OPTIMAL TRAINING FOR FREQUENCY OFFSET ESTIMATION IN CORRELATED-RICE FREQUENCY-SELECTIVE CHANNEL

Philippe Ciblat[(1)], *Pascal Bianchi*[(2)], *and Mounir Ghogho*[(3)]

[(1)] Département Comélec, ENST, Paris, France
[(2)] Département Télécom, Supélec, Gif-sur-Yvette, France
[(3)] School of Electrical Engineering, Leeds University, Leeds, United Kingdom
Email : philippe.ciblat@enst.fr, pascal.bianchi@supelec.fr, m.ghogho@leeds.ac.uk

ABSTRACT

We consider carrier frequency offset (CFO) estimation for single-carrier and single-user transmission over a frequency-selective channel. When training is solely devoted to frequency synchronization, it is important to design the training to optimize CFO estimation performance. In this paper we exhibit the training sequence that minimizes the Cramer-Rao bound associated with the carrier frequency offset and averaged over the channel statistics following a correlated Ricean fading channel model. Simulations show significant improvements compared to the standard pseudo-random white training sequence.

1. INTRODUCTION

In wireless communication, the transmitted signal is usually distorted by inter symbol interference due to the multipath channel and a carrier frequency offset (CFO) caused by a local oscillator drift or a Doppler effect. In order to retrieve the transmitted data accurately, the receiver needs to accurately estimate the CFO and the channel. In civil applications, symbols known to the receiver, called training symbols, are sent periodically by the transmitter. The training sequence can be employed to estimate the channel and the carrier frequency offset. Often, both parameters are estimated separately with two different kinds of training sequences. In this paper we focus on the carrier frequency offset estimation when the channel is still not a parameter of interest and is unknown at the receive and transmit sides.

The goal of this paper is to exhibit the "optimal" training sequence in the sense of optimizing CFO estimation performance. Toward this objective, we will minimize the Cramer-Rao bound associated with the CFO. In order for the solution to be independent of the channel realization, we average the CRB over the channel statistics model. In wireless environment, it is standard to consider the channel to be Rice distributed [1]. In the literature, the optimal training sequence for carrier frequency offset estimation has been developed in the case of uncorrelated Rayleigh channel [2]. In such a case, the best training sequence is white. In this paper, we will see that as soon as the channel has either a deterministic part or a correlated random part, the best training sequence is not white anymore.

As we consider a single-carrier and single-user communications scheme, the complex envelope of receive discrete-time signal $y(n)$ takes the following form

$$y(n) = e^{2i\pi f n}\sum_{l=0}^{L-1} h(l)t(n-l) + w(n), \tag{1}$$

where f denotes the carrier frequency offset and coefficients $h(0)\dots h(L-1)$ represent the channel coefficients. The training sequence $t(0), t(1), \dots, t(N_T-1)$ with length N_T devoted to CFO estimation is assumed to be transmitted periodically. Sequence $w(n)$ denotes a white complex-valued circular zero-mean Gaussian noise of variance $\sigma^2 = \mathbb{E}[|w(n)|^2]$.

In the sequel, we denote by $\mathbf{h} = [h(0)\dots h(L-1)]^{\mathrm{T}}$ the unknown channel vector and by $\mathbf{t} = [t(0), t(1)\dots t(N_T-1)]^{\mathrm{T}}$ the vector of training symbols. The superscript $(.)^{\mathrm{T}}$ stands for the transpose operator.

The channel is assumed to be Rice distributed *i.e.*,

$$\mathbf{h} = \sqrt{\frac{K}{K+1}}\mathbf{h}_d + \sqrt{\frac{1}{K+1}}\mathbf{h}_r \tag{2}$$

where $\mathbf{h}_d$, also called Line-Of-Sight (LOS) component, is a deterministic vector normalized in such a way that $\|\mathbf{h}_d\|^2 = 1$ and where $\mathbf{h}_r$, also called Non-Line-Of-Sight (NLOS) component, is a complex circular Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma} = \mathbb{E}[\mathbf{h}_r\mathbf{h}_r^{\mathrm{H}}]$, normalized in such a way that $\mathrm{Tr}\,(\mathbf{\Sigma}) = 1$. The superscript $(.)^{\mathrm{H}}$ stands for the conjugate transpose operator. Coefficient K is the so-called Ricean factor. We assume that K, $\mathbf{\Sigma}$, and $\mathbf{h}_d$ are known at both the transmitter and the receiver sides. This assumption is realistic in most wireless applications since the coherence time corresponding to K, $\mathbf{\Sigma}$, and $\mathbf{h}_d$ is much larger than the coherence time corresponding to $\mathbf{h}_r$.

In order to fix our optimal training sequence selection issue, the derivations and the optimization of the average CRB will be done under the following asymptotic regime: we first assume that the size N_T of the training sequence tends to infinity, and *then* assume the size L of the channel impulse response tends to infinity. The last assumption is equivalent to assuming $\lim_{N_T\to\infty} L/N_T = 0$ which means that $L \ll N_T$. It is worth pointing out that although our theoretical analysis is based on this asymptotic regime, realistic values for N_T and L are used in the simulation section.

When the number of available observations N_T is much greater than the number of unknown parameters $(L+1)$, it is well known that the Bayesian approach which takes into account the statistical model of the parameters (e.g., the Maximum A Posteriori based algorithm) does not provide significant improvement compared to the deterministic approach (e.g., the Maximum Likelihood based

This work has been supported by the Royal Society

algorithm). In the case of asymptotic regime, the performance of the Maximum Likelihood based algorithm is described by the deterministic Cramer-Rao bound [3, 4] associated with the joint estimation of the CFO and the channel $\mathbf{h}$. Therefore the use of the deterministic Cramer-Rao bound averaged over channel statistics as our criterion is well motivated.

The paper is organized as follows: in Section 2, we introduce the CRB associated with CFO and we derive its average over the channel statistics under the asymptotic regime. In Section 3, we show that the average CRB can be approximated by a convex function with respect to the training sequence spectrum. Section 4 presents simulations results which illustrate the merits of the proposed method.

2. CRITERION EVALUATION

The CRB associated with CFO depends on the training sequence. Consider the training sequence as a realization of a zero-mean stationary random sequence. Then, when N_T is large and when the channel $\mathbf{h}$ is unknown at the receiver and has to be estimated too, the CRB for the CFO can be expressed as a function of the second-order statistics of the training sequence (cf. [5]). This CRB is given by [5]

$$\gamma(f||\mathbf{h}\text{ fixed and unknown}) = \frac{3\sigma^2}{2\pi^2 N_T^3}\frac{1}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}} \qquad (3)$$

where $\mathbf{R}_t$ is the L-dimensional covariance matrix defined by $\{r(k-l)\}_{k,l=0,\cdots,L-1}$ with $r(k-l) = \mathbb{E}[t(n+k)\overline{t(n+l)}]$.

Our goal consists of selecting the correlation matrix that minimizes the CFO estimation error. Obviously, minimizing Eq. (3) with respect to $\mathbf{R}_t$ leads to an optimal color for the training sequence depending on the channel realization. This is unrealistic in practical systems since the channel is unknown. Therefore, we now focus on the expectation of the term (3) over the channel statistics. The resulting function, which will be our optimization criterion, only depends on the training sequence statistics and on the channel statistics and is defined as follows

$$J(\mathbf{R}_t) = s\mathbb{E}_{\mathbf{h}}\left[\frac{1}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}}\right] \quad \text{with} \quad s = \frac{3\sigma^2}{2\pi^2 N_T^3}. \qquad (4)$$

We would like to obtain a closed form expression for $J(\mathbf{R}_t)$. Such an expression is given in the next theorem.

Theorem 1 *Let $J(\mathbf{R}_t)$ be given by Eq. (4). If L is large, we have that*

$$J(\mathbf{R}_t) \approx s(K+1)\frac{Tr(\mathbf{R}_t\tilde{\mathbf{A}})}{(Tr(\mathbf{R}_t\tilde{\mathbf{A}}))^2 - Tr(\mathbf{R}_t\tilde{\mathbf{B}}\mathbf{R}_t\tilde{\mathbf{C}})} \qquad (5)$$

where

$$\begin{aligned}\tilde{\mathbf{A}} &= \mathbf{\Sigma} + K\mathbf{h}_d\mathbf{h}_d^{\mathrm{H}}\\ \tilde{\mathbf{B}} &= \mathbf{\Sigma}\\ \tilde{\mathbf{C}} &= \mathbf{\Sigma} + 2K\mathbf{h}_d\mathbf{h}_d^{\mathrm{H}},\end{aligned}$$

and where $Tr(.)$ stands for the Trace operator.

Proof– First of all, we write $\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}$ as a linear combination of non-central chi-square distribution. Let

$$\mathbf{x}_d = \mathbf{R}_t^{1/2}\sqrt{\frac{K}{K+1}}\mathbf{h}_d \quad \text{and} \quad \mathbf{x}_r = \mathbf{R}_t^{1/2}\sqrt{\frac{1}{K+1}}\mathbf{h}_r$$

Let $\mathbf{R}_x = \mathbb{E}[\mathbf{x}_r\mathbf{x}_r^{\mathrm{H}}]$ be the autocorrelation matrix of $\mathbf{x}_r$. The matrix can be diagonalized as follows

$$\mathbf{R}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}$$

where $\mathbf{U}$ is an unitary matrix in which are stacked the eigenvectors of $\mathbf{R}_x$ and where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_L)$ with λ_k being the k^{th} eigenvalue of $\mathbf{R}_x$. Let $\tilde{\mathbf{x}}_d$ and $\tilde{\mathbf{x}}_r$ be the following vectors

$$\tilde{\mathbf{x}}_d = \mathbf{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{x}_d \quad \text{and} \quad \tilde{\mathbf{x}}_r = \mathbf{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{x}_r.$$

$\tilde{\mathbf{x}}_d$ is a deterministic vector whereas $\tilde{\mathbf{x}}_r$ is a circularly Gaussian distributed vector with zero-mean and unit-variance. Then we get

$$\xi = \mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h} = \sum_{k=1}^{L}\lambda_k|\tilde{x}_d(k) + \tilde{x}_r(k)|^2$$

where $\tilde{x}_d(k)$ and $\tilde{x}_r(k)$ are the k^{th} component of $\tilde{\mathbf{x}}_d$ and $\tilde{\mathbf{x}}_r$ respectively.

Secondly, it is well known ([6] and references therein) that a weighted sum of non-central chi-square distribution of two degres of freedom can be well approximated by a central Gamma distribution. Let $p_G(t)$ be a Gamma distribution with standard parameters (p_1, p_2). Then we get

$$p_G(t) = \frac{p_1^{p_2}}{\Gamma(p_2)}t^{p_2-1}e^{-p_1 t}\mathbf{1}_{t\geq 0}$$

Notice that the mean (resp. variance) of a Gamma distribution is given by p_2/p_1 (resp. p_2/p_1^2). When L is large enough, the distribution of random variable ξ is close to that of a Gamma distribution of parameters (p_1, p_2) such that

$$\begin{aligned}p_2/p_1 &= \mathbb{E}[\xi]\\ p_2/p_1^2 &= \mathbb{E}[(\xi - \mathbb{E}[\xi])^2]\end{aligned}$$

After straightforward but tedious algebraic manipulations, we obtain

$$p_1 = \frac{\sum_{k=1}^{L}\lambda_k(1+|\tilde{x}_d(k)|^2)}{\sum_{k=1}^{L}\lambda_k^2(1+2|\tilde{x}_d(k)|^2)} \qquad (6)$$

$$p_2 = p_1\sum_{k=1}^{L}\lambda_k(1+|\tilde{x}_d(k)|^2) \qquad (7)$$

Thirdly, it remains to evaluate the expectation of $1/\xi$ when ξ is assumed to be Gamma distributed. One can easily check that

$$\mathbb{E}\left[\frac{1}{\xi}\right] \approx \int_0^\infty \frac{1}{t}p_G(t)dt = \frac{p_1}{p_2 - 1}$$

Thanks to Eqs (6)-(7), one can check that p_2 is larger than 1.

Finally, we need to evaluate p_1 and p_2 with respect to the channel statistics $(K, \mathbf{h}_d, \mathbf{\Sigma})$ and the training correlation matrix $\mathbf{R}_t$. For doing that, recall that

$$\tilde{\mathbf{x}}_d = \mathbf{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{R}_t^{1/2}\sqrt{\frac{K}{K+1}}\mathbf{h}_d$$

which implies that

$$\sum_{k=1}^{L}\lambda_k|\tilde{x}_d(k)|^2 = \frac{K}{K+1}\mathbf{h}_d^{\mathrm{H}}\mathbf{R}_t\mathbf{h}_d$$

and

$$\sum_{k=1}^{L}\lambda_k^2|\tilde{x}_d(k)|^2 = \frac{K}{(K+1)^2}\mathbf{h}_d^{\mathrm{H}}\mathbf{R}_t\mathbf{\Sigma}\mathbf{R}_t\mathbf{h}_d.$$

As $\mathbf{\Lambda}$ represents the diagonal matrix of eigenvalues of $\mathbf{R}_x$, we have

$$\sum_{k=1}^{L}\lambda_k = \mathrm{Tr}(\mathbf{R}_x) = \frac{1}{K+1}\mathrm{Tr}(\mathbf{R}_t\mathbf{\Sigma})$$

and

$$\sum_{k=1}^{L}\lambda_k^2 = \mathrm{Tr}(\mathbf{R}_x^2) = \frac{1}{(K+1)^2}\mathrm{Tr}(\mathbf{R}_t\mathbf{\Sigma}\mathbf{R}_t\mathbf{\Sigma})$$

which concludes the proof. ■

The expression provided in Theorem 1 is an extension of the one reported in [2] for the iid Rayleigh channel case (when $K=0$ and $\mathbf{\Sigma} = (1/L)\mathbf{Id}_L$). Given the expression of $J(\mathbf{R}_t)$, the optimization problem associated with the minimization of Eq. (5) with respect to the matrix $\mathbf{R}_t$ subject to power constraint seems to be intractable in closed-form. In next section, we will however see that under the asymptotic regime, $J(\mathbf{R}_t)$ can be well approximated by a convex function. This implies that our problem boils down to a standard convex optimization issue.

3. CRITERION OPTIMIZATION

By simplifying Eq. (5), the proposed criterion gives rise to a convex optimization problem for which we are able to obtain numerical solutions for the covariance matrix of the training sequence, or equivalently its spectrum.

First of all, let us focus on the simplification of $J(\mathbf{R}_t)$ given by Eq. (5). When L is large, the $(L\times L)$ Toeplitz matrix $\mathbf{R}_t$ can be almost diagonalized in the Fourier basis [7]. Consequently, we get

$$\mathbf{R}_t = \mathbf{F}\mathbf{\Lambda}_t\mathbf{F}^{\mathrm{H}}$$

where $\mathbf{F}$ is the $(L\times L)$ Fourier matrix and where $\mathbf{\Lambda}_t$ is a diagonal matrix whose l^{th} component is equal to the spectrum of $\mathbf{t}$ at FFT bin l/L. As a consequence, $J(\mathbf{R}_t)$ only depends on $\boldsymbol{\lambda}_t = [\lambda_1,\ldots,\lambda_L]$ and takes the following approximated form, denoted by $\boldsymbol{\lambda}_t \mapsto J_a(\boldsymbol{\lambda}_t)$,

$$J_a(\boldsymbol{\lambda}_t) = s(K+1)\frac{\sum_{l=1}^{L}\lambda_l a_{ll}}{(\sum_{l=1}^{L}\lambda_l a_{ll})^2 - \sum_{k,l=1}^{L}\lambda_k\lambda_l b_{kl}c_{lk}} \quad (8)$$

where a_{kl}, b_{kl}, and c_{kl} are the components of the k^{th} row and l^{th} column of matrices $\mathbf{A} = \mathbf{F}^{\mathrm{H}}\tilde{\mathbf{A}}\mathbf{F}$, $\mathbf{B} = \mathbf{F}^{\mathrm{H}}\tilde{\mathbf{B}}\mathbf{F}$, and $\mathbf{C} = \mathbf{F}^{\mathrm{H}}\tilde{\mathbf{C}}\mathbf{F}$, respectively.

In next theorem, we show that the criterion $\boldsymbol{\lambda}_t \mapsto J_a(\boldsymbol{\lambda}_t)$ is convex.

Theorem 2 *When L is large, $\mathbf{R}_t \mapsto J(\mathbf{R}_t)$ can be replaced with $\boldsymbol{\lambda}_t \mapsto J_a(\boldsymbol{\lambda}_t)$ defined by Eq. (8). Such a function is convex with respect to $\boldsymbol{\lambda}_t$ as soon as $\lambda_l \geq 0$ for all $l \in \{1,\cdots,L\}$.*

Proof– Instead of proving directly the convexity of $J_a(\boldsymbol{\lambda}_t)$, we prove more easily the convexity of $\phi(\boldsymbol{\lambda}_t) = -s(K+1)/J_a(\boldsymbol{\lambda}_t)$. As J_a is, by construction, positive whatever the matrix $\mathbf{R}_t$ and as the inverse function is also convex, the convexity of ϕ implies the convexity of J_a. Therefore we now concentrate on $\boldsymbol{\lambda}_t \mapsto \phi(\boldsymbol{\lambda}_t)$ which can be written as follows

$$\phi(\boldsymbol{\lambda}_t) = \frac{\sum_{k,l=1}^{L}\lambda_k\lambda_l b_{kl}c_{lk}}{\sum_{l=1}^{L}\lambda_l a_{ll}} - \sum_{l=1}^{L}\lambda_l a_{ll}$$

To prove the convexity of $\phi(\boldsymbol{\lambda}_t)$, we calculate its Hessian matrix, defined as follows

$$\mathcal{H} = \left[\frac{\partial^2\phi(\lambda_1,\cdots,\lambda_L)}{\partial\lambda_m\partial\lambda_n}\right]_{m,n=1,\cdots,L}$$

After tedious but straightforward algebraic manipulations, we have

$$\begin{aligned}\frac{\partial^2\phi}{\partial\lambda_m\partial\lambda_n} &= \frac{1}{S_{\boldsymbol{\lambda}_t}}\Big[b_{nm}c_{mn} + b_{mn}c_{nm}\Big]\\ &- \frac{1}{S_{\boldsymbol{\lambda}_t}^2}\Big[2b_{mm}c_{mm}a_{nn}\lambda_m + 2b_{nn}c_{nn}a_{mm}\lambda_n\\ &+ a_{mm}\sum_{\substack{k=1\\k\neq n}}^{L}\lambda_k(b_{kn}c_{nk} + b_{nk}c_{kn})\\ &+ a_{nn}\sum_{\substack{k=1\\k\neq m}}^{L}\lambda_k(b_{km}c_{mk} + b_{mk}c_{mk})\Big]\\ &+ \frac{1}{S_{\boldsymbol{\lambda}_t}^3}\Big[2a_{mm}a_{nn}\sum_{k,l=1}^{L}\lambda_k\lambda_l b_{kl}c_{lk}\Big]\end{aligned}$$

with $S_{\boldsymbol{\lambda}_t} = \sum_{l=1}^{L}\lambda_l a_{ll}$.

We need to prove that the Hessian matrix $\mathcal{H}$ is positive. Therefore we focus on the following term $c = \mathbf{x}^{\mathrm{T}}\mathcal{H}\mathbf{x}$ where $\mathbf{x} = [x_1,\cdots,x_L]^{\mathrm{T}}$ is a real-valued vector of length L. Let $\mathbf{T} = \mathbf{B}\odot\mathbf{C}^{\mathrm{T}}$ where $\odot$ stands for the element-by-element Hadamard product. We easily obtain that

$$c = \frac{2}{S_{\boldsymbol{\lambda}_t}}\left[\mathbf{x}^{\mathrm{T}}\mathbf{T}\mathbf{x} - \frac{(\boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}\mathbf{x} + \boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{x})S_{\mathbf{x}}}{S_{\boldsymbol{\lambda}_t}} + \frac{(\boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}\boldsymbol{\lambda}_t)S_{\mathbf{x}}^2}{S_{\boldsymbol{\lambda}_t}^2}\right]$$

with $S_{\mathbf{x}} = \sum_{l=1}^{L}x_l a_{ll}$.

Since $\mathbf{B}$ and $\mathbf{C}$ are hermitian positive, so is $\mathbf{T}$ [8]. As a consequence, we can apply the following Schwartz inequality

$$|\boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}\mathbf{x}|^2 \leq (\boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}\boldsymbol{\lambda}_t)(\mathbf{x}^{\mathrm{T}}\mathbf{T}\mathbf{x})$$

leading to

$$c \geq \frac{2}{S_{\boldsymbol{\lambda}_t}}\left[\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{T}\mathbf{x}} - \frac{\sqrt{\boldsymbol{\lambda}_t^{\mathrm{T}}\mathbf{T}\boldsymbol{\lambda}_t}\,|S_{\mathbf{x}}|}{S_{\boldsymbol{\lambda}_t}}\right]^2 \geq 0$$

which concludes the proof.■

Our optimization problem thus is convex since the function to be minimized is convex and the constraints are also convex

$$\mathrm{Tr}(\mathbf{R}_t) = \sum_{l=1}^{L}\lambda_l = LP \quad \text{and} \quad \lambda_l \geq 0,\ \forall l \quad (9)$$

To obtain numerical values for optimal $\boldsymbol{\lambda}_t$, we can use a standard gradient or Newton search algorithm. The convexity of our criterion ensures proper convergence of the latter algorithm.

Notice that minimizing the CRB of the CFO with respect to the training sequence has been partially treated in the literature. [9] selects the training sequence which minimizes the worst CRB (i.e., the maximum of the CRB over all normalized channel realizations) whereas [2] characterizes the training sequence which minimizes the average CRB when the channel is assumed to be i.i.d. Rayleigh (i.e., $K = 0$ and $\boldsymbol{\Sigma}$ proportional to the identity matrix). In both of the above references, the white training was found to be optimal. In [10], the training sequence was designed to render the exact (i.e. finite-sample) CRB of the CFO independent of the channel zeros. The developed design was shown to outperform the white training design when the length of the training sequence is small/moderate. For a large training sequence, the white sequence is still optimal when considering the criterion in [10]. Here we have derived the training sequence minimizing the average CRB when the channel is assumed to be Rice whatever the ricean factor K, the deterministic part $\mathbf{h}_d$ and the color of the random part $\boldsymbol{\Sigma}$. We have shown that white training may be far from being optimum.

In practice, a training sequence with optimal power spectrum $\boldsymbol{\lambda}_t$ can be generated as the output of a digital filter with relevant coefficients excited by a known white sequence. The generation of this optimal training sequence can be achieved without any additional computational complexity compared to a traditional (white) training sequence since the above filter has to be evaluated only when the channel statistics vary.

4. SIMULATIONS

Unless otherwise stated, we set $N_t = 50$, $L = 5$, and SNR=10dB. The carrier frequency offset is fixed to $f = 0.1$. All simulated points are averaged over 1000 Monte-Carlo runs for which we have modified the deterministic and random part of the channel. We assume that the correlation between two taps k and l of the channel impulse response is given by $\rho^{|k-l|}$ where $\rho \in [0, 1)$.

In Figure 1, we compare the CRB given by Eq. (4) with its approximation provided by Eq. (8) versus L. We have considered $\mathbf{R}_t = \mathbf{Id}_L$, $\boldsymbol{\lambda}_t = \mathrm{diag}(\mathbf{Id}_L)$, $K = 2$, $\rho = 0.75$. We remark that the approximation is tight enough since both curves are close to each other even for small value of L. This implies that the best training sequence deduced from the optimization done on Eq. (8) is still relevant for the original problem characterized by Eq. (4).

In Figure 2, we plot the approximate CRB versus K for uncorrelated channel ($\rho = 0$) and with white or optimal training sequence. We observe that for small K the white sequence is almost optimal. However, as soon as K is larger than 3, it is preferable to choose a training sequence different from the white one to estimate more accurately the frequency offset.

In Figure 3, we display $\boldsymbol{\lambda}_t$ the spectrum of the optimal training sequence associated with one specific realization of $\mathbf{h}_d$. The Ricean factor is fixed to $K = 10$. We observe that the energy of the spectrum is concentrated on the frequency maximizing the spectrum of $\mathbf{h}_d$. We remind that, if the channel $\mathbf{h}$ is perfectly known at the transmitter, the spectrum of the optimal training sequence is a spectral line at the frequency maximizing the spectrum of $\mathbf{h}$. Consequently, when K is large enough, the choice of the optimal training sequence spectrum can be done in a similar way but based on $\mathbf{h}_d$ instead of $\mathbf{h}$.

In Figure 4, we still display $\boldsymbol{\lambda}_t$ the spectrum of the optimal

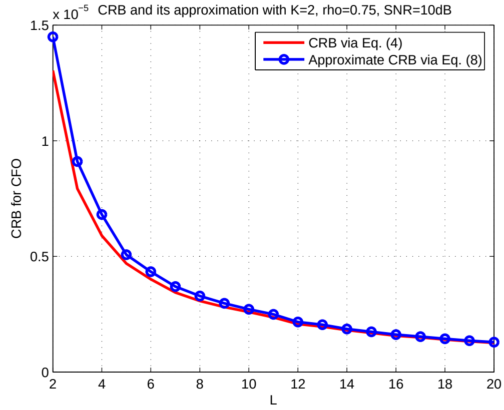

Fig. 1. J and J_a versus L

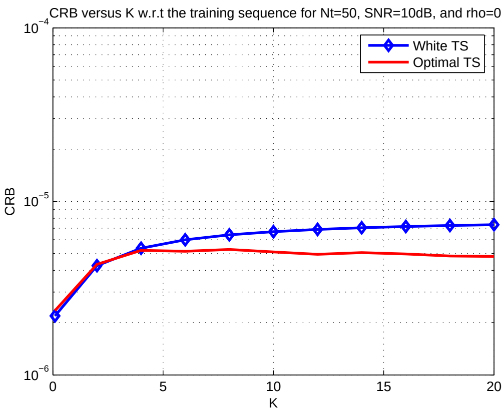

Fig. 2. Approximate CRB versus K ($\rho = 0$)

training sequence associated with one specific realization of $\mathbf{h}_d$ but with a low Ricean factor $K = 2$. In such a case, $\mathbf{h}_d$ does not provide enough information about $\mathbf{h}$. To avoid frequency fading of $\mathbf{h}$, the optimal training sequence is thus spread over all FFT bins.

In Figure 5, we plot the approximate CRB versus ρ for a Rayleigh channel ($K = 0$). The gain in performance is of interest for $\rho > 0.5$. The correlation thus needs to be strong to observe a difference between optimal training and white training.

In Figure 6, we display $\boldsymbol{\lambda}_t$ the spectrum of the optimal training sequence for Rayleigh channel with $\rho = 0.5$. We also plot the mean channel spectrum given by $\nu \mapsto \mathbb{E}[|h(\nu)|^2]$. We remark that the optimal training sequence spectrum has a shape similar to the mean spectrum which concentrates its energy at low frequencies.

5. CONCLUSION

We have inspected the training sequence design issue for carrier frequency offset estimation in the context of frequency-selective channel when the channel impulse response is assumed to be Rice-distributed with arbitrary correlation. As a criterion, we have chosen the Cramer-Rao bound for the carrier frequency offset, aver-

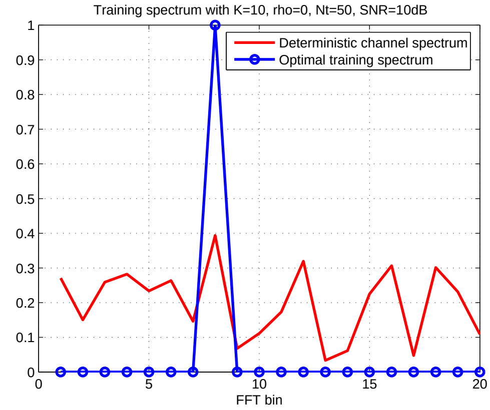

Fig. 3. Optimal $\boldsymbol{\lambda}_t$ for $K = 10$ and $\rho = 0$

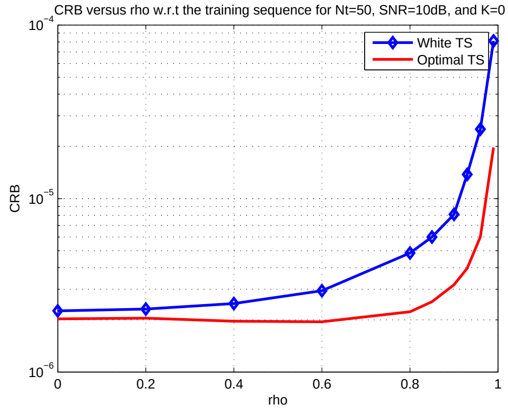

Fig. 5. Approximate CRB versus ρ ($K = 0$)

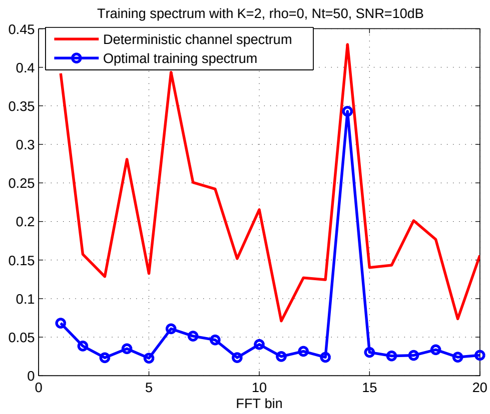

Fig. 4. Optimal $\boldsymbol{\lambda}_t$ for $K = 2$ and $\rho = 0$

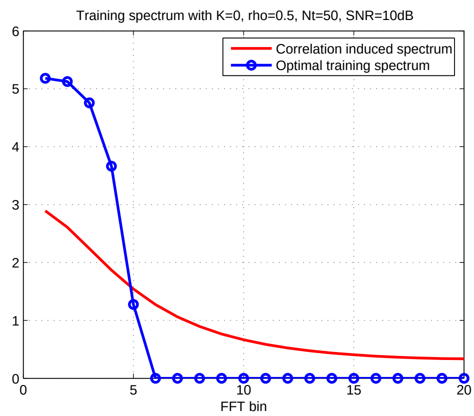

Fig. 6. Optimal $\boldsymbol{\lambda}_t$ for $\rho = 0.5$ and $K = 0$

aged over the channel statistics. The proposed criterion was simplified under an asymptotic assumption on the length of the training sequence and the length of the channel impulse response. Then we proved that the simplified criterion is convex which enables us to find numerically the optimal training sequence. Simulations showed that the gain was significant.

6. REFERENCES

[1] V. Erceg et al., "Channel models for fixed wireless applications", *IEEE 802.16 broadband wireless access working group report*, July 2001.

[2] H. Minn, X. Fu, and V. Bhargava, "Optimal periodic training signal for frequency offset estimation in frequency-selective fading channels", *IEEE Trans. on Comm.*, vol. 54, pp. 1081-1096, Jun. 2006.

[3] M. Morelli and U. Mengali, "Carrier frequency estimation for transmissions over selective channels", *IEEE Trans. on Comm.*, vol. 48, pp. 1580-1589, Sep. 2000.

[4] P. Ciblat and L. Vandendorpe, "On the Maximum-Likelihood based data-aided frequency offset and channel estimates", in *EUSIPCO*, Sep. 2002.

[5] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation", *IEEE Trans. on Comm.*, vol. 51, pp. 1910-1917, Nov. 2003.

[6] Q.T. Zhang and D.P. Liu, "A simple capacity formula for correlated diversity Ricean fading", *IEEE Communications Letters*, vol. 6, issue 11, pp. 481-483 , Nov. 2002.

[7] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, Univ. California (Berkeley) Press, 1958.

[8] R.A. Horn and C.R. Johnson, "Matrix Analysis", *Cambridge University Press*, 1985.

[9] O. Besson and P. Stoica, "Training sequence selection for frequency offset estimation in frequency selective channels", *Digital Signal Processing: A Review Journal*, vol. 13, pp.106-127, Jan. 2003.

[10] M. Ghogho and A. Swami, "Training Design for Multipath Channel and Frequency-Offset Estimation in MIMO Systems," *IEEE Trans. on Signal Processing*, vol. 54(10), pp. 3957-65, Oct. 2006.