}_N$:

$$\mathbf{J}_N = \begin{bmatrix} \mathbf{J}_{N,1,1} & \cdots & \mathbf{J}_{N,1,K} \\ \vdots & \ddots & \vdots \\ \mathbf{J}_{N,K,1} & \cdots & \mathbf{J}_{N,K,K} \end{bmatrix}. \tag{5}$$

For each $k, l = 1, \ldots, K$, $\mathbf{J}_{N,k,l}$ is the $(1+2LN_RN_T) \times (1+2LN_RN_T)$ matrix given by

$$\mathbf{J}_{N,k,l} = \frac{2}{\sigma^2}\begin{bmatrix} \alpha_{N,k,l} & \Im\left[\boldsymbol{\beta}_{N,l,k}^H\right] & \Re\left[\boldsymbol{\beta}_{N,l,k}^H\right] \\ -\Im\left[\boldsymbol{\beta}_{N,k,l}\right] & \Re\left[\mathbf{U}_{N,k,l}\right] & -\Im\left[\mathbf{U}_{N,k,l}\right] \\ \Re\left[\boldsymbol{\beta}_{N,k,l}\right] & \Im\left[\mathbf{U}_{N,k,l}\right] & \Re\left[\mathbf{U}_{N,k,l}\right] \end{bmatrix}, \tag{6}$$

where

$$\mathbf{U}_{N,k,l} = \mathbf{I}_{N_R} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}\right), \quad (7)$$

$$\boldsymbol{\beta}_{N,k,l} = \left[\mathbf{I}_{N_R} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{D}_N \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}\right)\right] \mathbf{h}_l, \quad (8)$$

$$\alpha_{N,k,l} = \Re\left[\mathbf{h}_k^H \left[\mathbf{I}_{N_R} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{D}_N^2 \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}\right)\right] \mathbf{h}_l\right], \quad (9)$$

and where $\mathbf{D}_N = \text{diag}\,(0, 1, \dots, N-1)$. 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 $\tilde{\boldsymbol{\theta}}$. 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 the total number of unknown parameters, *i.e.*, $N \gg K(LN_RN_T + 1)$. 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 our asymptotic analysis, we make the following assumptions:

*i)* 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 SAS, a certain number of subcarriers may not be modulated by user $k$. If $j$ is one of these subcarriers, we simply consider that $E[|s_{N,k}^{(t)}(j)|^2] = 0$.

*ii)* Furthermore, we assume that training sequences $(s_{N,k}^{(t)}(j))_j$ and $(s_{N,k}^{(t')}(j))_j$ 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 $E[s_{N,k}^{(t)}(j)s_{N,k}^{(t')}(j)^*]$ may be nonzero.

*iii)* We assume that eight-order moments of random variables $(s_{N,k}^{(t)}(j))_j$ are uniformly bounded, *i.e.*

$$\sup_N \max_j E\left[\left|s_{N,k}^{(t)}(j)\right|^8\right] < M \quad (10)$$

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

*iv)* 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 $\mathbf{CRB}_N = \mathbf{J}_N^{-1}$. Note that for finite values of $L$, it is reasonable to assume that the FIM is non-singular. Due to definitions of $\mathbf{U}_{N,k,l}$, $\boldsymbol{\beta}_{N,k,l}$ 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 $\mathbf{A}_{N,k}^H \mathbf{D}_N^u \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}$ for each $u = 0, 1, 2$ and for each $k, l = 1, \dots, K$. The following lemma provides a simpler expression of the latter matrices as $N$ increases.

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

$$\frac{u+1}{N^{u+1}} \mathbf{A}_{N,k}^H \mathbf{D}_N^u \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l} - \frac{\delta(k-l)}{N} \sum_{j=0}^{N-1} E\left[\mathbf{s}_{N,k}(j)\mathbf{s}_{N,k}(j)^H\right]^* \otimes \left[\mathbf{e}(\frac{j}{N})\mathbf{e}(\frac{j}{N})^H\right] \xrightarrow{N} 0 \quad \text{a.s.} \tag{11}$$

where notation $\xrightarrow{N} 0$ 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 $\mathbf{s}_{N,k}(j) = [s_{N,k}^{(1)}(j), \dots, s_{N,k}^{(N_T)}(j)]^T$ 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 $\mathbf{J}_{N,k,l}$ for $k \neq l$ of the FIM $\mathbf{J}_N$ become significantly smaller than the corresponding elements of diagonal blocks $\mathbf{J}_{N,k,k}$ as $N$ increases. In order to characterize the asymptotic behavior of the desired matrices $\frac{u+1}{N^{u+1}} \mathbf{A}_{N,k}^H \mathbf{D}_N^u \mathbf{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}$, it is convenient to rewrite the second term of (11) using

$$\frac{1}{N} \sum_{j=0}^{N-1} E\left[\mathbf{s}_{N,k}(j)\mathbf{s}_{N,k}(j)^H\right]^* \otimes \left[\mathbf{e}(\frac{j}{N})\mathbf{e}(\frac{j}{N})^H\right] = \int_0^1 \boldsymbol{\mu}_{N,k}(df) \otimes \left[\mathbf{e}(f)\mathbf{e}(f)^H\right], \tag{12}$$

where $\boldsymbol{\mu}_{N,k}$ denotes the following matrix-valued measure [16] defined for any Borel set $A$ of $[0, 1]$ by

$$\boldsymbol{\mu}_{N,k}(A) = \frac{1}{N} \sum_{j=0}^{N-1} E\left[\mathbf{s}_{N,k}(j)\mathbf{s}_{N,k}(j)^H\right]^* \mathcal{I}_A(\frac{j}{N}), \tag{13}$$

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 $\mu_{N,k}^{(1,1)}(A) = \frac{1}{N}\sum_{j=0}^{N-1} E\left[|s_{N,k}^{(1)}(j)|^2\right] \mathcal{I}_A(\frac{j}{N})$. 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 $\boldsymbol{\mu}_{N,k}$ is in some sense equivalent to the (conjugate) covariance profile of the multi-dimensional training sequence $\mathbf{s}_{N,k}(j)$.

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 $\boldsymbol{\mu}_k$ such that $\boldsymbol{\mu}_{N,k}$ converges weakly to $\boldsymbol{\mu}_k$ as $N$ tends to infinity. Equivalently, for any continuous function $f \to F(f)$, the integral $\int_0^1 F(f)\boldsymbol{\mu}_{N,k}(df)$ converges to $\int_0^1 F(f)\boldsymbol{\mu}_k(df)$ as $N$ tends to infinity.

In the sequel, we refer to $\boldsymbol{\mu}_k$ 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 SAS and power allocation strategies for OFDMA systems. Moreover, we shall see below that the asymptotic CRB associated with $\tilde{\boldsymbol{\theta}}$ depends on the training strategy only via $\boldsymbol{\mu}_k$. Hence, the asymptotic covariance profile is sufficient to characterize the asymptotic CRB. No further assumptions on the particular SAS, 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 $A \to \boldsymbol{\mu}_k(A)$, focus for instance on its first component $A \to \mu_k^{(1,1)}(A)$ (*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 $N_T \times N_T$ matrix-valued measure $\boldsymbol{\mu}_k$ has a matrix density $\mathbf{P}_k(f)$, (*i.e.*, $\boldsymbol{\mu}_k(df) = \mathbf{P}_k(f)df$), then $\mathbf{P}_k(f)$ can be interpreted as the power density spectrum of the $N_T$-dimensional sequence sent at all $N_T$ antennas of user $k$.

Lemma 1 along with Assumption 1 and equality (12) immediately leads to the following lemma.

*Lemma 2:* For each $k$, denote by $\mathbf{R}_k$ the following $LN_T \times LN_T$ matrix

$$\mathbf{R}_k = \int_0^1 \boldsymbol{\mu}_k(df) \otimes \left[\mathbf{e}(f)\mathbf{e}(f)^H\right]. \tag{14}$$

Then, for each $u = 0, 1, 2$, for each $k, l = 1, \ldots, K$,

$$\frac{u+1}{N^{u+1}} \mathbf{A}_{N,k}^H \mathbf{D}_N^u \boldsymbol{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l} \xrightarrow{N} \delta(k-l)\ \mathbf{R}_k \quad \text{a.s.} \tag{15}$$

Using Lemma 2, one can now easily characterize the asymptotic behavior of each block (6) of the FIM. For example, due to Lemma 2 and (7), the matrix $\mathbf{U}_{N,k,l}$ has the same asymptotic behavior as $N\delta(k-l)\mathbf{I}_{N_R}\otimes\mathbf{R}_k$ as $N$ tends to infinity. Similarly, from (9), the coefficient $\alpha_{N,k,l}$ has the same asymptotic behavior as $\frac{N^3}{3}\delta(k-l)\gamma_k$, where

$$\begin{aligned}\gamma_k &= \mathbf{h}_k^H\left[\mathbf{I}_{N_R}\otimes\mathbf{R}_k\right]\mathbf{h}_k \\ &= \sum_{r=1}^{N_R}\int_0^1 \boldsymbol{h}_k^{(r)}(f)^H\boldsymbol{\mu}_k(df)\boldsymbol{h}_k^{(r)}(f). \end{aligned} \tag{16}$$

Here, $\boldsymbol{h}_k^{(r)}(f) = \sum_{l=0}^{L-1}[h_k^{(1,r)}(l),\ldots,h_k^{(N_T,r)}(l)]^T e^{-2\imath\pi lf}$ may be interpreted as the overall frequency response of the channel "seen" at receive antenna $r$. Following these kind of ideas, it is straightforward to characterize the asymptotic behavior of $\mathbf{J}_N$ and thus of $\mathbf{CRB}_N$. Note that Lemma 2 suggests rather to study the asymptotic behavior of the *normalized* CRB. We define $\overline{\mathbf{CRB}}_N = \mathbf{W}_N\mathbf{CRB}_N\mathbf{W}_N$ where $\mathbf{W}_N$ is the $K(1+2LN_RN_T)\times K(1+2LN_RN_T)$ diagonal matrix defined by $\mathbf{W}_N = \mathrm{diag}(\mathbf{w}_N^T,\ldots,\mathbf{w}_N^T)$ 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 $\overline{\mathbf{CRB}}_N$ converges almost surely to the block-diagonal matrix $\overline{\mathbf{CRB}}$ given by $\overline{\mathbf{CRB}} = \mathrm{diag}\left(\mathbf{C}_1,\ldots,\mathbf{C}_K\right)$. For each $k=1,\ldots,K$, $\mathbf{C}_k$ is the $(1+2LN_RN_T)\times(1+2LN_RN_T)$ matrix equal to

$$\mathbf{C}_k = \frac{\sigma^2}{2}\begin{bmatrix} \frac{12}{\gamma_k} & \frac{6\,\mathbf{h}_{I,k}^T}{\gamma_k} & -\frac{6\,\mathbf{h}_{R,k}^T}{\gamma_k} \\ \frac{6\,\mathbf{h}_{I,k}}{\gamma_k} & \Re\left[\left(\mathbf{I}_{N_R}\otimes\mathbf{R}_k^{-1}\right)\right]+3\frac{\mathbf{h}_{I,k}\mathbf{h}_{I,k}^T}{\gamma_k} & -\Im\left[\left(\mathbf{I}_{N_R}\otimes\mathbf{R}_k^{-1}\right)\right]-3\frac{\mathbf{h}_{I,k}\mathbf{h}_{R,k}^T}{\gamma_k} \\ \frac{-6\,\mathbf{h}_{R,k}}{\gamma_k} & \Im\left[\left(\mathbf{I}_{N_R}\otimes\mathbf{R}_k^{-1}\right)\right]-3\frac{\mathbf{h}_{R,k}\mathbf{h}_{I,k}^T}{\gamma_k} & \Re\left[\left(\mathbf{I}_{N_R}\otimes\mathbf{R}_k^{-1}\right)\right]+3\frac{\mathbf{h}_{R,k}\mathbf{h}_{R,k}^T}{\gamma_k} \end{bmatrix}. \tag{17}$$

*Proof:* Using Lemma 2, it can be shown that the *normalized* FIM $\overline{\mathbf{J}}_N = \mathbf{W}_N^{-1}\mathbf{J}_N\mathbf{W}_N^{-1}$ converges a.s. to a block-diagonal matrix $\overline{\mathbf{J}} = \mathrm{diag}\left(\mathbf{J}_1,\ldots,\mathbf{J}_K\right)$ 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 $\overline{\mathbf{CRB}} = \overline{\mathbf{J}}^{-1} = \mathrm{diag}\left(\mathbf{J}_1^{-1},\ldots,\mathbf{J}_K^{-1}\right)$. 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 [10], 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, $E_N[\,.\,]$ denotes the conditional expectation w.r.t. training sequences, *i.e.*, $E_N[X] = E\left[X / \left(\mathbf{s}_{N,1}(j)\right)_j, \ldots, \left(\mathbf{s}_{N,K}(j)\right)_j\right]$ 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 $\hat{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, Theorem 1 implies that the following inequalities hold with probability one (w.p.1).

$$\liminf_{N\to\infty} N^3 E_N\left[\left(\hat{\omega}_{N,k} - \omega_k\right)^2\right] \geq \frac{6\sigma^2}{\gamma_k} \tag{18}$$

$$\liminf_{N\to\infty} N E_N\left[\left\|\hat{\mathbf{h}}_{N,k} - \mathbf{h}_k\right\|^2\right] \geq N_R \sigma^2 \mathrm{tr}\left(\mathbf{R}_k^{-1}\right) + \frac{3\sigma^2}{2}\frac{\mathbf{h}_k^H \mathbf{h}_k}{\gamma_k} \tag{19}$$

  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, 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 $\frac{1}{N^3}$.
- Theorem 1 states that the CRB matrix converges *almost surely* to a deterministic matrix $\overline{\mathbf{CRB}}$ whereas most previous results in the field of CRB studies and training sequence design only focus on the convergence in probability. The almost sure convergence is a much more powerful result. It means that *for almost any particular realization of the training sequence*, the normalized CRB converges to the deterministic expression given by Theorem 1. It allows furthermore the statement of important inequalities such as (18) and (19).
- It is interesting to remark that the expression of the asymptotic CRB (18) associated with the frequency offset is similar to the bound obtained by [9] in the single-user single-carrier case. In both cases, the asymptotic CRB is proportional to a certain constant $\gamma_k$ which can be interpreted as the *useful* power received by the base station and transmitted by user $k$. In other words, the asymptotic CRB $\frac{6\sigma^2}{\gamma_k}$ is proportional to the inverse of the *signal to noise ratio* associated with a given transmitter.
- For a given user $k$, the bounds on the MSE for estimates of $\boldsymbol{\theta}_k$ do not depend on the values of parameters $\boldsymbol{\theta}_l$ associated with other users $l \neq k$ and 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 obtained in the absence of other users $l \neq k$.

- Asymptotic bounds for the MSE of unbiased estimates of $\boldsymbol{\theta}$ depend on the training scheme only via the asymptotic covariance profiles $\boldsymbol{\mu}_k$. Different training schemes may have identical asymptotic covariance profile $\boldsymbol{\mu}_k$. Hence, different training schemes may lead to similar estimation performance. In order to illustrate this claim, we consider the following two training examples. The first one in a non-orthogonal[1] training strategy which we call $T_1$. In this case, each user $k$ modulates all subcarriers $j = 0, \ldots, N-1$, with equal power $\mathcal{P}_k$. In other words, 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 having the indices $\left\{iK + k - 1 \middle/ i = 0, \ldots, \frac{N}{K} - 1\right\}$ 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 SAS as *interleaved* OFDMA [5]. 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 $\boldsymbol{\mu}_k(A) = \mathcal{P}_k \mathbf{I}_{N_T} \lambda(A)$, 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 SAS both have the same asymptotic performance bounds, as long as they have an identical asymptotic covariance profile $\boldsymbol{\mu}_k$. 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 $\boldsymbol{\theta}_k$.
- The normalized CRB tends to the limit given by Theorem 1 when $N$ tends to infinity, while all other parameters remain constant. This means in particular that $\frac{N}{KL(N_T+N_R)}$ tends to infinity. In practice, for finite values of $N$, it is therefore reasonable to conjecture that the normalized CRB is close to the limit of Theorem 1 provided that $\frac{N}{KL(N_T+N_R)}$ is large enough. Section V provides more details on this point.

## IV. Further Simplifications, Training Sequence Selection

Theorem 1 provides 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 $\boldsymbol{\mu}_k$. Thus, following the approach of previous studies [8][10], it is natural to search for the training strategies which minimize the CRB on the frequency offset and on the channel, the latter

[1]A given subcarrier is possibly modulated by several users

being at least measured in a part of the available frequency band as we shall see below. 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.

It is worth noticing that for a given $k$, The asymptotic CRB only depends on the training sequence of user $k$ and depends neither on the training strategies $\boldsymbol{\mu}_l$ corresponding to other users $l \neq k$ nor on parameters $\boldsymbol{\theta}_l$ associated with other users. This remark is of practical importance. Indeed, it implies that the CRB based selection of the training sequence of a given user $k$ does not require the knowledge of other users' training strategies and parameters.

Of course, one would expect from an ideal training strategy that it simultaneously minimizes bounds on both the frequency offset and the channel. Unfortunately, results of [8][10][12] tend to show that no single training sequence is likely to jointly minimize both bounds. In order to overcome this problem, [10] 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 [17]. This issue is out of the scope of the present paper. Here, we focus on separate minimization of both bounds. We begin with the asymptotic CRB on $\omega_k$, written

$$asCRB_{\omega_k} = \frac{6\sigma^2}{\gamma_k} \tag{20}$$

as inequality (18) shows.

### A. Minimization of $asCRB_{\omega_k}$

We minimize $asCRB_{\omega_k}$ under the constraint that the total power transmitted per time sample by user $k$ does not exceed $\mathcal{P}_k$. The latter power constraint is equivalent to the following inequality:

$$\mathrm{tr}\left(\boldsymbol{\mu}_k([0,1])\right) \leq \mathcal{P}_k. \tag{21}$$

As mentioned in Section III, the above expression of the power constraint is a direct consequence of (13). Inequality (21) can be motivated by noticing that for finite $N$, the total transmit power coincides with $\mathrm{tr}(\boldsymbol{\mu}_{N,k}([0,1]))$. In addition to the power constraint, as in [9] we have to put a condition that ensures that $\mathbf{R}_k$ is a regular matrix for the CRB expression (17) to be meaningful. One such condition is to assume that $\boldsymbol{\mu}_k$ has the form

$$\boldsymbol{\mu}_k(df) = (\epsilon/N_T)\mathbf{I}_{N_T} df + \boldsymbol{\rho}_k(df) \ , \tag{22}$$

where $\epsilon$ is a small number $\epsilon > 0$ and $\boldsymbol{\rho}_k(df)$ is a matrix-valued positive measure on $[0,1]$. With this assumption, $\mathbf{R}_k$ has the form

$$\mathbf{R}_k = \frac{\epsilon}{N_T}\mathbf{I}_{LN_T} + \int_0^1 \boldsymbol{\rho}_k(df) \otimes \left[\mathbf{e}(f)\mathbf{e}(f)^H\right] \ . \tag{23}$$

As the integral in the right hand side is a non negative matrix, $\mathbf{R}_k$ is regular. Training sequences that minimize the bound $asCRB_{\omega_k}$ can be selected according to Proposition 1:

*Proposition 1:* For each $f \in [0,1]$, recall that $\boldsymbol{h}_k^{(r)}(f) = \sum_{l=0}^{L-1}[h_k^{(1,r)}(l), \ldots, h_k^{(N_T,r)}(l)]^T e^{-2\imath\pi l f}$. Denote by $\lambda_{k,max}(f)$ the largest eigenvalue of matrix $\sum_{r=1}^{N_R} \boldsymbol{h}_k^{(r)}(f)\boldsymbol{h}_k^{(r)}(f)^H$. Define $f_k^{(opt)} = \arg\max_f \lambda_{k,max}(f)$. Under power constraint (21), along with (22), $asCRB_{\omega_k}$ is minimum if and only if

$$\boldsymbol{\rho}_k = (\mathcal{P}_k - \epsilon)\left(\boldsymbol{\nu}_k^{(opt)}\boldsymbol{\nu}_k^{(opt)H}\right)\delta_{f_k^{(opt)}}, \tag{24}$$

where $\boldsymbol{\nu}_k^{(opt)}$ is the (unit norm) eigenvector associated with $\lambda_{k,max}(f_k^{(opt)})$ and where $\delta_{f_k^{(opt)}}$ is the Dirac measure at $f_k^{(opt)}$.

The proof of Proposition 1 is provided in Appendix II. 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 $\boldsymbol{\mu}_k$ defined by (24). In other words, Proposition 1 describes the best use of the available power $\mathcal{P}_k$ 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 almost all available power at the frequency for which the largest eigenvalue of $\sum_{r=1}^{N_R} \boldsymbol{h}_k^{(r)}(f)\boldsymbol{h}_k^{(r)}(f)^H$ is maximum. In the single antenna case, *i.e.*, $N_T = 1$, $N_R = 1$, optimal asymptotic covariance profile (24) simply reduces to the scalar measure $(\mathcal{P}_k - \epsilon)\delta_{f_k^{(opt)}}$ where $f_k^{(opt)} = \arg\max_{f\in[0,1]}\left|\sum_{l=0}^{L-1} h_k(l)e^{-2\imath\pi f l}\right|^2$. In this case, Proposition 1 suggests to allocate most of the power at the subcarriers which are close to the optimal frequency $f_k^{(opt)}$. Note that there is of course a close link between the above guideline and the training strategy obtained by [9] in case of a classical single-carrier transmission. Indeed, when the training sequence is transmitted in the time-domain, [9] suggests to choose a (stationary) training sequence whose power spectrum is maximum at frequency $f_k^{(opt)}$.
- Proposition 1 also states that $asCRB_{\omega_k}$ is minimum when the covariance matrix between training sequences sent at different transmit antennas coincides with $\boldsymbol{\nu}_k^{(opt)}\boldsymbol{\nu}_k^{(opt)H}$ up to a multiplicative factor (and to the small additive factor $(\epsilon/N_T)\mathbf{I}_{LN_T}$). In practice, for finite values of $N$ and for a given subcarrier $j$ whose frequency $j/N$ is close to the optimal frequency $f_k^{(opt)}$, this guideline may be followed by defining the multi-dimensional training sequence $\mathbf{s}_{N,k}(j) = [s_{N,k}^{(1)}(j), \ldots, s_{N,k}^{(N_T)}(j)]^T$

as $\mathbf{s}_{N,k}(j) = \boldsymbol{\nu}_k^{(opt)} w_{N,k}(j)$, 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 is non zero only at subcarriers $j$ such that $\frac{j}{N}$ is close to $f_k^{(opt)}$.

- The training strategy suggested by Proposition 1 requires some limited channel knowledge consisting in $f_k^{(opt)}$ and $\boldsymbol{\nu}_k^{(opt)}$. As explained and discussed in [9], 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 $\boldsymbol{\rho}_k$ is chosen as in (24), then one can show by inspecting the proof in Appendix II that $asCRB_{\omega_k}$ increases with $\epsilon$. But for small values of $\epsilon$, matrix $\mathbf{R}_k$ given by (23) becomes nearly singular. In this case, the asymptotic bound on channel estimates given by the righthand side of (19) becomes very large. 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. *Frequency domain estimation of channel parameters*

In the present subsection, we further investigate the performance of estimates of channel coefficients. As mentioned in Section II, one should 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 $\mathbf{h}_k$. 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 [14]. In such a situation, estimation of the channel outside the useful frequency band is of little 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 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

$$g_{N,k}^{(t,r)}(i) = \sum_{l=0}^{L-1} h_k^{(t,r)}(l) e^{-2\imath\pi l j_i / N} \tag{25}$$

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

$$\mathbf{\Phi}_{N,k} = \left(e^{-2\imath\pi l j_i/N}\right)_{\substack{0\leq i\leq N_k-1\\ 0\leq l\leq L-1}} . \tag{26}$$

The desired parameter vector for a given user $k$ is defined by $\boldsymbol{g}_{N,k} = [\boldsymbol{g}_{N,k}^{(1,1)^T}, \ldots, \boldsymbol{g}_{N,k}^{(N_T,1)^T}, \ldots, \boldsymbol{g}_{N,k}^{(1,N_R)^T}, \ldots, \boldsymbol{g}_{N,k}^{(N_T,N_R)^T}]^T$ and obtained as $\boldsymbol{g}_{N,k} = (\mathbf{I}_{N_T N_R} \otimes \mathbf{\Phi}_{N,k})\,\mathbf{h}_k$. Parameter $\boldsymbol{g}_{N,k}$ is thus a linear function of the initial set of parameters. Recall that the CRB associated with any linear function $\mathbf{G}\boldsymbol{\theta}$ (where $\mathbf{G}$ is any matrix) of the initial parameter vector $\boldsymbol{\theta}$ is given by $\mathbf{G}(\mathbf{CRB}_N)\mathbf{G}^T$ (see [15]). Using this result, it is straightforward to show that for any unbiased estimator $\hat{\boldsymbol{g}}_{N,k}$ of $\boldsymbol{g}_{N,k}$, the normalized MSE on channel frequency taps verifies the following inequality:

$$\frac{1}{N_k}E_N\left[\left\|\hat{\boldsymbol{g}}_{N,k} - \boldsymbol{g}_{N,k}\right\|^2\right] \geq \frac{1}{N}\mathrm{tr}\left(\overline{\mathbf{CRB}}_{N,k,k}\begin{bmatrix} 0 & \mathbf{0}_{LN_TN_R}{}^T & \mathbf{0}_{LN_TN_R}{}^T \\ \mathbf{0}_{LN_TN_R} & \mathbf{I}_{N_TN_R}\otimes\Re\left[\mathbf{T}_{N,k}\right] & -\mathbf{I}_{N_TN_R}\otimes\Im\left[\mathbf{T}_{N,k}\right] \\ \mathbf{0}_{LN_TN_R} & \mathbf{I}_{N_TN_R}\otimes\Im\left[\mathbf{T}_{N,k}\right] & \mathbf{I}_{N_TN_R}\otimes\Re\left[\mathbf{T}_{N,k}\right]\end{bmatrix}\right), \tag{27}$$

where $\overline{\mathbf{CRB}}_{N,k,k}$ represents the $k$th $(2LN_TN_R+1)\times(2LN_TN_R+1)$ diagonal block of the normalized CRB matrix associated with initial parameter vector $\tilde{\boldsymbol{\theta}}$, *i.e.*, $\overline{\mathbf{CRB}}_N = \left(\overline{\mathbf{CRB}}_{N,k,l}\right)_{k,l=1,\ldots,K}$, and where $\mathbf{T}_{N,k} = \frac{1}{N_k}\mathbf{\Phi}_{N,k}{}^H\mathbf{\Phi}_{N,k}$. Notation $\mathbf{0}_{LN_TN_R}$ stands for the $LN_TN_R \times 1$ null vector. We now simplify (27). Indeed, the righthand side of (27) only depends on two matrices $\overline{\mathbf{CRB}}_{N,k,k}$ and $\mathbf{T}_{N,k}$. Theorem 1 states that $\overline{\mathbf{CRB}}_{N,k,k}$ converges a.s. toward matrix $\mathbf{C}_k$ given by (17) as $N$ increases. In order to obtain a simple expression of (27), the only task is therefore to study the asymptotic behavior of $\mathbf{T}_{N,k}$. 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 $\mathbf{T}_{N,k}$ can be written as

$$\frac{1}{N_k}\sum_{j\in\Xi_{N,k}} e^{2\imath\pi(p-q)\frac{j}{N}} = \int_0^1 e^{2\imath\pi(p-q)f}\nu_{N,k}(df)\ , \tag{28}$$

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

$$\nu_{N,k}(A) = \frac{1}{N_k}\sum_{j\in\Xi_{N,k}} \mathcal{I}_A(\frac{j}{N})\ . \tag{29}$$

Measure $\nu_{N,k}$ is introduced in order to simplify the asymptotic study of matrix $\mathbf{T}_{N,k}$. 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, $\nu_{N,k}([0,1]) = 1$. 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. Equivalently, for any continuous function $f \to F(f)$, the integral $\int_0^1 F(f)\boldsymbol{\nu}_{N,k}(df)$ converges to $\int_0^1 F(f)\boldsymbol{\nu}_k(df)$ as $N$ tends to infinity.

Based on the above assumption, we can directly write that (28) converges to $\int_0^1 e^{2\imath\pi(p-q)f}\nu_k(df)$ as $N$ tends to infinity. Finally, $\mathbf{T}_{N,k}$ converges to the following $L \times L$ matrix $\mathbf{T}_k$ defined by

$$\mathbf{T}_k = \int_0^1 \mathbf{e}(f)\mathbf{e}(f)^H \nu_k(df), \tag{30}$$

where we recall that $\mathbf{e}(f) = [1, e^{2\imath\pi f}, \ldots, e^{2\imath\pi f(L-1)}]^T$. In the following, 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 other hand, 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 SAS, 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 $\mathcal{D}_k$. In this case, $\mathbf{T}_k$ coincides with $\mathbf{T}_k = \frac{1}{\Delta_k}\int_{\mathcal{D}_k} \mathbf{e}(f)\mathbf{e}(f)^H df$. Based on Theorem 1, it is straightforward to show that (27) reduces to:

$$\liminf_{N\to\infty} \frac{N}{N_k} E_N\left[\left\|\hat{\boldsymbol{g}}_{N,k} - \boldsymbol{g}_{N,k}\right\|^2\right] \geq \sigma^2 N_R \mathrm{tr}\left(\mathbf{R}_k^{-1}\left(\mathbf{I}_{N_T} \otimes \mathbf{T}_k\right)\right) + \frac{3\sigma^2}{2}\frac{\beta_k}{\gamma_k}, \tag{31}$$

where $\beta_k = \frac{1}{\Delta_k}\sum_{r=1}^{N_R}\int_{\mathcal{D}_k}\|\mathbf{h}_k^{(r)}(f)\|^2 df$. Before commenting this result, it is worth noting that inequality (31) 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 (31) also increase. Then, as we shall see below, the first term of the righthand side of (31) 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 (31). Ultimately, such an analysis will lead us to obtain a lower bound on $\liminf_{L\to\infty}\left(\liminf_{N\to\infty}\frac{N}{N_k}E_N\left[\left\|\hat{\boldsymbol{g}}_{N,k} - \boldsymbol{g}_{N,k}\right\|^2\right]\right)$. 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 $\boldsymbol{\mu}_k$ has a matrix-valued density $\mathbf{P}_k(f)$ w.r.t. the Lebesgue measure on $[0,1]$ (*i.e.*, $\boldsymbol{\mu}_k(df) = \mathbf{P}_k(f)df$). For each $f$, denote by $\lambda_k^{(1)}(f) < \cdots <$

$\lambda_k^{(N_T)}(f)$ the (non-negative) eigenvalues of $\mathbf{P}_k(f)$. Assume that $\lambda_k^{(1)}(f) \geq \epsilon$ for each $f \in \mathcal{D}_k$ and for some $\epsilon > 0$. Then,

$$\liminf_{N\to\infty} \frac{N}{N_k} E_N \left[ \left\| \hat{\boldsymbol{g}}_{N,k} - \boldsymbol{g}_{N,k} \right\|^2 \right] \geq \frac{\sigma^2 L N_R}{\Delta_k} \sum_{t=1}^{N_T} \int_{\mathcal{D}_k} \frac{1}{\lambda_k^{(t)}(f)} df + o_L(L) \tag{32}$$

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 III. □

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 $\mathbf{P}_k(f)$ 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 $\hat{\boldsymbol{g}}_{N,k}$ in the righthand side of equation (32) becomes

$$\frac{\sigma^2 L N_R}{\Delta_k} \int_{\mathcal{D}_k} \frac{1}{P_k(f)} df + o_L(L).$$

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

We now explore the training sequences that minimize the asymptotic CRB

$$asCRB_{\boldsymbol{g}_{N,k}} = \frac{\sigma^2 L N_R}{\Delta_k} \sum_{t=1}^{N_T} \int_{\mathcal{D}_k} \frac{1}{\lambda_k^{(t)}(f)} df \tag{33}$$

that stems from inequality (32).

### C. *Minimization of* $asCRB_{\boldsymbol{g}_{N,k}}$

We now study the training sequences which minimize the bound $asCRB_{\boldsymbol{g}_{N,k}}$ given by (33) on the MSE associated with estimates of the (useful) channel frequency taps. Again, the minimization is achieved under power constraint (21). In order to fit the assumption of Theorem 2 we assume that $\boldsymbol{\mu}_k$ has a matrix density $\mathbf{P}_k(f)$ whose eigenvalues are denoted by $\lambda_k^{(1)}(f) < \cdots < \lambda_k^{(N_T)}(f)$.

*Proposition 2:* Under power constraint (21), $asCRB_{\boldsymbol{g}_{N,k}}$ is minimum if and only if for each $f \in [0,1]$, $\mathbf{P}_k(f) = \frac{\mathcal{P}_k}{N_T \Delta_k} \mathbf{I}_{N_T} \mathcal{I}_{\mathcal{D}_k}(f)$, where $\mathbf{I}_{N_T}$ is the $N_T \times N_T$ identity matrix and where $\mathcal{I}_{\mathcal{D}_k}(f)$ is the indicator function of the useful bandwidth $\mathcal{D}_k$.

*Proof:* Based on definition (33), the minimization of $asCRB_{\boldsymbol{g}_{N,k}}$ under power constraint (21) is equivalent to the minimization of $\sum_{t=1}^{N_T} \int_{\mathcal{D}_k} \frac{1}{\lambda_k^{(t)}(f)} df$. For each $t = 1, \ldots, N_T$, Cauchy-Schwarz inequality implies that

$$\int_{\mathcal{D}_k} \frac{1}{\lambda_k^{(t)}(f)} df \int_{\mathcal{D}_k} \lambda_k^{(t)}(f) df \geq \Delta_k^2. \tag{34}$$

Note that equality holds in (34) if and only if $\lambda_k^{(t)}(f)$ is a constant w.r.t. $f$ in $\mathcal{D}_k$. Using (34) and Jensen's inequality, we obtain

$$\begin{aligned} asCRB_{\boldsymbol{g}_{N,k}} &\geq \sigma^2 L N_R \Delta_k \sum_{t=1}^{N_T} \frac{1}{\int_{\mathcal{D}_k} \lambda_k^{(t)}(f) df} \\ &\geq \sigma^2 L N_R \Delta_k \frac{N_T^2}{\sum_{t=1}^{N_T} \int_{\mathcal{D}_k} \lambda_k^{(t)}(f) df} \ . \end{aligned} \quad (35)$$

Equality holds in (35) if and only if *i)* $\lambda_k^{(t)}(f)$ is a constant w.r.t. $f \in \mathcal{D}_k$ and *ii)* all elements of $\left(\int_{\mathcal{D}_k} \lambda_k^{(t)}(f) df\right)_{t=1,\dots,N_T}$ are identical. This is equivalent to: $\forall f \in \mathcal{D}_k, \forall t, \lambda_k^{(t)}(f) = \lambda_k$ where $\lambda_k$ is a certain constant. In this case, matrix $\mathbf{P}_k(f)$ is proportional to the identity matrix, *i.e.*, $\mathbf{P}_k(f) = \lambda_k \mathbf{I}_{N_T}$ for each $f \in \mathcal{D}_k$. Finally, the denominator of the righthand side of (35) coincides with tr$(\int_{\mathcal{D}_k} \mathbf{P}_k(f) df)$. Thus, it is less than or equal to power constraint $\mathcal{P}_k$ and achieves $\mathcal{P}_k$ if and only if $\lambda_k = \frac{\mathcal{P}_k}{N_T \Delta_k}$ and $\mathbf{P}_k(f) = 0$ for $f \notin \mathcal{D}_k$. Thus, $asCRB_{\boldsymbol{g}_{N,k}}$ is minimum if and only if $\mathbf{P}_k(f) = \frac{\mathcal{P}_k}{N_T \Delta_k} \mathbf{I}_{N_T} \mathcal{I}_{\mathcal{D}_k}(f)$. □

**Comments**

- We recall that the *non* diagonal elements of matrix $\mathbf{P}_k(f)$ 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 $\mathbf{P}_k(f)$ 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 $\mathbf{P}_k(f)$ should be constant w.r.t. $f$ in the bandwidth of interest $\mathcal{D}_k$ so as to minimize $asCRB_{\boldsymbol{g}_{N,k}}$. 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.
- Propositions 1 and 2 lead to very different training sequences. This confirms the claim of [8] that no training sequence is jointly optimal for both channel and frequency offset estimation. Of course, the final goal would be to construct one single training sequence which would be relevant for both channel and CFO estimation. However, the selection of such a training sequence is difficult because it crucially relies on the particular receiver's architecture. One example of possible approach is for instance to follow the design criterion of [10]. In our case, this approach would consist in selecting the training sequence in order to minimize the objective function $\max_{\|\mathbf{h}_k\|=\rho} \left(asCRB_{\omega_k} + asCRB_{\mathbf{g}_{N,k}}\right)$ for a fixed channel norm $\rho$. We refer to [10] for more details. Ultimately, an even more relevant approach would be to select the training sequence which maximizes the capacity or minimize the bit error rate. An other relevant approach would be to select the training strategy which maximizes the

signal to noise ratio in data mode. Such optimizations are however out of the scope of the present paper and will be the subject of future works.

## V. Simulation Results

In order to sustain our claims, we present simulation results considering MIMO-OFDMA system with QPSK signaling. We considered either SISO case or MIMO case with $N_R = 2$ and $N_T = 1$ or $2$ depending on the context. 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 $[0, 5T]$. Due to the length of the impulse response of the shaping filter, the maximum channel length is about $8T$. Hence, we put $L = 8$. On the other hand, for each user $k$, the value of $\omega_k$ is randomly chosen in the interval $[-0.01, 0.01]$. 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 $\mathcal{P}_k$ are equal for all users. All results are averaged over $1000$ realizations of the training sequences and the channel parameters.

### A. Comparison of exact and asymptotic CRB

First, we study the values of $N$ for which our asymptotic results provide an accurate approximation of the exact CRB. The previously mentioned three 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 SAS, assigned subcarriers are modulated with equal power, training sequences sent at different antennas are uncorrelated). Finally, we introduce $T_3$ which is the optimal training strategy for frequency offset estimation as depicted in Section IV. In the single transmit antenna case, it consists in transmitting the highest power at the frequency $j$ such that $\frac{j}{N}$ is close to $f_1^{(opt)}$. In the multiple transmit antenna case, relevant correlation between antennas given by vector $\boldsymbol{\nu}_1^{(opt)}$ is furthermore introduced. We have taken the value of $\epsilon$ as $10^{-6}$ to ensure the regularity of $\mathbf{R}_k$. Note that $T_3$ is not a proper choice for channel estimation and will be used just for frequency offset estimation. The ratio $\frac{\mathcal{P}_1}{\sigma^2}$ between the transmitted energy per symbol and the noise variance is set to $20$ dB.

Fig. 1a compares exact and asymptotic CRB associated to *channel* parameters $\mathbf{g}_{N,1}$, for different values of the number $N$ of subcarriers. The number of users is equal to $2$ or $4$. The solid line represents the value $\frac{1}{N} asCRB_{L,\boldsymbol{g}_{N,1}}$, where $asCRB_{L,\boldsymbol{g}_{N,1}}$ denotes the asymptotic CRB on channel frequency taps equal to the righthand side of (31). Note that the value of $\frac{1}{N} asCRB_{L,\boldsymbol{g}_{N,1}}$ is identical for $T_1$ and $T_2$. Indeed, the asymptotic CRB only depends on the training sequence via the asymptotic covariance profile $\boldsymbol{\mu}_1$ and 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 (27). As shown by Fig. 1a, the exact CRBs depend on the particular training strategy $T_1$, $T_2$. Fig. 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 fits the asymptotic CRB. This sustains the claim that both training strategies have a similar performance for large $N$. The dotted line represents the value of $\frac{1}{N} asCRB_{\boldsymbol{g}_{N,1}}$, where $asCRB_{\boldsymbol{g}_{N,1}}$ is given by (33). Due to Theorem 2, $asCRB_{\boldsymbol{g}_{N,1}}$ approximates the asymptotic CRB $asCRB_{L,\boldsymbol{g}_{N,1}}$ of large values of $L$. In Fig. 1a, 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, we have $L = 8$. 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.

Fig. 1b compares exact and asymptotic CRB for frequency offset estimation as a function of $N$. Now, we add the performance curves for $T_3$. Again, the solid line represents $\frac{1}{N^3} asCRB_{\omega_1}$, where $asCRB_{\omega_1}$ is defined by (20). Other curves represent the exact CRB for $\omega_1$, defined as the coefficient at the first row and the first column of $\mathbf{CRB}_N$. As expected, exact CRBs corresponding respectively to training strategies $T_1$ and $T_2$ tend to be identical when $N$ increases. Moreover, all the exact CRBs fit the respective asymptotic bounds for large $N$. It is also worth noting that the exact CRB is close to the asymptotic CRB even for moderate values of $N$ in both cases.

Another interesting point is the behavior of the presented bounds as the number of users increase. Fig. 2 shows such a relation for both channel and frequency offset variances using $T_2$. We depict the curves for both $N = 512$ and $N = 1024$. As expected, the exact bounds fit the asymptotic ones for both SISO and MIMO cases as long as the number of subcarriers is larger than the number of unknown parameters (*i.e.*, $N$ larger than $K(N_R N_T L + 1)$). Particularly, in the SISO case, the difference between the exact and asymptotic bounds is negligible even in the presence of $32$ users. On the other hand, the asymptotic regime may no longer be reached in MIMO case for large number of users. For instance, when $N = 512$, the exact bounds deviates from the asymptotic bounds for the number of users greater than $16$.

Similar conclusions can be drawn for all training strategies including the ones which introduce a correlation between antennas. Fig. 3 shows such an example for an arbitrary sequence $T_4$. This training strategy is defined as follows. For each user $k$, we put $\mathbf{s}_{N,k}(j) = \mathbf{\Omega}\mathbf{d}_{N,k}(j)$ where the vector $\mathbf{d}_{N,k}(j)$ has i.i.d. entries and where

$$\mathbf{\Omega} = \frac{1}{\sqrt{1+\sin^2(\theta)}} \begin{bmatrix} 1 & \sin(\theta) \\ \sin(\theta) & 1 \end{bmatrix},$$

for $\theta = \frac{\pi}{6}$. As seen in Fig. 3, similar convergence behavior can be deduced for $T_4$. For the sake of comparison, we have also included the bounds for $T_2$ and $T_3$ which are found to be optimal training strategies respectively for channel and frequency estimation. The interesting point is the remarkable difference between the bounds of $T_4$ and respective optimal training strategies. These results simply imply that the arbitrarily selected training strategy $T_4$ is a successful choice neither for channel nor for frequency offset estimation.

### *B. Estimation Performance*

We now study the estimation performance associated to the frequency offset and channel parameters 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. Fig. 4 represents the corresponding MSE $E_N[(\hat{\omega}_{N,1} - \omega_1)^2]$ as a function of ratio $\frac{\mathcal{P}_1}{\sigma^2}$, for the training strategies $T_2$ and $T_3$. The MSE is compared with the asymptotic CRB $\frac{1}{N^3} asCRB_{\omega_k}$ in solid lines. Figs. 4a and 4b illustrate the performance when the number of transmit antennas is set to $N_T = 1$ or $N_T = 2$ respectively. For both training strategies, Fig. 4 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. Fig. 4a 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 $2.6$ dB in terms of $\frac{\mathcal{P}_1}{\sigma^2}$ 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 a 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 $2.6$ dB when $N_T = 1$. Fig. 4b 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 $5$ dB compared to the case of uncorrelated training sequences.

Fig. 5 illustrates the estimation performance associated with $\mathbf{g}_{N,1}$ as a function of $\frac{\mathcal{P}_1}{\sigma^2}$. Since $T_3$ is not relevant for channel estimation, results are given just for $T_2$. Similar to the previous case, the ML estimates coincide with exact CRBs and they are very close to the asymptotic bounds both for $N_T = 1$ and $N_T = 2$.

## 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. 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 $\frac{1}{N^3}$. 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 I
## PROOF OF LEMMA 1

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

$$\begin{aligned}\zeta_{N,k,l}^{(t,t')u}(p,q) &= \frac{u+1}{N^{u+1}} \sum_{n=0}^{N-1} n^u a_{N,k}^{(t)}(n-p)^* a_{N,l}^{(t')}(n-q) e^{2\imath\pi n \Delta f_{lk}} \\ &= \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} s_{N,k}^{(t)}(i)^* s_{N,l}^{(t')}(j) e^{2\imath\pi \frac{(pi-qj)}{N}} \psi_{N,u}(j - i + N\Delta f_{lk}), \qquad (36)\end{aligned}$$

where for each $x$, $\psi_{N,u}(x) = \frac{u+1}{N^{u+1}} \sum_{n=0}^{N-1} n^u e^{2\imath\pi \frac{nx}{N}}$. The following proof relies on the observation that for each $x$ such that $0 < |x| \leq \frac{N}{2}$,

$$|\psi_{N,u}(x)| \leq \frac{C_u}{|x|}, \qquad (37)$$

where $C_u$ is a constant. This inequality was recently used by [11]. For the sake of completeness, we provide a sketch of the proof. The claim can be easily shown for $\psi_{N,0}(x) = e^{\imath\pi\frac{(N-1)x}{N}}\frac{1}{N}\frac{\sin\pi x}{\sin\frac{\pi x}{N}}$ using the fact that $\left|\sin\frac{\pi x}{N}\right| \geq 2\left|\frac{x}{N}\right|$ for $|x| \leq \frac{N}{2}$. When $u = 1$, it can be shown that

$$\psi_{N,1}(x) = \frac{\imath}{N\sin\left(\frac{\pi x}{N}\right)}\left(e^{2\imath\pi\frac{(2N-1)x}{2N}} - \frac{e^{\imath\pi x}\sin(\pi x)}{N\sin\left(\frac{\pi x}{N}\right)}\right). \tag{38}$$

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 $1 + \left|e^{i\pi x}\frac{\sin(\pi x)}{N\sin(\frac{\pi x}{N})}\right| \leq 2$. The case $u = 2$ can be treated using similar arguments. The 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 $p, q$,

$$\chi_{N,k,k} = \zeta_{N,k,k}^{(t,t'),u} - \frac{1}{N}\sum_{i=0}^{N-1} E\left[s_{N,k}^{(t)}(i)^* s_{N,k}^{(t')}(i)\right] e^{2\imath\pi\frac{(p-q)i}{N}} \xrightarrow{N} 0 \quad \text{a.s.} \tag{39}$$

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

$$\chi_{N,k,k}^a = \frac{1}{N}\sum_{i=0}^{N-1}\left(s_{N,k}^{(t)}(i)^* s_{N,k}^{(t')}(i)\psi_{N,u}(0) - E\left[s_{N,k}^{(t)}(i)^* s_{N,k}^{(t')}(i)\right]\right) e^{2\imath\pi\frac{(p-q)i}{N}} \tag{40}$$

$$\chi_{N,k,k}^b = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{\substack{j=0\\ j\neq i}}^{N-1} s_{N,k}^{(t)}(i)^* s_{N,k}^{(t')}(j) e^{2\imath\pi\frac{(pi-qj)}{N}}\psi_{N,u}(j-i), \tag{41}$$

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 [11]. 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 righthand side of (41) except that the inner sum w.r.t. $j$ is restricted to the set $\mathcal{E}_i^1 = \{j = 0, \ldots, N-1/ j \neq i, |j - i| \leq \frac{N}{2}\}$. 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 $\mathcal{E}_i^2 = \{j = 0, \ldots, N-1/j \neq i, \frac{N}{2} < j - i \leq N-1\}$ and $\mathcal{E}_i^3 = \{j = 0, \ldots, N-1/ j \neq i, -N+1 \leq j - i < -\frac{N}{2}\}$ respectively. We now prove that $\chi_{N,k,k}^{b,1} \xrightarrow{a.s.} 0$. To that end, we show that

$E[|\chi^{b,1}_{N,k,k}|^4] \leq \frac{C}{N^2}$, where $C$ is a constant. We first expand $E[|\chi^{b,1}_{N,k,k}|^4]$ as follows.

$$E[|\chi^{b,1}_{N,k,k}|^4] = \frac{1}{N^4} \sum_{(i_1,i_2,i_3,i_4,j_1,j_2,j_3,j_4)\in\mathcal{V}} E\left[\prod_{n=1}^{4} \overline{s^{(t)}_{N,k}(i_n)^* s^{(t')}_{N,k}(j_n)}^{(n)}\right] \prod_{n=1}^{4} \overline{e^{\frac{2\imath\pi(pi_n - qj_n)}{N}} \psi_{N,u}(j_n - i_n)}^{(n)} \tag{42}$$

$$\leq \frac{1}{N^4} \sum_{(i_1,i_2,i_3,i_4,j_1,j_2,j_3,j_4)\in\mathcal{V}} \left| E\left[\prod_{n=1}^{4} \overline{s^{(t)}_{N,k}(i_n)^* s^{(t')}_{N,k}(j_n)}^{(n)}\right]\right| \prod_{n=1}^{4} |\psi_{N,u}(j_n - i_n)| \tag{43}$$

$$\leq \frac{C}{N^4} \sum_{(i_1,i_2,i_3,i_4,j_1,j_2,j_3,j_4)\in\mathcal{V}} \prod_{n=1}^{4} \frac{1}{|j_n - i_n|}\ , \tag{44}$$

where $\overline{x}^{(n)}$ is equal to $n$ if $x$ is odd, and to $x^*$ if $n$ is even. Inequality (43) comes from the triangular inequality. Then, applying Cauchy-Schwarz inequality successively and using (37) along with the assumption (10) leads to (44). Indeed,

$$\begin{aligned}
\left| E\left[\prod_{n=1}^{4} \overline{s^{(t)}_{N,k}(i_n)^* s^{(t')}_{N,k}(j_n)}^{(n)}\right]\right| &= \left| E\left[\prod_{n=1}^{4} \overline{s^{(t)}_{N,k}(i_n)^*}^{(n)} \prod_{n=1}^{4} \overline{s^{(t')}_{N,k}(j_n)}^{(n)}\right]\right| \\
&\leq \left( E\left[\prod_{n=1}^{4} |s^{(t)}_{N,k}(i_n)|^2\right] E\left[\prod_{n=1}^{4} |s^{(t')}_{N,k}(j_n)|^2\right]\right)^{\frac{1}{2}} \\
&\leq \left( \prod_{n=1}^{4} E\left[|s^{(t)}_{N,k}(i_n)|^8\right] \prod_{n=1}^{4} E\left[|s^{(t')}_{N,k}(j_n)|^8\right]\right)^{\frac{1}{8}} \\
&\leq \left(C^4 \times C^4\right)^{\frac{1}{8}} = C.
\end{aligned}$$

The sum in (42) is considered w.r.t. to all 8-uplet $(i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4) \in \mathcal{V}$ where $\mathcal{V}$ denotes the set of values of $i_1, \ldots, i_4, j_1, \ldots, j_4$ such that *i)* for each $n = 1, \ldots, 4$, $j_n \in \mathcal{E}^1_{i_n}$ 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. This is nothing but a direct result of the independence of the training symbols sent at different subcarriers. For instance, some terms of the sum (44) 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

$$\frac{C}{N^4} \sum_{i_1,i_3} \sum_{j_1,j_3} \frac{1}{|j_1 - i_1|^2} \frac{1}{|j_3 - i_3|^2} = \frac{C}{N^4} \left( \sum_{i=0}^{N-1} \sum_{j\in\mathcal{E}^1_i} \frac{1}{|j - i|^2}\right)^2 \leq \frac{C}{N^4} \left( \sum_{i=0}^{N-1} \sum_{l\neq 0} \frac{1}{|l|^2}\right)^2 \leq \frac{C'}{N^2}$$

where $C'$ is a constant. Other terms can be treated similarly. After some algebra, we obtain that $E[|\chi^{b,1}_{N,k,k}|^4] \leq \frac{C''}{N^2}$. Using the Borel-Cantelli Lemma, this implies that $\chi^{b,1}_{N,k,k} \xrightarrow{a.s.} 0$. By the same approach and the fact that $\psi_{N,u}$ is an $N$-periodic function, $\chi^{b,2}_{N,k,k}$ and $\chi^{b,3}_{N,k,k}$ can be shown to converge almost surely to zero. The proof of the a.s. convergence to zero of $\chi^{a}_{N,k,k}$ is similar. It can easily be seen using assumption (10) that $E[|\chi^a_{N,k,k}|^4]$ can be written as $\frac{1}{N^4}$ times (less than) $3N^2$ bounded terms. Thus $E[|\chi^a_{N,k,k}|^4] \leq \frac{C}{N^2}$, where $C$ is a constant. 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 $\zeta^{(t,t'),u}_{N,k,l}(p,q) = \zeta^{(1)}_{N,k,l} + \zeta^{(2)}_{N,k,l} + \zeta^{(3)}_{N,k,l}$ where $\zeta^{(1)}_{N,k,l}$ correspond to the righthand side of (36) except that the inner sum w.r.t. $j$ is restricted to $j \in \mathcal{F}_i$, where for each $i = 0, \ldots, N-1$, $\mathcal{F}_i = \{j = 0, \ldots, N-1 / |j - i + N\Delta f_{lk}| \leq \frac{N}{2}\}$. Similarly, $\zeta^{(2)}_{N,k,l}$ and $\zeta^{(3)}_{N,k,l}$ 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 $\zeta^{(1)}_{N,k,l} \overset{a.s.}{\to} 0$, it can be shown as in (44), by expanding $E[|\zeta^{(1)}_{N,k,l}|^4]$ and using (37), that

$$E[|\zeta^{(1)}_{N,k,l}|^4] \leq \frac{C}{N^4} \sum_{i_1,i_2,i_3,i_4} \sum_{j_1,\ldots,j_4} \prod_{n=1}^{4} \mathcal{M}(|j_n - i_n + N\Delta f_{lk}|), \tag{45}$$

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 $\mathcal{F}_{i_1} \times \ldots \times \mathcal{F}_{i_4}$ 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 $E[|\zeta^{(1)}_{N,k,l}|^4] \leq \frac{C'}{N^2}$. This proves that $\zeta^{(1)}_{N,k,l} \overset{a.s.}{\to} 0$. Terms $\zeta^{(2)}_{N,k,l}$ and $\zeta^{(3)}_{N,k,l}$ can be treated using the same approach.

## APPENDIX II
## PROOF OF PROPOSITION 1

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

$$\gamma_k = \frac{\epsilon}{N_T} \|\mathbf{h}_k\|^2 + \sup_{(A_i)} \sum_i \inf_{f \in A_i} \left[ \sum_{r=1}^{N_R} \boldsymbol{h}_k^{(r)}(f)^H \boldsymbol{\rho}_k(A_i) \boldsymbol{h}_k^{(r)}(f) \right], \tag{46}$$

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 $\boldsymbol{\rho}_k(A_i)$ is a non-negative Hermitian matrix. Based on this remark, we now make use of the following lemma.

*Lemma 3:* Denote by $(\mathbf{x}_r)_{r=1,\ldots,N_R}$ a sequence of $N_R$ complex column vectors of size $N_T \times 1$. Denote by $\lambda_{max}$ the largest eigenvalue of matrix $\sum_{r=1}^{N_R} \mathbf{x}_r {\mathbf{x}_r}^H$ and by $\boldsymbol{\nu}$ the corresponding eigenvector. For each $N_T \times N_T$ non-negative Hermitian matrix $\mathbf{M}$,

$$\sum_{r=1}^{N_R} {\mathbf{x}_r}^H \mathbf{M} \mathbf{x}_r \leq \lambda_{max} \mathrm{tr}(\mathbf{M}), \tag{47}$$

with equality if and only if (iff) $\mathbf{M}$ has the form $\mathbf{M} = \beta\,\boldsymbol{\nu}\boldsymbol{\nu}^H$ 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 $\mathbf{M} = \boldsymbol{\rho}_k(A_i)$. For each $f \in [0,1]$, define $\lambda_{k,max}(f)$ as the maximum eigenvalue of $\left[\sum_{r=1}^{N_R} \boldsymbol{h}_k^{(r)}(f)^H \boldsymbol{h}_k^{(r)}(f)\right]$ and $f_k^{(opt)} = \arg\max_f(\lambda_{k,max}(f))$. Then, using Lemma 3, we have

$$\sum_i \inf_{\{f\in A_i\}} \left[\sum_{r=1}^{N_R} \boldsymbol{h}_k^{(r)}(f)^H \boldsymbol{\rho}_k(A_i) \boldsymbol{h}_k^{(r)}(f)\right] \le \sum_i \inf_{\{f\in A_i\}} [\mathrm{tr}(\boldsymbol{\rho}_k(A_i))\lambda_{k,max}(f)] \tag{48}$$

$$\le \sum_i \mathrm{tr}(\boldsymbol{\rho}_k(A_i))\lambda_{k,max}(f_k^{(opt)}) \tag{49}$$

$$= \mathrm{tr}\left(\boldsymbol{\rho}_k([0,1])\right)\ \lambda_{k,max}(f_k^{(opt)}).$$

Equation (48) holds with equality iff for each $A_i$ and for each $f$, $\boldsymbol{\rho}_k(A_i)$ has the form $\boldsymbol{\rho}_k(A_i) = \beta(A_i)\boldsymbol{\nu}_k(f)\boldsymbol{\nu}_k(f)^H$, where $\boldsymbol{\nu}_k(f)$ represents the eigenvector associated with $\lambda_{k,max}(f)$. As a consequence, equation (49) holds with equality iff *i)* decomposition $(A_i)_i$ contains the one element set $\{f_k^{(opt)}\}$, *ii)* $\boldsymbol{\rho}_k(\{f_k^{(opt)}\}) = \beta(f_k^{(opt)})\boldsymbol{\nu}_k^{(opt)}\boldsymbol{\nu}_k^{(opt)H}$ where $\boldsymbol{\nu}_k^{(opt)} = \boldsymbol{\nu}_k(f_k^{(opt)})$, and *iii)* $\boldsymbol{\rho}_k(A_i) = 0$ if $f_k^{(opt)} \notin A_i$. Considering the supremum of the lefthand side of (48), we obtain that $\gamma_k \le \frac{\epsilon}{N_T}\|\mathbf{h}_k\|^2 + \mathrm{tr}(\boldsymbol{\rho}_k([0,1]))\lambda_{k,max}(f_k^{(opt)})$ with equality iff $\boldsymbol{\rho}_k$ has the form $\boldsymbol{\rho}_k = \beta(f_k^{(opt)})\boldsymbol{\nu}_k^{(opt)}\boldsymbol{\nu}_k^{(opt)H}\delta_{f_k^{(opt)}}$. Finally, introducing the power constraint (21), $\gamma_k \le \frac{\epsilon}{N_T}\|\mathbf{h}_k\|^2 + (\mathcal{P}_k - \epsilon)\lambda_{k,max}(f_k^{(opt)})$ with equality iff $\boldsymbol{\rho}_k = (\mathcal{P}_k - \epsilon)\boldsymbol{\nu}_k^{(opt)}\boldsymbol{\nu}_k^{(opt)H}\delta_{f_k^{(opt)}}$.

## APPENDIX III
## PROOF OF THEOREM 2

We study the behavior of the righthand side of (31) as the length $L$ of the channel increases. It is worth keeping in mind that (31) 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 $\mathbf{R}_k^{-1}(\mathbf{I}_{N_T} \otimes \mathbf{T}_k)$. For this, we make use of classical results on the behavior of large Toeplitz matrices [18] [20] [21]. More precisely, the proof requires the use of results on large *block*-Toeplitz matrices, which are direct generalizations of [18] (see [22] and references therein). Note however that the proof of Theorem 2 requires slightly more general results than those of [22].

Firstly, we study separately the asymptotic behaviors of $\mathbf{R}_k$ and $\mathbf{T}_k$. Secondly, we deduce from this the asymptotic behavior of $\mathrm{tr}(\mathbf{R}_k^{-1}(\mathbf{I}_{N_T} \otimes \mathbf{T}_k))$. We denote by $\mathbf{P}_k(f)$ the density of complex matrix-valued measure $\boldsymbol{\mu}_k$ w.r.t. the Lebesgue measure and by $P_k^{(t,u)}(f)$ the element of the $t$th row and the

$u$th column of $\mathbf{P}_k(f)$. We assume that each component $P_k^{(t,u)}(f)$ is a bounded function on $[0,1]$. For each $t,u = 1,\dots,N_T$, we denote by $\mathbf{R}_k^{(t,u)}$ the $L \times L$ block of coordinates $(t,u)$ of matrix $\mathbf{R}_k$, *i.e.*, $\mathbf{R}_k = \left(\mathbf{R}_k^{(t,u)}\right)_{t,u=1,\dots,N_T}$. We denote by $\mathbf{F}$ the $L \times L$ Fourier matrix $\mathbf{F} = \left(\frac{1}{\sqrt{L}} e^{-2\imath\pi\frac{ij}{L}}\right)_{i,j=0,\dots,L-1}$. Classical results on the asymptotic behavior of large Toeplitz matrices [18] imply that matrix $\mathbf{R}_k^{(t,u)}$ is asymptotically equivalent to matrix $\mathbf{\Theta}_k^{(t,u)}$ defined by

$$\mathbf{\Theta}_k^{(t,u)} = \mathbf{F}^H \mathbf{\Pi}_k^{(t,u)} \mathbf{F}, \tag{50}$$

where $\mathbf{\Pi}_k^{(t,u)} = \text{diag}\left(P_k^{(t,u)}(0), P_k^{(t,u)}(\frac{1}{L}), \dots, P_k^{(t,u)}(\frac{L-1}{L})\right)$. By *asymptotically equivalent*, we mean that spectral norms of $\mathbf{R}_k^{(t,u)}$ and $\mathbf{\Theta}_k^{(t,u)}$ are both bounded as $L \to \infty$ and that the normalized Frobenius norm $|\mathbf{R}_k^{(t,u)} - \mathbf{\Theta}_k^{(t,u)}|$ of $\mathbf{R}_k^{(t,u)} - \mathbf{\Theta}_k^{(t,u)}$ tends to zero as $L$ tends to infinity [21]. We recall that the normalized Frobenius norm $|\mathbf{X}|$ of a given $n \times n$ matrix $\mathbf{X} = (x_{ij})_{i,j=1,\dots,n}$ verifies $|\mathbf{X}|^2 = \frac{1}{n}\sum_i \sum_j |x_{ij}|^2$. From the above claim, it is straightforward to show that $\mathbf{R}_k$ is asymptotically equivalent to $\mathbf{\Theta}_k = \left(\mathbf{\Theta}_k^{(t,u)}\right)_{t,u=1,\dots,N_T}$ as $L \to \infty$. Based on (50), it is useful to note that $\mathbf{\Theta}_k$ can be written as

$$\mathbf{\Theta}_k = \left(\mathbf{I}_{N_T} \otimes \mathbf{F}^H\right) \mathbf{\Pi}_k \left(\mathbf{I}_{N_T} \otimes \mathbf{F}\right), \tag{51}$$

where $\mathbf{\Pi}_k = \left(\mathbf{\Pi}_k^{(t,u)}\right)_{t,u=1,\dots,N_T}$. Each block $\mathbf{\Pi}_k^{(t,u)}$ of $\mathbf{\Pi}_k$ is a diagonal matrix. "Renumbering" the elements of $\mathbf{\Pi}_k$ as in [22], it is straightforward to show that $\mathbf{\Pi}_k$ is equivalent to a block diagonal matrix up to a permutation of its rows and columns:

$$\mathbf{\Pi}_k = \mathbf{Q}^T \, \text{diag}\left(\mathbf{P}_k(0), \dots, \mathbf{P}_k(\frac{L-1}{L})\right) \mathbf{Q}, \tag{52}$$

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 $\mathbf{R}_k^{-1}$ and $\mathbf{T}_k$, 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 [20] 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 [11], 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 $\mathbf{X} = \mathbf{V}^H \text{diag}(\lambda_1, \dots, \lambda_n)\mathbf{V}$ where $\mathbf{V}$ is unitary, we generalize the above pseudo-inverse function by $F_\epsilon(\mathbf{X}) = \mathbf{V}^H \text{diag}(F_\epsilon(\lambda_1), \dots, F_\epsilon(\lambda_n))\mathbf{V}$. 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 [20], it is straightforward to show that $F_\epsilon(\mathbf{R}_k)$ is asymptotically equivalent to $F_\epsilon(\mathbf{\Theta}_k)$ as $L \to \infty$. Using (51) and (52), we obtain

$$F_\epsilon(\mathbf{\Theta}_k) = \left(\mathbf{I}_{N_T} \otimes \mathbf{F}^H\right) \mathbf{Q}^T \text{diag}\left(F_\epsilon\left(\mathbf{P}_k(0)\right), \ldots, F_\epsilon(\mathbf{P}_k(\frac{L-1}{L}))\right) \mathbf{Q}\left(\mathbf{I}_{N_T} \otimes \mathbf{F}\right). \tag{53}$$

We now study the asymptotic behavior of matrix $\mathbf{T}_k = \frac{1}{\Delta_k}\int_{\mathcal{D}_k} \mathbf{e}(f)\mathbf{e}(f)^H df$. Matrix $\mathbf{T}_k$ is an $L \times L$ Toeplitz matrix. Using again [18] along with the fact that function $\mathcal{I}_{\mathcal{D}_k}(f)$ is bounded, we immediately obtain that $\mathbf{T}_k$ is asymptotically equivalent to matrix $\mathbf{F}^H \mathbf{\Upsilon} \mathbf{F}$ where $\mathbf{\Upsilon} = \frac{1}{\Delta_k}\text{diag}\left(\mathcal{I}_{\mathcal{D}_k}(0), \ldots, \mathcal{I}_{\mathcal{D}_k}(\frac{L-1}{L})\right)$. As a consequence, $\mathbf{I}_{N_T} \otimes \mathbf{T}_k$ is asymptotically equivalent to

$$\text{diag}\left(\mathbf{F}^H\mathbf{\Upsilon}\mathbf{F}, \ldots, \mathbf{F}^H\mathbf{\Upsilon}\mathbf{F}\right) = \left(\mathbf{I}_{N_T} \otimes \mathbf{F}^H\right)\left(\mathbf{I}_{N_T} \otimes \mathbf{\Upsilon}\right)\left(\mathbf{I}_{N_T} \otimes \mathbf{F}\right). \tag{54}$$

Permuting rows and columns of $\mathbf{\Upsilon}$ in the same way as that of $\mathbf{\Pi}_k$, we obtain $\mathbf{I}_{N_T} \otimes \mathbf{\Upsilon} = \mathbf{Q}^T\left(\mathbf{\Upsilon} \otimes \mathbf{I}_{N_T}\right)\mathbf{Q}$, where $\mathbf{Q}$ is the same permutation matrix found in (53). Plugging this equality into (54) and putting (53) and (54) together, we conclude that matrix $F_\epsilon(\mathbf{\Theta}_k)\left(\mathbf{I}_{N_T} \otimes (\mathbf{F}^H\mathbf{\Upsilon}\mathbf{F})\right)$ coincides with the following matrix

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

Consequently, $F_\epsilon(\mathbf{R}_k)(\mathbf{I}_{N_T} \otimes \mathbf{T}_k)$ is asymptotically equivalent to the above matrix. Classical results [18] finally yield

$$\begin{aligned}
\lim_{L\to\infty} \frac{1}{L}\text{tr}\left(F_\epsilon(\mathbf{R}_k)(\mathbf{I}_{N_T} \otimes \mathbf{T}_k)\right) &= \lim_{L\to\infty}\frac{1}{L}\text{tr}\left(\mathbf{\Psi}\right) \\
&= \lim_{L\to\infty}\frac{1}{\Delta_k L}\sum_{j=0}^{L-1}\mathcal{I}_{\mathcal{D}_k}(\frac{j}{L})\text{tr}\left(F_\epsilon(\mathbf{P}_k(\frac{j}{L}))\right) \\
&= \frac{1}{\Delta_k}\int_{\mathcal{D}_k}\text{tr}\left(F_\epsilon(\mathbf{P}_k(f))\right) df \\
&= \frac{1}{\Delta_k}\sum_{t=1}^{N_T}\int_{\mathcal{D}_k} F_\epsilon(\lambda_k^{(t)}(f)) df,
\end{aligned} \tag{55}$$

where $\lambda_k^{(1)}(f) < \cdots < \lambda_k^{(N_T)}(f)$ are the eigenvalues of $\mathbf{P}_k(f)$. We now use the fact that for a given (finite) value of $L$, $\mathbf{R}_k$ and $\mathbf{T}_k$ are positive definite Hermitian matrices. This implies that $\text{tr}\left(\mathbf{R}_k^{-1}(\mathbf{I}_{N_T} \otimes \mathbf{T}_k)\right) \geq \text{tr}\left(F_\epsilon(\mathbf{R}_k)(\mathbf{I}_{N_T} \otimes \mathbf{T}_k)\right)$. Using (31) and (55), we finally obtain

$$\liminf_{L\to\infty}\frac{1}{L}\left(\liminf_{N\to\infty}\frac{N}{N_k}E\left[\left\|\hat{\boldsymbol{g}}_{N,k} - \boldsymbol{g}_{N,k}\right\|^2\right]\right) \geq \frac{\sigma^2 N_R}{\Delta_k}\sum_{t=1}^{N_T}\int_{\mathcal{D}_k} F_\epsilon(\lambda_k^{(t)}(f)) df. \tag{56}$$

The above equation holds for arbitrarily small values of $\epsilon$. Due to the assumptions of Theorem 2, $\int_{\mathcal{D}_k}\frac{1}{\lambda_k^{(1)}(f)}df$ exists. Therefore, the righthand side of (55) has a limit when $\epsilon \to 0$.
This limit coincides with $\frac{\sigma^2 N_R}{\Delta_k}\sum_{t=1}^{N_T}\int_{\mathcal{D}_k}\frac{1}{\lambda_k^{(t)}(f)}df$. This proves Theorem 2.

## References


[1] Y. Yao and G. B. Giannakis, "Blind carrier frequency offset estimation in SISO, MIMO, and multiuser OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 173-183, Jan. 2005.

[2] L. Hanzo, M. Münster, B. J. Choi, and T. Kellser, *OFDM and MC-CDMA for broadband multi-user communications, WLAN's and broadcasting*, IEEE Press, Wiley, 2003.

[3] S. Barbarossa, M. Pompili, and G.B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 2, pp. 474–486, Feb. 2002.

[4] J. J. Van de Beek *et al.*, "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1900–1914, Nov. 1999.

[5] Z. Cao, U. Tureli, and Y. D. Yao, "Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1585–1594, Sep. 2004.

[6] M. Morelli, "Timing and frequency synchronization for the Uplink of an OFDMA system," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 296–306, Feb. 2004.

[7] M. Pun, M. Morelli and C. J. Kuo, "Maximum Likelihood Synchronization and Channel Estimation for OFDMA Uplink Transmissions", *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.

[8] P. Ciblat and L. Vandendorpe, "On the maximum-likelihood based data-aided joint frequency offset and channel estimation," in *Proc. EUSIPCO'02*, pp. 627–630, Toulouse, France, Sept. 3-6, 2002.

[9] O. Besson and P. Stoica, "Training sequence selection for frequency offset estimation in frequency- selective channels," *Digital Signal Processing: Rev. J*, vol. 13, no. 1, pp. 106–127, Jan. 2003.

[10] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.

[11] S. Gault, W. Hachem, and P. Ciblat, "Joint sampling clock offset and channel estimation for OFDM signals: Cramér-Rao Bound and Algorithms," accepted for publication in *IEEE Trans. Signal Processing.*

[12] T. Cui and C. Tellambura, "Joint channel and frequency offset estimation and training sequence design for MIMO systems over frequency selective channels," in *Proc. Globecom'04*, Dallas, TX, USA, Nov. 29-Dec. 03, 2004.

[13] C. Ibars and Y. Bar-Ness, "Inter-Carrier Interference Cancellation for OFDM Systems with Macrodiversity and Multiple Frequency Offsets," *Wireless Personal Commun.*, no. 26(4), pp. 285-304, Sept. 2003.

[14] IEEE Std. 802.16a: *IEEE Standard for local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems-Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz*, 2003.

[15] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: Wiley, 1968, pt. I.

[16] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, New York, 1995.

[17] A. Vosoughi and A. Scaglione, "Everything You Always Wanted To Know About Training: Guidelines Derived Using The Affine Precoding Framework and The CRB," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 940–954, Mar. 2006.

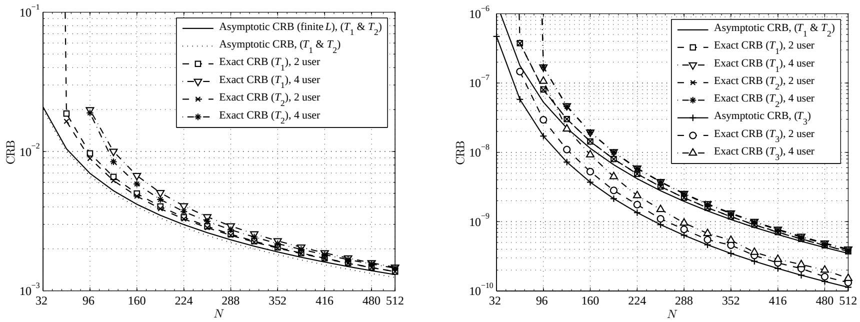
[18] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, University of California Press, Berkeley and Los Angeles, 1958.

[19] S. Sezginer and P. Bianchi, "Joint Frequency Offset and Channel Estimation in the OFDMA Uplink: Cramér-Rao Bounds and Training Sequence Design," in *Proc. SPAWC'05*, New York, USA, June. 5-8, 2005.

[20] H. Widom, "Toeplitz Matrices," in *Studies in Real and Complex Analysis*, edited by I.I. Hirschmann, Jr., *MAA Studies in Mathematics*, Prentice-Hall, Englewood Cliffs, NJ, 1965.

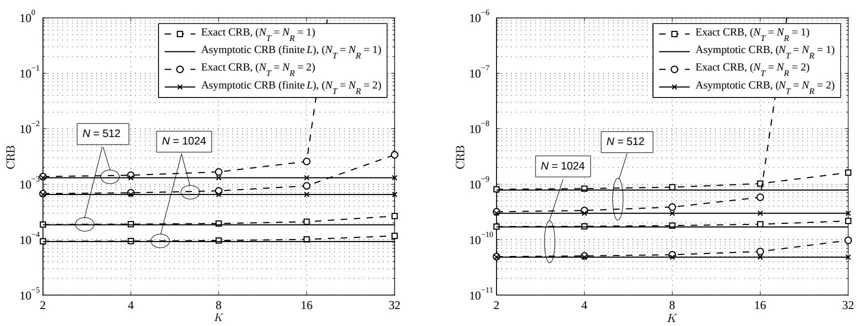
[21] R. M. Gray. (2000, Mar.) Toeplitz and circulant matrices: A review. Information Theory Laboratory, Stanford Univ., Stanford, CA. [Online]. Available: http://www-isl.stanford.edu/ gray/compression.html

[22] H. Gazzah, P. A. Regalia, and J. P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1243–1251, Mar. 2001.



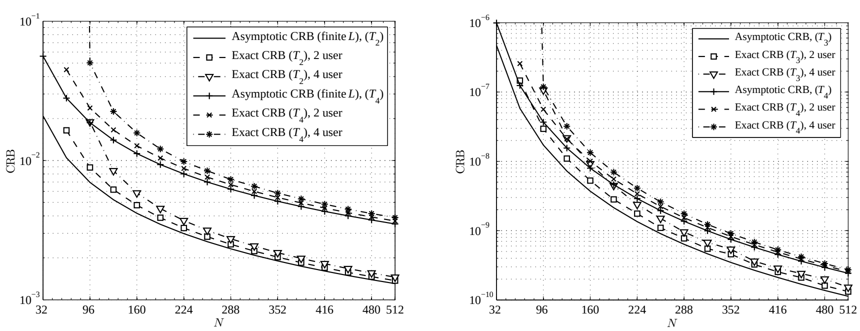

(a) CRB for $\mathbf{g}_{N,1}$ (b) CRB for $\omega_1$

Fig. 1. Exact and asymptotic CRB as a function of $N$ – $\frac{\mathcal{P}_1}{\sigma^2} = 20$ dB, $N_T = N_R = 2$.



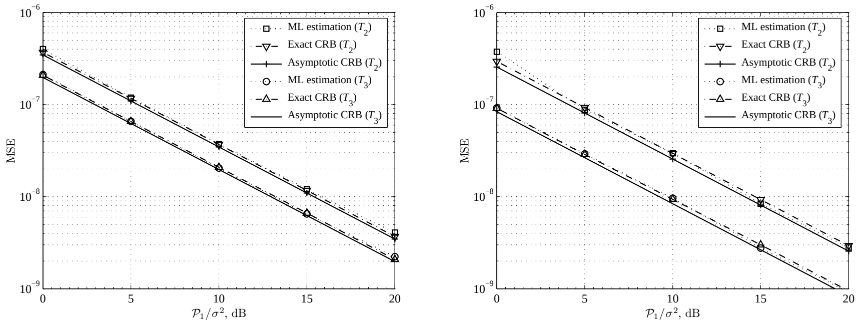

(a) CRB for $\mathbf{g}_{N,1}$ (b) CRB for $\omega_1$

Fig. 2. Exact and asymptotic CRB as a function of $K$ – $\frac{\mathcal{P}_1}{\sigma^2} = 20$ dB.




(a) CRB for $\mathbf{g}_{N,1}$ (b) CRB for $\omega_1$

Fig. 3. Exact and asymptotic CRB as a function of $N$ – $\frac{\mathcal{P}_1}{\sigma^2} = 20$ dB, $N_T = N_R = 2$.




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

Fig. 4. MSE of the ML estimator of $\omega_1$ and CRB as a function of $\frac{\mathcal{P}_1}{\sigma^2}$ – $N = 256$, $K = 2$.

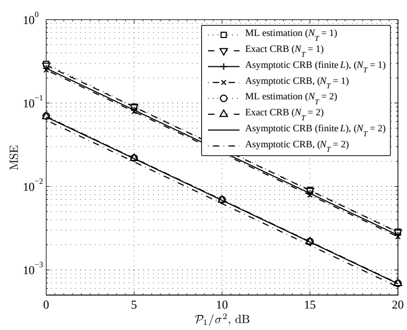



Fig. 5. MSE of the ML estimator of $\mathbf{g}_{N,1}$ and CRB as a function of $\frac{\mathcal{P}_1}{\sigma^2}$ – $N = 256$, $K = 2$.