TRAINING SEQUENCE DESIGN FOR JOINT CHANNEL AND FREQUENCY OFFSET ESTIMATION WITH PARTIAL CHANNEL STATE INFORMATION

Pascal Bianchi(1) *and Philippe Ciblat*(2)

(1) Département Télécom, Supélec, Gif-sur-Yvette, France
(2) Département Comélec, ENST, Paris, France
Email : `pascal.bianchi@supelec.fr, philippe.ciblat@enst.fr`

ABSTRACT

We consider signal-carrier and single-user transmission over a frequency-selective channel. We focus on the data-aided joint estimation of the dispersive channel and the frequency offset. We propose a new training sequence selection strategy relevant for both parameters of interest in the context of a Ricean fading channel. Our strategy relies on the minimization of the Mean-Square Error on data symbols at the output of a Wiener equalizer after frequency offset compensation. Simulations based on bit error rate confirm our claim.

1. INTRODUCTION

Usually the transmitted signal can be affected by inter symbol interference due to the multipath channel and a frequency offset caused by a Doppler effect or a local oscillator drift. Before applying an offset correction and an equalizer, the channel and the frequency offset, which are unknown, have to be estimated. In many applications, parameter estimation is achieved via the transmission of known (training) symbols prior to the unknown data symbols. A natural question is therefore: how to select the training sequence at the transmitter side, so that relevant estimates of the unknown parameters can be obtained at the receiver side ? Whereas the design of optimal training sequence is fixed when only the channel is unknown or when only the frequency offset is unknown [2, 1], the design of the optimal or relevant training sequence associated with the joint estimation issue is still an open problem. This is the concern of this paper.

Let us consider our signal model. We focus on a single-carrier and single-user communications scheme. Assume that a training sequence $t(0), t(1), \ldots, t(N_T-1)$ with length N_T is transmitted. The received signal $y(n)$ has the form:

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

where parameter f denotes the frequency offset and where coefficients $h(0) \ldots h(L-1)$ represent the channel coefficients. Sequence $w(n)$ denotes a white complex-valued circular zero-mean Gaussian noise of variance $\sigma^2 = \mathbb{E}[|w(n)|^2]$. Here, $t(n)$ represents the training sequence. In the sequel, we denote by $\mathbf{h} = [h(0) \ldots h(L-1)]^{\mathrm{T}}$ the unknown channel vector and by $\mathbf{t} = [t(0), t(1) \ldots t(N_T-1)]^{\mathrm{T}}$ the vector of training symbols. Training symbols are assumed to be known by the receiver. Here, the superscript $(.)^{\mathrm{T}}$ represents the transpose operator. For the sake of simplicity, we assume that training sequence $t(n)$ is a realization of a random stationary sequence, possibly correlated. 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 $\mathbf{h}_d$ is a deterministic vector normalized in such a way that $\|\mathbf{h}_d\|^2 = 1$ and where $\mathbf{h}_r$ is a complex circular Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma} = E[\mathbf{h}_r\mathbf{h}_r^{\mathrm{H}}]$, 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. In the sequel, we respectively refer to the first and the second term of the righthand side of (2) as the line of sight (LOS) and the non line of sight (NLOS) components of the channel.

In the present paper, we assume that the LOS component $\mathbf{h}_d$ of the channel is known at both the transmitter and the receiver sides. This is motivated by the fact that in most wireless applications, the coherence time corresponding to the LOS component is much larger than the coherence time corresponding to the NLOS component. Note that in this case, the estimation of $\mathbf{h}$ is actually equivalent to the estimation of the NLOS component $\mathbf{h}_r$. We denote the unknown parameter vector as $\boldsymbol{\theta} = [f, \mathbf{h}_R, \mathbf{h}_I]^{\mathrm{T}}$, where $\mathbf{h}_R$ and $\mathbf{h}_I$ respectively represent the real and the imaginary part of $\mathbf{h}$. Although we assume a stochastic channel model, we focus on a standard deterministic estimation approach for the sake of simplicity (our results can however be generalized to a Bayesian estimation approach without difficulty). Therefore we concentrate on the use of the Maximum Likelihood (ML) estimator, defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\tilde{\boldsymbol{\theta}}} \log p(y(0), y(1) \ldots y(N_T-1) \| \tilde{\boldsymbol{\theta}}). \tag{3}$$

The implementation and performance of the above estimate of $\boldsymbol{\theta}$ has been extensively studied in the literature [2, 1, 4, 3]. In the sequel, $\hat{f}$ and $\hat{\mathbf{h}}$ respectively denote the estimates of the frequency offset and the channel.

The aim of the present paper is to propose a method allowing to select training sequence $\mathbf{t}$ in a relevant way. It is natural to search for the training sequences $\mathbf{t}$ which minimize the estimation error. For instance, when the number N_T of training symbols is large, [4] and [2] characterize the training sequences minimizing the (limit of the) mean square error (MSE) on the channel:

$$\mathbb{E}\left[\|\Delta\mathbf{h}\|^2 \;\|\mathbf{t}\right] \tag{4}$$

where $\Delta\mathbf{h} = \hat{\mathbf{h}} - \mathbf{h}$ and, on the otherhand, characterize the training sequences minimizing the MSE on the frequency offset:

$$\mathbb{E}\left[(\Delta f)^2 \;\|\mathbf{t}\right], \tag{5}$$

where $\Delta f = \hat{f} - f$. Here, $\mathbb{E}\left[.\;\|\mathbf{t}\right]$ denotes the conditional expectation w.r.t. $\mathbf{t}$ and $\|.\|$ denotes the Euclidian norm. Unfortunately, it turns out that the training sequences minimizing (4) are completely different from the training sequences minimizing (5). In other words, there is *no* training sequence $\mathbf{t}$ allowing to jointly minimize the estimation error on the channel and on the frequency offset. In order to overcome this problem and to exhibit a single training sequence selection strategy, [1] investigates the minimization of the sum $\mathbb{E}[\|\Delta\mathbf{h}\|^2\|\mathbf{t}] + \mathbb{E}[(\Delta f)^2\|\mathbf{t}]$. However, channel estimation errors will have a different impact, e.g., on the bit error rate, than frequency offset estimation errors. Consequently, it is more reasonable to minimize a weighted sum of the MSE such as $w_h\mathbb{E}[\|\Delta\mathbf{h}\|^2\|\mathbf{t}]+w_f\mathbb{E}[(\Delta f)^2\|\mathbf{t}]$ where w_h and w_f are respectively chosen in accordance with the impact of the channel and the frequency offset estimation errors on the overall system performance e.g., the bit error rate. An even more global approach suggested by [1] is to search for the training sequences $\mathbf{t}$ which minimize the criterion

$$\mathrm{Tr}\left(\mathbf{W}\mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H \;\|\mathbf{t}\right]\right) \tag{6}$$

where $\mathbf{W}$ is a weighting matrix which enables to place different weights on channel estimation errors and frequency offset estimation errors. Unfortunately, the choice of weighting matrix $\mathbf{W}$ is a difficult task. To our knowledge, it has not been addressed in the literature.

In this paper, our goal is to propose a relevant Training Sequence Design (TSD) criterion and to exhibit training sequences minimizing the latter criterion Furthermore, the resulting "optimal training" sequences should not depend on parameter $\boldsymbol{\theta}$, since $\boldsymbol{\theta}$ is unknown at both the transmitter and the receiver. Solutions should only depend on the prior knowledge on $\boldsymbol{\theta}$, namely K, $\boldsymbol{\Sigma}$ and the LOS component $\mathbf{h}_d$.

2. THE PROPOSED CRITERION

Clearly, the selection of a TSD criterion crucially depends on the receiver's architecture. Indeed, each particular receiver may be more or less sensitive to channel estimation errors / frequency offset estimation errors. Therefore, we propose to construct our TSD criterion based on the simple receiver structure depicted at subsection 2.1: First, the receiver compensates for the frequency offset using the estimated value $\hat{f}$ of f. Second, it compensates for the channel distortion using a classical Wiener filter based of the estimated value $\hat{\mathbf{h}}$ of $\mathbf{h}$. Next, we derive a TSD criterion in accordance with the receiver structure of interest. Basically, our TSD strategy consists in searching for the training sequences such that the mean square error on data symbols at the output of the Wiener filter, is minimum. Of course, our "optimal" training sequences will be relevant when the receiver coincides with the one depicted at subsection 2.1. Nevertheless, in the case where a different receiver is used, it is reasonable to believe that the proposed training strategy is still likely to improve the system performance compared to more standard training strategies.

2.1. Receiver structure

We assume that the transmission consists in a training mode during which training sequence $\mathbf{t}$ with length N_T is transmitted, followed by a data mode during which data sequence $\mathbf{d} = [d(0), d(1), \ldots, d(N_D-1)]^T$ with length N_D is transmitted. For the sake of simplicity, $d(n)$ is assumed to be an independent identically distributed (i.i.d.) sequence with variance $\sigma_d^2 = \mathbb{E}[|d(n)|^2]$. Note however that our results can be generalized to the case where $d(n)$ is a colored sequence. Parameter vector $\boldsymbol{\theta}$ is estimated using (3). The receiver first compensates for the value of the frequency offset: it generates signal $y_c(n) = e^{-2\imath\pi\hat{f}n}y(n)$. For each $n \geq N_T + L$,

$$y_c(n) = e^{-2\imath\pi\Delta f n}\sum_{l=0}^{L-1} h(l)d(n - N_T - l) + w(n).$$

Then, a linear equalizer with coefficients $\mathbf{g} = [g(-L_g), \ldots, g(L_g)]$ is used on the received signal. The output equalizer $z(n)$ is defined by

$$z(n) = \sum_{k=-L_g}^{L_g} g(k)y_c(n-k).$$

Finally, a hard detector is used on the equalizer output in order to recover the transmitted data symbols.

2.2. Training Sequence Design criterion

A natural approach would be to exhibit the training strategy which leads to the minimum bit error rate at the detector output. Unfortunately, such a criterion is very difficult to express as a simple function of the training strategy. Here, we propose to minimize the MSE at the output equalizer. In the sequel, we define

$$\mathrm{MSE}(n, \mathbf{t}) = \mathbb{E}\left[|z(n) - d(n - N_T)|^2 \;\|\mathbf{t}\right].$$

It is worth noting that $z(n)$ is a non stationary sequence due to the presence of factor $e^{-2\imath\pi\Delta f n}$. Therefore, the above expression of the MSE depends on index n. It is of course impractical to minimize the MSE for all possible values of n. Here, we propose to minimize the average MSE w.r.t. all data symbols:

$$\mathrm{MSE}(\mathbf{t}) = \frac{1}{N_D}\sum_{n=N_T}^{N_T+N_D-1} \mathrm{MSE}(n, \mathbf{t}). \tag{7}$$

The above criterion depends on the training sequence $\mathbf{t}$ via the estimation errors on parameters $\mathbf{h}$ and f. The objective of the next section is to express $\mathrm{MSE}(\mathbf{t})$ in closed-form.

3. THE CRITERION EVALUATION

In order to express (7) in a more convenient way, the first step is to provide a simple expression of $\text{MSE}(\mathbf{t})$ as a function of the estimation error on parameters $\mathbf{h}$ and f. The second step is to relate the latter estimation errors to the training strategy. The third step is to obtain a simplified version of Eq. (7).

Step 1: MSE as a function of the estimation error After straightforward but tedious algebraic manipulations, we obtain

$$\text{MSE}(\mathbf{t}) = \mathbb{E}_{\mathbf{h}}[\text{MSE}(\mathbf{t}\|\mathbf{h})]$$

where

$$\begin{aligned}\text{MSE}(\mathbf{t}\|\mathbf{h}) = &\ \sigma_d^2 + \sigma^2 \int_0^1 |(g+\Delta g)(\nu)|^2 d\nu \\ &+ \sigma_d^2 \int_0^1 |h(\nu)|^2 |(g+\Delta g)(\nu - \Delta f)|^2 d\nu \\ &- 2\sigma_d^2 \Re\left[\mathcal{S}(\Delta f)\int_0^1 (g+\Delta g)(\nu-\Delta f) h(\nu) d\nu\right]\end{aligned}$$

with

$$\mathcal{S}(\nu) = \frac{1}{N_D}\sum_{n=N_T}^{N_T+N_D-1} e^{-2\imath\pi\nu n}$$

and $\mathbf{g}$ (resp. $\mathbf{g}+\Delta\mathbf{g}$) is the Wiener filter associated with $\mathbf{h}$ (resp. $\mathbf{h}+\Delta\mathbf{h}$). In the above expression, we use the notation $p(\nu) = \sum_{l=-L_1}^{L_2} p(l)e^{-2\imath\pi\nu n}$ for any vector $\mathbf{p} = [p(-L_2), \cdots, p(L_1)]^{\mathrm{T}}$ of length (L_1+L_2+1). The notation $\text{MSE}(\mathbf{t}\|\mathbf{h})$ stands for the MSE given a realization of $\mathbf{h}$. $\Re[.]$ represents the real part of a complex-valued term.

In order to obtain a simple link between $\text{MSE}(\mathbf{t}\|\mathbf{h})$ and the estimation error, we consider the "asymptotic" regime, *i.e.*, we assume that *both* the size N_T of the training sequence and the size N_D of the data sequence tends to infinity, while the ratio N_D/N_T converges to a constant. We assume that

$$\lim_{N_T\to\infty} N_D/N_T = \alpha$$

where α is a constant depending on the system of interest. We recall (cf. [4, 2]) that the MSE of channel estimation is of order $1/N_T$ while the MSE of frequency offset estimation is of order $1/N_T^3$. Consequently, function $\mathcal{S}(f)$ can be decomposed as follows

$$\begin{aligned}\mathcal{S}(\Delta f) &= 1 - \imath\pi(2+\alpha)N_T\Delta f \\ &- 2\pi^2(1+\alpha+\alpha^2/3)N_T^2(\Delta f)^2 + o_p(1/N_T)\end{aligned}$$

where $o_p(1/N_T)$ is negligible w.r.t. $1/N_T$ in probability. Based on the above decomposition, we are able to show that

$$\text{MSE}(\mathbf{t}\|\mathbf{h}) = e_0 + e_1 + e_2 + e_3 + o_p(1/N_T)$$

where

$$\begin{aligned}e_0 &= \sigma_d^2 + \int_0^1 (\sigma_d^2|h(\nu)|^2 + \sigma^2)|g(\nu)|^2 d\nu \\ &- 2\sigma_d^2\Re\left[\int_0^1 h(\nu)g(\nu)d\nu\right] \\ e_1 &= \int_0^1 (\sigma_d^2|h(\nu)|^2+\sigma^2)\gamma_{g,g}(\nu)d\nu \\ e_2 &= 2\sigma_d^2(2+\alpha)\pi N_T \Im\left[\int_0^1 h(\nu)\gamma_{g,f}(\nu)d\nu\right] \\ e_3 &= 4\sigma_d^2\pi^2(1+\alpha+\alpha^2/3)N_T^2\gamma_{f,f}\Re\left[\int_0^1 g(\nu)h(\nu)d\nu\right]\end{aligned}$$

where $\Im[.]$ represents the imaginary part and where

$$\gamma_{g,g}(\nu) = \mathbb{E}[|\Delta g(\nu)|^2] \quad (8)$$

$$\gamma_{g,f}(\nu) = \mathbb{E}[\Delta g(\nu)\Delta f] \quad (9)$$

$$\gamma_{f,f} = \mathbb{E}[(\Delta f)^2]. \quad (10)$$

The term e_0 represents the error due to the Wiener filter based receiver when $\mathbf{h}$ and f are known. Error e_1 (resp. e_3) is the extra term associated with the mis-estimation of $\mathbf{h}$ (resp. f). Finally e_3 is the supplementary error caused by the mis-estimation of both $\mathbf{h}$ and f.

Our aim is now to express the error on the Wiener filter as a function of the error on the channel filter. Under the assumption of an infinite length Wiener filter (*i.e.*, $L_g \to \infty$) , it is easy to check that

$$\Delta g(\nu) = \frac{-\sigma_d^4\overline{h(\nu)}^2\Delta h(\nu) + \sigma_d^2\sigma^2\overline{\Delta h(\nu)}}{(\sigma_d^2|h(\nu)|^2+\sigma^2)^2}$$

Then we have

$$\begin{aligned}\gamma_{g,g}(\nu) &= \frac{(\sigma_d^8|h(\nu)|^4 + \sigma_d^4\sigma^4)\gamma_{h,h}(\nu)}{(\sigma_d^2|h(\nu)|^2+\sigma^2)^4} \\ &- \frac{2\Re\left[\sigma_d^6\sigma^2\overline{h(\nu)}^2\tilde{\gamma}_{h,h}(\nu)\right]}{(\sigma_d^2|h(\nu)|^2+\sigma^2)^4}\end{aligned} \quad (11)$$

$$\gamma_{g,f}(\nu) = \frac{\sigma_d^2\sigma^2\overline{\gamma_{h,f}(\nu)} - \sigma_d^4\overline{h(\nu)}^2\gamma_{h,f}(\nu)}{(\sigma_d^2|h(\nu)|^2+\sigma^2)^2} \quad (12)$$

where

$$\tilde{\gamma}_{h,h}(\nu) = \mathbb{E}[\Delta h(\nu)^2]$$

and where $\gamma_{h,h}(\nu)$ (resp. $\gamma_{h,f}(\nu)$) is defined similarly to Eq. (8) (resp. Eq. (9)).

Step 2: MSE as a function of the TS statistics In the sequel, we express $\gamma_{h,h}(\nu)$, $\gamma_{h,f}(\nu)$ and $\gamma_{f,f}(\nu)$ as a function of the training sequence, actually, as a function of the training sequence statistics. Before going further, we consider the training sequence as a realization of a zero-mean stationary random sequence.

When N_T is large, it is known that (cf. [1])

$$\begin{aligned}
\mathbb{E}[\Delta\mathbf{h}\Delta\mathbf{h}^{\mathrm{H}}] &= \frac{\sigma^2}{N_T}\left(\mathbf{R}_t^{-1} + \frac{3}{2}\frac{\mathbf{h}\mathbf{h}^{\mathrm{H}}}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}}\right) \\
\mathbb{E}[\Delta\mathbf{h}\Delta\mathbf{h}^{\mathrm{T}}] &= \frac{3\sigma^2}{2N_T}\frac{\mathbf{h}\mathbf{h}^{\mathrm{T}}}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}} \\
\mathbb{E}[\Delta\mathbf{h}\Delta f] &= -\imath\frac{3\sigma^2}{2\pi N_T^2}\frac{\mathbf{h}}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}} \\
\mathbb{E}[(\Delta f)^2] &= \frac{3\sigma^2}{2\pi^2 N_T^3}\frac{1}{\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h}}
\end{aligned}$$

where $\mathbf{R}_t$ is the L-dimensional covariance matrix defined by $\{r(k-l)\}_{k,l=0,\cdots,L-1}$ with $r(k-l) = \mathbb{E}[t(n+k)\overline{t(n+l)}]$. If we assume that the sequence $\{r(k); k = 0, \pm 1, \cdots\}$ is absolutely summable, then we can define the following spectrum associated with the training sequence

$$S_{tt}(\nu) = |q(\nu)|^2 = \sum_{k\in\mathbb{Z}} r(k)e^{-2\imath\pi k\nu}$$

It is easy to check that

$$r(k) = \int_0^1 |q(u)|^2 e^{2\imath\pi k u}\mathrm{d}u$$

This implies that

$$\mathbf{h}^{\mathrm{H}}\mathbf{R}_t\mathbf{h} = \int_0^1 |q(u)|^2|h(u)|^2\mathrm{d}u$$

Consequently, we have

$$\begin{aligned}
\tilde{\gamma}_{h,h}(\nu) &= -\frac{3\sigma^2}{2N_T}\frac{(h(\nu))^2}{\int_0^1 |q(u)|^2|h(u)|^2\mathrm{d}u} \\
\gamma_{h,f}(\nu) &= -\imath\frac{3\sigma^2}{2\pi N_T^2}\frac{h(\nu)}{\int_0^1 |q(u)|^2|h(u)|^2\mathrm{d}u} \\
\gamma_{f,f} &= \frac{3\sigma^2}{2\pi^2 N_T^3}\frac{1}{\int_0^1 |q(u)|^2|h(u)|^2\mathrm{d}u}.
\end{aligned}$$

A closed-form expression for $\gamma_{h,h}(\nu)$ is more complicated since we have to handle matrix $\mathbf{R}_t^{-1}$. More precisely, $\gamma_{h,h}(\nu)$ depends on $\mathbf{d}_L(\nu)\mathbf{R}_t^{-1}\mathbf{d}_L(\nu)^{\mathrm{H}}$ where $\mathbf{d}_L(\nu) = [1, \cdots, e^{-2\imath\pi(L-1)\nu}]$. In order to express the latter quantity as a simple function of $q(\nu)$, we notice that, when the channel length L is large enough, $\mathbf{R}_t$ becomes a large Toeplitz matrix for which the inverse $\mathbf{R}_t^{-1}$ can be well approximated by a circulant matrix (cf. [5]) described by its first row

$$\left[\int_0^1 \frac{1}{|q(u)|^2}e^{2\imath\pi k u}\mathrm{d}u\right]_{k=0,\cdots,L-1}$$

Based on this approximation, we obtain

$$\gamma_{h,h}(\nu) = \frac{\sigma^2}{N_T}\left(\frac{L}{|q(\nu)|^2} + \frac{3}{2}\frac{|h(\nu)|^2}{\int_0^1 |q(u)|^2|h(u)|^2\mathrm{d}u}\right)$$

When both N_T and L become large, the MSE finally becomes

$$\mathrm{MSE}(\mathbf{t}) = \mathrm{MSE}_0 + \frac{\sigma_d^2\sigma^2}{N_T}J(q) \tag{13}$$

where MSE_0 is a constant which represents the average MSE that one would have observed if the estimation was perfect, and where the "excess MSE" $J(q)$ has the following form

$$J(q) = \int_0^1 \frac{c_1(\nu)}{|q(\nu)|^2}d\nu + \beta\mathbb{E}\left[\frac{1}{\int_0^1 c_2(\nu)|q(\nu)|^2 d\nu}\right]. \tag{14}$$

with $\beta = 27/2 + 9\alpha + 2\alpha^2$. $c_1(\nu)$ is a deterministic function defined by

$$c_1(\nu) = L\sigma_d^2\mathbb{E}\left[\frac{\sigma_d^4|h(\nu)|^4 + \sigma^4}{(\sigma_d^2|h(\nu)|^2 + \sigma^2)^3}\right]. \tag{15}$$

Finally $c_2(\nu)$ is a random process defined by

$$c_2(\nu) = \frac{|h(\nu)|^2}{\int_0^1 \frac{\sigma_d^2|h(u)|^2}{\sigma_d^2|h(u)|^2+\sigma^2}du}. \tag{16}$$

Our approach consists in selecting the value of the power spectrum $|q(\nu)|^2$ of the training sequence which minimizes the excess MSE $J(q)$. To that end, we should obtain a closed form expression of $J(q)$. In particular, the expectation in the second term of the righthand side of (14) should be derived in closed form. This task is however involved and, due to the lack of space, will only be investigated in an extended version of this paper. Here, instead of directly minimizing (14), we rather minimize a simplier criterion $J_d(q)$ which can be interpreted as an approximation of the initial criterion $J(q)$.

Step 3: Approximated MSE As explained above, we now propose a simplier training sequence design criterion based on (14). Our criterion is based on the following observation. For any continuous functional $\mathcal{F}$ of the channel coefficients $\mathbf{h}$, $\mathbb{E}[\mathcal{F}(\mathbf{h})]$ converges to $\mathcal{F}(\mathbf{h}_d)$ as the Ricean factor K tends to infinity. Following this idea, the proposed simplified criterion consists in replacing each occurence of $h(\nu)$ in (14) with the transfert function $h_d(\nu)$ of the LOS component. The simplified training sequence design criterion is defined by

$$J_d(q) = \int_0^1 \frac{c_1^d(\nu)}{|q(\nu)|^2}d\nu + \beta\frac{1}{\int_0^1 c_2^d(\nu)|q(\nu)|^2 d\nu}. \tag{17}$$

where $c_1^d(\nu)$ is obtained by removing the mathematical expectation and by replacing $h(\nu)$ with $h_d(\nu)$ in Eq. (15), and where $c_2^d(\nu)$ is obtained by replacing $h(\nu)$ with $h_d(\nu)$ in Eq. (16). Of course, the above criterion $J_d(q)$ is likely to be a relevant approximation of $J(q)$ provided that the Ricean factor K is large enough. The training strategy proposed in the forthcoming section is therefore appropriate when K is large. On the otherhand, the minimization of the initial criterion $J(q)$ which is likely to provide better results for moderate values of K, will be investigated in future works.

4. OPTIMAL TRAINING SEQUENCE

It can be shown that the minimization of $J_d(q)$ w.r.t. q reduces to a convex optimization problem. Using Lagrange optimization method, we prove the following result. The proof is omitted due to the lack of space.

Theorem 1 *Criterion J_d defined by Eq. (17) is minimum under power constraint $\int_0^1 |q(\nu)|^2 d\nu \leq P$ for*

$$|q(\nu)|^2 = P \frac{\sqrt{c_1^d(\nu)/(\mu - c_2^d(\nu))}}{\int_0^1 \sqrt{c_1^d(u)/(\mu - c_2^d(u))} du} \quad (18)$$

where μ is such that $\int_0^1 c_2^d(u)\sqrt{c_1^d(u)/(\mu - c_2^d(u))} du = \sqrt{\beta}$.

We now make the following comments.

• In practice, a training sequence with power spectrum (18) can be very simply generated as the output of a digital filter with relevant coefficients excited by a known sequence.

• The generation of a training sequence following (18) can be achieved without any additional computational complexity compared to a traditional (white) training sequence since the above filter has to be evaluated only once by transmission.

• When K is large but cannot be considered as infinite, as mentioned in previous section, we propose to still make use of the proposed training sequence (18). Indeed, simulations show that performance gains can be obtained by using the "colored" training sequence (18) when compared to an uncorrelated training sequence.

5. SIMULATIONS

We consider $N_t = 50$, $L = 5$ and $\alpha = 10$. The carrier frequency offset is fixed to $f = 0.1$. All simulated points are averaged over 100 Monte-Carlo runs for which we have modified the deterministic and random part of the channel at each trial with respect to the zero-mean unit-variance Gaussian distribution.

In Figure 1, we display the theoretical MSE versus SNR (with $K = 5$) when the parameters are perfectly known and when the parameters have to be estimated with either a white training sequence or the suggested training sequence. In Figure 2, we also plot the theoretical MSE versus K (with SNR=15dB). We observe that the gain in terms of MSE exists but is small between the white and colored training sequence.

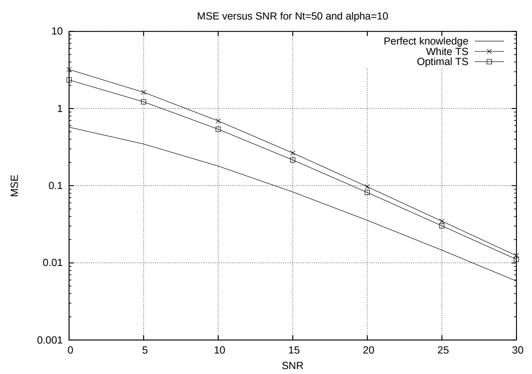

Fig. 1. MSE versus SNR

In Figure 3, we plot the Bit Error Rate (BER) versus SNR (with $K = 5$) when a frequency compensation and a Wiener

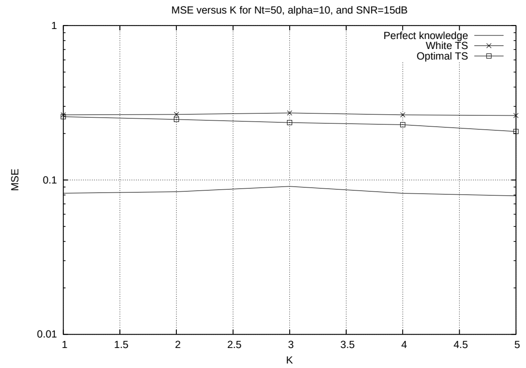

Fig. 2. MSE versus K

equalizer based on the estimated values of the parameters are employed. We remark that the gain in performance is of interest. Moreover we guess that the gain may be more important if we carry out the training sequence design relying on the true criterion (14) instead of on the simplified Eq.(17).

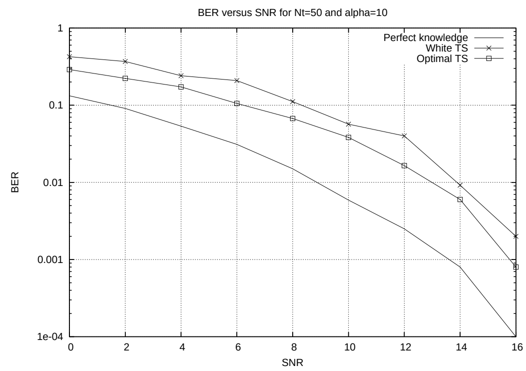

Fig. 3. BER versus SNR

6. REFERENCES

[1] 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.

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

[3] 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.

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

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