CRAMÉR-RAO BOUND AND TRAINING SEQUENCE SELECTION
FOR MIMO-OFDMA TRANSMISSIONS IMPAIRED BY FREQUENCY OFFSETS

S. Sezginer and P. Bianchi

Supélec, Telecommunications Dept., Plateau du Moulon, 91192 Gif-sur-Yvette, France
e-mail: {serdar.sezginer, pascal.bianchi}@supelec.fr

ABSTRACT

In this paper, we investigate the data-aided estimation of frequency offsets and channel coefficients in the uplink transmission of MIMO-OFDMA systems. A compact and informative expression of the Cramér-Rao Bound (CRB) is derived for large training sequence sizes. It is shown that the asymptotic performance bounds do not depend on the number of users. 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 send at different antennas. On the otherhand, accurate channel estimation is achieved when training sequences send at different antennas are uncorrelated.

1. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) has recently become very popular in wireless communications and appears as a potential candidate for 4G mobile cellular systems. In this paper, we consider an uplink MIMO-OFDMA transmission involving N_t transmit (Tx) antennas per user and N_r receive (Rx) antennas at the base station. Each user modulates a certain group of subcarriers, following a given Carrier Assignment Scheme (CAS). The signal transmitted by a given user is impaired by a frequency selective channel and by a frequency offset due to Doppler effect and oscillator drifts.

The crucial impact of channels and frequency offsets estimation on further detection step in OFDMA has motivated a number of recent works (see,e.g., [6] and references therein). In this paper, we study the Cramér-Rao Bound (CRB) for the joint data-aided estimation of the K frequency offsets and the K MIMO channels, where K denotes the number of users. Unfortunately, the expression of the exact CRB turns out to be complicated. Thus, we investigate the case where the number N of subcarriers increases. In this case, a compact and informative expression of the (asymptotic) CRB can be obtained. Such an asymptotic analysis of the CRB has been proposed in [3] and [4] in case of a single-user single-carrier single-antenna system. The case of frequency offset estimation performance for MIMO systems has been investigated by [10] in case of a flat fading channel. Recently, the asymptotic CRB for clock offset and channel estimation has been studied by [5] in case of a single-user single antenna OFDM system.

Of course, the resulting CRB crucially depends on the training strategy employed by each user. Thus, following the general idea of [4], we bring to the fore the training design strategies which minimize the latter asymptotic performance bounds, with the goal of providing accurate estimates of parameters. Such a study has been recently provided by [7] in the single antenna case. However, it is worth noting

that the MIMO case is far from being a straightforward generalization of [7]. In particular, in the MIMO case, the design of relevant training sequences implies the selection of relevant beamformers at the transmitter side. In other words, the correlation between training sequences send at different Tx antennas of a given user should be selected so as to minimize the (bounds on the) estimation errors.

The paper is organized as follows. Section 2 describes the signal model. In Section 3, we derive the exact and the asymptotic CRB for the joint estimation of all parameters. In Section 4, we address the problem of training sequence design. Simulation results of Section 5 sustain our claims.

2. SIGNAL MODEL

Consider an uplink MIMO-OFDMA transmission, with N_t Tx antennas per user and N_r Rx antennas at the base station. Assume that K users share N subcarriers. We denote by $\frac{1}{NT}$ the spacing between two consecutive subcarriers. Hence, the complex envelope of the received signal is assumed to be sampled at rate T. After cyclic prefix removal, received samples at Rx antenna r can be written for each $n = 0, \ldots, N-1$ as

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

For each $k = 1, \ldots, K$, parameter ω_k is defined as $\omega_k = 2\pi\delta f_k T$, where δf_k denotes the frequency offset corresponding to user k. Since the difference between frequency offsets observed at different Tx-Rx pairs is usually negligible for a given user k, we assume without restriction that δf_k is a constant w.r.t. antenna pairs (t, r). Parameter $h_k^{(t,r)}(l)$ represents lth tap of the channel impulse response between t^{th}Tx antenna of user k and r^{th}Rx antenna. Each channel is assumed to have no more than L non zero 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. Finally, sequence $\{a_{N,k}^{(t)}(n)\}_n$ is the Inverse FFT of the training sequence $\{s_{N,k}^{(t)}(j)\}_j$ transmitted in the frequency domain by user k at Tx antenna t. Thus, for each n and each antenna t,

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

where j denotes the subcarrier index. Remark that, we simply put $s_{N,k}^{(t)}(j) = 0$ in the case where subcarrier j is not modulated by user k for the tth antenna. However, we do not specify any subcarrier assignment strategy at this point. Note also that training sequences $s_{N,k}^{(t)}(j)$ and $s_{N,k}^{(t')}(j)$ send at different antennas t and t' are possibly different.

In order to obtain a compact matricial representation of (1), 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 $(.)^T$ represents the transpose operator. Stacking all above channel parameters in a single vector, we define $\mathbf{h}_k = [\mathbf{h}_k^{(1,1)^T}, \ldots, \mathbf{h}_k^{(N_t,1)^T}, \mathbf{h}_k^{(1,2)^T}, \ldots, \mathbf{h}_k^{(N_t,N_r)^T}]^T$. For each user k, we define the $N \times N_t L$ matrix $\mathbf{A}_{N,k} = [\mathbf{A}_{N,k}^{(1)}, \ldots, \mathbf{A}_{N,k}^{(N_t)}]$, where the tth $N \times L$ block is equal to $[\mathbf{A}_{N,k}^{(t)}]_{i,j} = a_{N,k}^{(t)}(i-j)$ for $0 \le i \le N-1$ and $0 \le j \le L-1$. Then, stacking all $N_r N$ received samples (1) into a single received vector $\mathbf{y}_N$, we finally obtain:

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

where $\mathbf{I}_{N_r}$ denotes the $N_r \times N_r$ identity matrix, $\otimes$ denotes the Kronecker product and where $\mathbf{\Gamma}_N(\omega_k) = \text{diag}(1, e^{\imath\omega_k} \ldots e^{\imath\omega_k(N-1)})$. Here $\mathbf{v}_N$ is a Gaussian noise vector with independent and identically distributed entries of variance σ^2. In the sequel, we investigate the data aided estimation of the unknown parameter vector $\boldsymbol{\theta} = [\omega_1, \mathbf{h}_1^T, \ldots, \omega_K, \mathbf{h}_K^T]^T$.

3. ASYMPTOTIC CRAMÉR-RAO BOUND

We now study the CRB associated to parameter vector $\boldsymbol{\theta}$. 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.

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 $\mathbf{h}_k$. Using an approach similar to [4], the Fisher Information Matrix (FIM) for parameter $\tilde{\boldsymbol{\theta}}$ can be obtained as the following $K(1+2LN_rN_t) \times K(1+2LN_rN_t)$ matrix:

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

where $\boldsymbol{\eta}_N = \sum_{k=1}^{K} [\mathbf{I}_{N_r} \otimes (\mathbf{\Gamma}_N(\omega_k)\mathbf{A}_{N,k})] \mathbf{h}_k$. Here, superscript $(\cdot)^H$ denotes the transpose-conjugate, $\Re[x]$ (resp. $\Im[x]$) denotes the real (resp. imaginary) part of x. After some algebra, $\mathbf{J}_N$ can be written as the block matrix $\mathbf{J}_N = (\mathbf{J}_{N,k,l})_{1\le k,l \le K}$, where for each $k,l = 1 \ldots K$, $\mathbf{J}_{N,k,l}$ coincides with the following $(1+2LN_rN_t) \times (1+2LN_rN_t)$ matrix:

$$\frac{2}{\sigma^2} \begin{bmatrix} \alpha_{N,k,l} & \Im\left[\boldsymbol{\beta}_{N,l,k}^H\right] & \Re\left[\boldsymbol{\beta}_{N,l,k}^H\right] \\ -\Im\left[\boldsymbol{\beta}_{N,k,l}\right] & \Re\left[\mathbf{R}_{N,k,l}\right] & -\Im\left[\mathbf{R}_{N,k,l}\right] \\ \Re\left[\boldsymbol{\beta}_{N,k,l}\right] & \Im\left[\mathbf{R}_{N,k,l}\right] & \Re\left[\mathbf{R}_{N,k,l}\right] \end{bmatrix} \qquad (5)$$

where

$$\mathbf{R}_{N,k,l} = \mathbf{I}_{N_r} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{\Gamma}_N(\omega_l - \omega_k) \mathbf{A}_{N,l} \right), \qquad (6)$$

$$\boldsymbol{\beta}_{N,k,l} = \left[\mathbf{I}_{N_r} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{D}_N \mathbf{\Gamma}_N(\omega_l - \omega_k) \mathbf{A}_{N,l} \right) \right] \mathbf{h}_l, \qquad (7)$$

$$\alpha_{N,k,l} = \mathbf{h}_k^H \left[\mathbf{I}_{N_r} \otimes \left(\mathbf{A}_{N,k}^H \mathbf{D}_N^2 \mathbf{\Gamma}_N(\omega_l - \omega_k) \mathbf{A}_{N,l} \right) \right] \mathbf{h}_l, \quad (8)$$

and where $\mathbf{D}_N = \text{diag}(0, 1, 2, \ldots, N-1)$. Exact CRB is defined as the inverse of $\mathbf{J}_N$. However, for finite values of N, it seems difficult to obtain an informative expression of $\mathbf{J}_N^{-1}$ directly from (5).

Thus, we investigate the case where the number N of subcarrier increases. In this case, compact and informative expressions of the CRB can be obtained.

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

In order to simplify the following analysis, we assume that for a given antenna t of a given user k, $\{s_{N,k}^{(t)}(j)\}_{j=0\ldots N-1}$ is a sequence of independent random variables. Note that this assumption encompasses usual OFDMA training strategies [5]. However, we do *not* assume that training symbols are identically distributed. In particular, the variance $E[|s_{N,k}^{(t)}(j)|^2]$ of the jth training symbol depends on j. This is motivated by the observation that in practical OFDM systems, different powers may be allocated to different subcarriers. Moreover, for a given CAS, a certain number of subcarriers may not be modulated by user k. If j is one of these subcarriers, we simply consider that $E[|s_{N,k}^{(t)}(j)|^2] = 0$. Finally, 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 non zero.

For a given training strategy (*i.e.* for a given CAS, a given power allocation and a given correlation between antennas), we introduce the following matricial measure defined for any Borel set A of $[0,1]$

$$\boldsymbol{\mu}_{N,k}(A) = \frac{1}{N} \sum_{j=0}^{N-1} E\left[\mathbf{s}_{N,k}(j)\mathbf{s}_{N,k}(j)^H \right] I_A(\frac{j}{N}), \qquad (9)$$

where I_A stands for the indicator function of set A and where $\mathbf{s}_{N,k}(j) = [s_{N,k}^{(1)}(j) \ldots s_{N,k}^{(N_t)}(j)]^T$ is the vector of containing all training symbols modulating subcarrier j. We denote by $\mu_{N,k}^{(t,t')}(A)$ the coefficient of the tth row and the t'^{th} column of (9). In order to have some insights on the meaning of (9), it is interesting to remark that the first coefficient $\mu_{N,k}^{(1,1)}(A)$ verifies $\mu_{N,k}^{(1,1)}(A) = \frac{1}{N}\sum_{j=0}^{N-1} E\left[|s_{N,k}^{(1)}(j)|^2\right] I_A(\frac{j}{N})$. Therefore, measure $\mu_{N,k}^{(1,1)}$ can be interpreted as the power profile of the training sequence send 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 Tx antenna pair (t,t'), $\mu_{N,k}^{(t,t')}$ can be interpreted as the cross-correlation profile of the training sequences respectively send at antennas t and t'. Finally, matricial measure $\boldsymbol{\mu}_{N,k}$ is in some sense equivalent to the covariance profile of the vectorial training sequence $\mathbf{s}_{N,k}(j)$. We now make the following assumption.

Assumption 1 *For each k, we assume that there is a matricial measure $\boldsymbol{\mu}_k$ such that $\boldsymbol{\mu}_{N,k}$ converges weakly to $\boldsymbol{\mu}_k$ as $N \to \infty$.*

We will refer to $\boldsymbol{\mu}_k$ as the asymptotic covariance profile of the training sequence of user k. It is worth noting that the introduction of the above covariance profiles allows to thoroughly simplify the asymptotic analysis of the CRB. Moreover, (9) does not imply any restriction on the particular CAS, so that our results are valid for a large number of OFDMA training strategies. Note also that $\boldsymbol{\mu}_k$ jointly encompasses the CAS, the power allocation policy and the correlation between Tx antennas (due to possible beamforming).

Using (6), (7) and (8), $\mathbf{J}_N$ can be written as a continuous function of random matrices $\mathbf{A}_{N,k}^H \mathbf{D}_N^u \boldsymbol{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l}$ for $u = 0, 1, 2$ and for $k, l = 1 \dots K$. Thus, the asymptotic behavior of the CRB matrix is solely characterized by the asymptotic behaviors of the latter random matrices. After some algebra and the use of classical asymptotic analysis tools [1], one can show the following lemma.

Lemma 1 *For each $u = 0, 1, 2$, for each $k, l = 1 \dots K$,*

$$\frac{u+1}{N^{u+1}} \mathbf{A}_{N,k}^H \mathbf{D}_N^u \boldsymbol{\Gamma}_N(\omega_l - \omega_k)\mathbf{A}_{N,l} \xrightarrow{P} \delta(k-l) \, \mathbf{R}_k, \quad (10)$$

where $\delta(k - l)$ is equal to 1 if $k = l$ and equal to zero otherwise. Notation $\xrightarrow{P}$ stands for the (componentwise) convergence in probability as $N \to \infty$. We define the $LN_t \times LN_t$ matrix $\mathbf{R}_k$

$$\mathbf{R}_k = \int_0^1 \boldsymbol{\mu}_k(df) \otimes \left[\mathbf{e}(f)\mathbf{e}(f)^H \right], \quad (11)$$

where $\mathbf{e}(f) = [1, e^{2\imath \pi f}, \dots, e^{2\imath \pi f(L-1)}]^T$.

The proof is omitted due to the lack of space.

In accordance with (10), we rather focus on the normalized CRB, $\overline{\mathbf{CRB}}_N = \mathbf{W}_N \mathbf{CRB}_N \mathbf{W}_N$ where $\mathbf{W}_N$ is the $K(1 + 2LN_rN_t) \times K(1 + 2LN_rN_t)$ diagonal matrix defined by $\mathbf{W}_N = \mathrm{diag}(\mathbf{w}_N^T, \dots, \mathbf{w}_N^T)$ where $\mathbf{w}_N^T$ denotes the $(1 + 2LN_rN_t)$ row vector $\mathbf{w}_N^T = [N^{3/2}, N^{1/2}, \dots, N^{1/2}]$. Using (10), we obtain the following result.

Proposition 1 *As N tends to infinity, $\overline{\mathbf{CRB}}_N$ converges in probability to the block-diagonal matrix $\overline{\mathbf{CRB}}$ given by*

$$\overline{\mathbf{CRB}} = \mathrm{diag}\left(\mathbf{C}_1, \dots, \mathbf{C}_K\right). \quad (12)$$

For each $k = 1, \dots, K$, $\mathbf{C}_k$ is the $(1 + 2LN_rN_t) \times (1 + 2LN_rN_t)$ matrix equal to

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

where

$$\gamma_k = \sum_{r=1}^{N_r} \int_0^1 \mathbf{H}_k^{(r)}(f)^H \boldsymbol{\mu}_k(df) \mathbf{H}_k^{(r)}(f).$$

Here, $\mathbf{H}_k^{(r)}(f) = \sum_{l=0}^{L-1}[h_k^{(r,1)}(l), \dots, h_k^{(r,N_t)}(l)]^T e^{-2\imath \pi l f}$ may be interpreted as the overall frequency response of the channel "seen" at Rx antenna r.

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 are non correlated as $N \to \infty$. Looking more carefully at diagonal block $\mathbf{C}_k$, it can be easily seen that the estimation errors associated to the parameters of a user k depend neither on the parameters of other users $l \neq k$ nor on the number K of users. The asymptotic CRB is also independent from the values of the frequency offsets.

We now further study the bounds on the Mean Square Errors (MSE) for the parameters of a given user k. We define

$$\overline{CRB}_{\omega,k} = \frac{6\sigma^2}{\gamma_k} \quad (13)$$

$$\overline{CRB}_{h,k} = N_r\sigma^2 \mathrm{tr}\left(\mathbf{R}_k^{-1}\right) + \frac{3\sigma^2}{2}\frac{\mathbf{h}_k^H\mathbf{h}_k}{\gamma_k} \quad (14)$$

where $\mathrm{tr}(\mathbf{X})$ stands for the trace of $\mathbf{X}$. For any unbiased estimate $\hat{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, Proposition 1 implies that

$$\liminf_{N\to\infty} N^3 E\left[(\hat{\omega}_{N,k} - \omega_k)^2\right] \geq \overline{CRB}_{\omega,k} \quad (15)$$

$$\liminf_{N\to\infty} NE\left[\left\|\hat{\mathbf{h}}_{N,k} - \mathbf{h}_k\right\|^2\right] \geq \overline{CRB}_{h,k} \quad (16)$$

where $\hat{\omega}_{N,k}$ and $\hat{\mathbf{h}}_{N,k}$ respectively denote the estimates of the (angular) frequency offset ω_k and channel coefficients $\mathbf{h}_k$.

4. TRAINING SEQUENCE SELECTION

In this Section, we characterize the training sequences which are likely to provide accurate estimates of parameters. It was shown in [3], [4] and [11] that no training sequence is likely to jointly minimize the bounds on the MSE of all parameters. In order to overcome this problem, [4] 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 delicate, and is out of the scope of the present paper. Here, we separately study the minimization of $\overline{CRB}_{\omega,k}$ and $\overline{CRB}_{h,k}$.

Proposition 1 indicates that the design of all K training sequences can be achieved separately, without restriction. In words, each user k is likely to select its own training sequence without taking other users into account. As a consequence, **we simply drop index k in the sequel**. Proposition 2 characterizes the training strategies that minimize (13).

Proposition 2 *For each $f \in [0, 1]$, denote by $\lambda_{max}(f)$ the maximum eigenvalue of $\sum_{r=1}^{N_r} \mathbf{H}^{(r)}(f)\mathbf{H}^{(r)}(f)^H$. Define $f_{opt} = \arg\max_f \lambda_{max}(f)$. Assuming power constraint $tr(\boldsymbol{\mu}([0,1])) \leq \mathcal{P}$, $\overline{CRB}_\omega$ is minimum if and only if $\boldsymbol{\mu} = \mathcal{P}\boldsymbol{\nu}_{opt}\boldsymbol{\nu}_{opt}^H \delta_{f_{opt}}$ where $\boldsymbol{\nu}_{opt}$ is the eigenvector associated to $\lambda_{max}(f_{opt})$ and $\delta_{f_{opt}}$ represents the Dirac measure at f_{opt}.*

The proof of the above result make use of classical results on integration [1] and is omitted due to the lack of space. The optimum training strategy for accurate frequency offset estimation is determined by the covariance profile $\boldsymbol{\mu}$ defined by Proposition 2. In particular, this suggests that an accurate estimate of ω can be obtained by transmitting all available power at the frequency for which the largest eigenvalue of $\sum_{r=1}^{N_r} \mathbf{H}^{(r)}(f)\mathbf{H}^{(r)}(f)^H$ is maximum. Moreover, the cross-correlation between training sequences send at different antennas is dictated by eigenvector $\boldsymbol{\nu}_{opt}$. In other words, $\boldsymbol{\nu}_{opt}$ can be interpreted as the optimal beamformer to be used at transmitter side.

Of course, the above training strategy requires the channel knowledge at the transmitter side (or at least some limited channel knowledge consisting in f_{opt} and $\boldsymbol{\nu}_{opt}$). Thus, such a strategy is only applicable under certain channel feedback conditions. However, in certain applications, channel information at the transmitter side may be limited to some channel statistics. Although this particular case is somewhat out of the scope of the paper, it is of practical interest to provide general ideas that can be used for designing training sequences in the case where the channel information is limited to the covariance matrix $\boldsymbol{\Sigma} = E\left[\mathbf{hh}^H\right]$ of the channel. In the sequel, we assume that the channel is Rayleigh distributed. We propose to minimize the expectation w.r.t. $\mathbf{h}$ of $\overline{CRB}_\omega$, i.e. $E_\mathbf{h}\left[\overline{CRB}_\omega\right]$. According to (15), this is equivalent to the minimization of the lower bound on the average MSE of the frequency offset estimate. The direct calculation of $E_\mathbf{h}\left[\overline{CRB}_\omega\right]$ is unfortunately a difficult task. Following an idea recently used by [8], we propose to approximate the distribution of γ

in (13) with a Gamma distribution. After some algebra, we obtain

$$E_{\mathbf{h}}\left[\overline{CRB}_\omega\right] \simeq \frac{6\sigma^2 \mathrm{tr}((\mathbf{I}_{N_r}\otimes \mathbf{R})\mathbf{\Sigma})}{[\mathrm{tr}((\mathbf{I}_{N_r}\otimes \mathbf{R})\mathbf{\Sigma})]^2 - \mathrm{tr}\left([(\mathbf{I}_{N_r}\otimes \mathbf{R})\mathbf{\Sigma}]^2\right)}. \quad (17)$$

The issue of minimizing (17) w.r.t. the training strategy is however out of the scope of the paper and will be investigated in future works.

We now study the training sequences which minimize $\overline{CRB}_h$ given in (14). In order to simplify what follows, we assume that $\boldsymbol{\mu}$ has a density and we denote this density by $\mathbf{P}(f)$, *i.e.* $\boldsymbol{\mu}(df) = \mathbf{P}(f)df$. We also assume for the sake of simplicity that for each f, density $\mathbf{P}(f)$ is positive definite. Since the minimization of (14) in the general case seems to be difficult, we rather focus on the case where the length L of the channel increases (note that we still assume $N \gg L$). Using recent results on large block Toeplitz matrices [9], it can be shown that

$$\overline{CRB}_h = L\,N_r\sigma^2 \sum_{t=1}^{N_t} \int_0^1 \frac{1}{\lambda_t(\mathbf{P}(f))} df + O_L(1) \quad (18)$$

where $\lambda_t(\mathbf{X})$ denotes the t^{th}eigenvalue of matrix $\mathbf{X}$. After some algebra, one can show the following Proposition, which holds for sufficiently large values of L.

Proposition 3 *Under power constraint* $tr(\int_0^1 \mathbf{P}(f)df) \leq \mathcal{P}$, $\overline{CRB}_h$ *is minimum iff for each* $f \in [0,1]$, $\mathbf{P}(f) = \frac{\mathcal{P}}{N_t}\mathbf{I}_{N_t}$.

As a consequence, accurate estimates of channel coefficients are likely to be obtained when the following training strategy is used. Firstly, $\mathbf{P}(f)$ is a diagonal matrix for each f. This suggests to send uncorrelated training sequences at different antennas. Secondly, $\mathbf{P}(f)$ does not depend on frequency f. This suggests to select training sequences having a flat power profile.

5. SIMULATION RESULTS

We consider MIMO-OFDMA system with $K = 2$ users and $N = 64$ subcarriers. The number of Rx antennas is $N_r = 2$ and the number of Tx antennas N_t is either 1 or 2. Results are averaged over 500 realizations of the training sequence. We used QPSK symbols. The following results have been obtained using one fixed realization of a Rayleigh channel with length $L = 3$. For each realization, the values of frequency offsets ω_1 and ω_2 are randomly chosen in the interval $[-0.4, 0.4]$. Transmitted powers $\mathcal{P}_1$ and $\mathcal{P}_2$ are supposed to be equal. Due to the lack of space, we focus on frequency offset estimation. Figure 1 represents the MSE associated to the frequency offset of the first user (i.e. $E[(\hat{\omega}_{N,1} - \omega_1)^2]$) as a function of the transmitted signal-to-noise ratio (SNR) $\frac{\mathcal{P}_1}{\sigma^2}$. Results are given for both "white" (T1) and optimum (T2) training strategies. Both training strategies T1 and T2 are realized using interleaved CAS, i.e. each user $k = 1, 2$ modulates $\frac{N}{2}$ subcarriers. In case of training strategy T_1, training sequences corresponding to different antennas are uncorrelated and equal power $2\frac{\mathcal{P}_1}{N_t}$ is allocated to each of the $\frac{N}{2}$ modulated subcarriers. On the otherhand, T2 is chosen so as to approximate the optimal training strategy suggested by Proposition 2. Maximum power is allocated to the subcarrier whose index coincides with the integer part of Nf_{opt}. A cross-correlation dictated by eigenvector $\boldsymbol{\nu}_{opt}$ is introduced between sequences send at different antennas. Solid lines represent the exact CRB on ω_1 (i.e. the coefficient at the first row and the first column of $\mathbf{J}_N^{-1}$). Dotted lines represent the asymptotic bound $\frac{1}{N^3}\overline{CRB}_{\omega,1}$. For both training strategies, the performance of the ML estimate of ω_1 is close to the asymptotic CRB and almost fit the exact
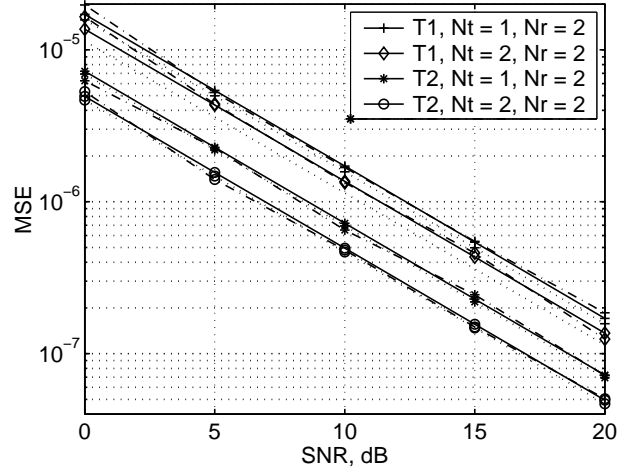
Fig. 1. ML estimate and CRBs for ω - Comparison between white sequence (T1) and optimum sequence (T2).
— Exact CRB   −· ML estimation   ··· Asymptotic CRB

bounds. Also for both training strategies, we note a performance improvement of about 1.5dB when introducing a second Tx antenna. Comparing both training strategies, T2 provides about 3.4 dB gain w.r.t. T1 in the case $N_t = 1$, $N_r = 2$. This gap increases when the number of Tx antennas increases. As a conclusion, relevant training sequence selection results in considerable performance gain.

6. REFERENCES

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

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

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

[4] 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.

[5] 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.

[6] M. Pun, M. Morelli and C. J. Kuo, "Maximum Likelihood Synchronization and Channel Estimation for OFDMA Uplink Transmissions", submitted to IEEE Trans. Commun.

[7] 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.

[8] H. Minn, S. Xing, and V. K. Bhargava, "Optimal periodic training signal for frequency offset estimation in frequency-selective fading channels," in Proc. ICC'04, Paris, France, pp. 488-492, June 20-24, 2004.

[9] 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.

[10] P. Stoica and O. Besson, "On the parameter estimation of MIMO flat-fading channels with frequency offsets," IEEE Trans. Signal Processing, vol. 51, no. 3, pp. 602-613, Mar. 2003.

[11] 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.