Joint Sampling Clock Offset and Channel Estimation for OFDM Signals : Cramér-Rao Bound and Algorithms

Sophie Gault, Walid Hachem, and Philippe Ciblat (contact author)

Abstract

We consider the problem of sampling clock synchronization and channel estimation for Orthogonal Frequency Division Multiplex (OFDM) systems. In such systems, when the number of subcarriers is large, a sampling clock frequency mismatch between the transmitter and the receiver dramatically degrades the performance. So far, the literature proposes ad-hoc estimation algorithms. However, a complete performance analysis, especially the Cramér-Rao Bound (CRB) derivation, remains to be done. Obviously, the channel impulse response is unknown at the receiver and also needs to be estimated. Therefore we evaluate theoretically the Cramér-Rao Bound associated with this joint estimation. When the number of subcarriers and the channel degree are large, very compact closed-form expressions for the CRB are obtained. Furthermore, along with the ML estimator, we introduce sub-optimal estimation algorithms and compare them with some existing approaches and with the CRB.

Keywords

Cramér-Rao bound, OFDM, Sampling clock offset estimation, Channel estimation, VDSL, PLT

Edics

3-PERF, 3-SYNC, 2-PERF

Sophie Gault and Walid Hachem are with Département Télécommunications, Supélec, Gif-sur-Yvette, France (sophie.gault@supelec.fr, walid.hachem@supelec.fr)

Philippe Ciblat is with Département Communications et Electronique, Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France (philippe.ciblat@enst.fr).

I. INTRODUCTION

The detection of Orthogonal Frequency Division Multiplexing (OFDM) symbols cannot be done properly without a reliable clock synchronization. One synchronization step consists in estimating the OFDM symbol timing, which is the delay between the transmitted and the received OFDM symbols. In a certain number of applications where these symbols are short, estimating this delay is enough. However, as soon as the number of samples per OFDM symbol (or equivalently, the number of subcarriers) becomes large, the frequency offset between the transmitter's sampling clock and the receiver's sampling clock in its free oscillation mode has to be considered too. Indeed this offset leads to a sampling delay that drifts linearly in time over the OFDM symbol. Without any compensation, this drift hampers the receiver's performance as soon as the product of the relative clock frequency offset with the number of subcarriers becomes non negligible in comparison with one ([1]). For instance, in Very high speed Digital Subscriber Lines (VDSL) transmissions, these two quantities can respectively reach 10^{-4} and 4096 ([2]), making the clock frequency offset compensation mandatory. As an other example, Power Line Transmissions (PLT) in the band $[1\ \mathrm{MHz}, 20\ \mathrm{MHz}]$ ([3]) show a similar behavior with respect to this phenomenon.

As it is well known, the part of the OFDM symbol that enters the Fast Fourier Transform (FFT) device at the receiver comes after a cyclic prefix. As the latter has a length comparable to the channel impulse response length, it is precisely when the channel is long that a long duration has to be chosen for the useful part of the OFDM symbol, in order to reduce the impact of the cyclic prefix on the spectral efficiency. It is therefore worth considering the problem of the joint estimation of the clock frequency offset and of the channel impulse response, particularly in these situations where the observation window has to be rather large.

The literature proposes several data-aided algorithms (in the sense that one or several OFDM symbols are devoted to training) to perform the estimation of the clock frequency offset ([4], [5], [6], [7], [8], [9]). In some of these approaches, the channel is implicitly assumed perfectly known while in others, the knowledge of the channel is not required to perform the frequency offset estimation. In this paper, in order to better understand the interactions between these two estimations, we begin in section III by giving the Cramér-Rao Bound (CRB) associated with this joint estimation problem. In section IV, we simplify the closed-form expressions of the CRB when the observation window length grows large. It appears that these expressions can be simplified further when the channel degree is large. This is in particular the case of VDSL or the PLT wireline channels, whose degree is often of the order 100. Section V deals with practical estimation algorithms. We begin by the Maximum-Likelihood (ML) estimator for which

we propose a simplified version. Because the ML algorithm remains complicated even in its simplified version, we study simple estimation algorithms that require OFDM training symbols having particular structures. In section VI, the ML algorithm as well as sub-optimal algorithms are tested and compared to the CRBs. Concluding remarks are drawn in Section VII.

In the sequel, $\mathbb{E}$ is the expectation operator and $\mathbb{P}$ is the probability measure. $\mathbf{I}_P$ stands for the $P \times P$ identity matrix and $\mathbf{F}_{P,Q}$ is the $P \times Q$ matrix which element at the p^{th} row and q^{th} column is $\frac{1}{\sqrt{P}} e^{-\frac{2i\pi}{P}pq}$ for $p = 0, \ldots, P-1$ and $q = 0, \ldots, Q-1$. The Kronecker product between matrices is denoted $\otimes$. The argument of a complex-valued scalar is denoted $\angle$.

II. System model

Let us consider the reception of one standard OFDM block which has passed through a non-flat fading channel. After removing the guard interval, the observation window size is $T_0 = NT$ where N is the number of subcarriers, T is the sampling period at the transmitter and $1/T_0$ the spacing between two adjacent subcarriers. Consequently the continuous-time received signal $y_N^{(a)}(t)$ writes as follows :

$$y_N^{(a)}(t) = \sum_{k \in \mathbb{Z}} d_{N,k} g^{(a)}(t - kT) + v^{(a)}(t) \tag{1}$$

where $(d_{N,k})_{k=0,\ldots,N-1}$ represents the output of the Nfold Inverse FFT (IFFT) device of the transmitter. This OFDM symbol is devoted to training and therefore, is assumed to be known at the receiver. As usual, N is a power of 2. The unknown impulse response $g^{(a)}(t)$ represents the complete channel that includes the transmit filter, the propagation channel, and the receiver low-pass filter. Finally $v^{(a)}(t)$ is an additive noise independent of the data.

Because of the oscillators' imperfection, the transmitter's and receiver's clocks are not synchronized. Therefore $y_N^{(a)}(t)$ is sampled at $(1+\delta)T$ instead of T, where δ is an unknown offset lying in the known interval $[-\delta_{\max}, \delta_{\max}]$. The parameter $\delta_{\max}$ is related to the precision of the oscillators used in the transmission chain. The ASDL/VDSL norms [2], for instance, recommend that $\delta_{\max}$ be equal to 10^{-4}. The discrete-time received signal $y_N(n) = y_N^{(a)}(n(1+\delta)T)$ is then written

$$y_N(n) = \sum_{m \in \mathbb{Z}} d_{N,n-m} g^{(a)}(mT + n\delta T) + v(n) \tag{2}$$

where $v(n) = v^{(a)}(n(1+\delta)T)$ is assumed white Gaussian circular with zero-mean and known variance $\sigma^2 = \mathbb{E}[|v(n)|^2]$. As usual, $g^{(a)}(t)$ is assumed time limited with the time support included in $[0, LT)$ where L is a known integer. We thus write $g_l = g^{(a)}(lT)$ for $l = 0, \ldots, L-1$. The Fourier transform

$G^{(a)}(f)$ of $g^{(a)}(t)$ is furthermore assumed to have an effective frequency support included in the interval $[0, 1/T]$. With these assumptions, it is possible to make the useful approximation

$$\forall k \in \{0, \ldots, N-1\}, \sum_m g^{(a)}\left((m+n\delta)T\right) e^{-\frac{2i\pi}{N}mk} = e^{\frac{2i\pi}{N}kn\delta} \sum_{l=0}^{L-1} g_l e^{-\frac{2i\pi}{N}lk} \tag{3}$$

that can be justified by the following argument: k being a chosen integer, let $g^{(a)}_{k,n\delta}(t) = g^{(a)}(t + n\delta T)e^{-\frac{2i\pi}{N}k\frac{t}{T}}$, a function which Fourier Transform is $G^{(a)}\left(f + \frac{k}{NT}\right) e^{2i\pi(f+\frac{k}{NT})n\delta T}$. Using Poisson summation formula, we then have

$$\sum_m g^{(a)}_{k,n\delta}(mT) = \frac{1}{T}\sum_l G^{(a)}\left(\frac{l}{T} + \frac{k}{NT}\right) e^{2i\pi(l+\frac{k}{N})n\delta} .$$

The left hand side of this equation is precisely the left hand side of (3). Moreover, as the effective support of $G^{(a)}(f)$ belongs to $[0, 1/T]$, then the right hand side is $\frac{1}{T}G^{(a)}\left(\frac{k}{NT}\right) e^{\frac{2i\pi}{N}kn\delta}$. By the sampling theorem, this quantity coincides with $e^{\frac{2i\pi}{N}kn\delta}\sum_l g_l e^{-\frac{2i\pi}{N}lk}$, hence equation (3).

As resulting from an inverse FFT operation, the transmitted samples $(d_{N,k})$ write

$$d_{N,k} = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} D_{N,n'} e^{\frac{2i\pi}{N}kn'}$$

where $(D_{N,n'})_{n'=0,\ldots,N-1}$ are the training symbols in the frequency domain, sometimes referred to as the pilot "subcarriers". Plugging this equation and the approximation (3) into (2), the received signal writes

$$y_N(n) = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} \sum_{l=0}^{L-1} D_{N,n'} g_l e^{\frac{2i\pi}{N}n'(n(1+\delta)-l)} + v(n) . \tag{4}$$

Let the superscript $^{\mathrm{T}}$ be the transposition operator. Putting $\mathbf{y}_N = [y_N(0), \ldots, y_N(N-1)]^{\mathrm{T}}$, $\mathbf{v}_N = [v(0), \ldots, v(N-1)]^{\mathrm{T}}$ and $\mathbf{g} = [g_0, \ldots, g_{L-1}]^{\mathrm{T}}$, we can then write

$$\mathbf{y}_N = \mathbf{R}_N(\delta)\mathbf{g} + \mathbf{v}_N \tag{5}$$

where the element (n, l) of the matrix $\mathbf{R}_N(\delta)$ is $[\mathbf{R}_N(\delta)]_{n,l} = \frac{1}{\sqrt{N}}\sum_{n'=0}^{N-1} D_{N,n'} e^{\frac{2i\pi}{N}n'(n(1+\delta)-l)}$ for $n = 0, \ldots, N-1$ and $l = 0, \ldots, L-1$. After removing the guard interval (which duration is assumed greater than LT), the vector output of the FFT device at the receiver is then

$$\mathbf{Y}_N = \mathbf{F}_{N,N}\mathbf{y}_N = \mathbf{F}_{N,N}\mathbf{R}_N(\delta)\mathbf{g} + \mathbf{F}_{N,N}\mathbf{v}_N \tag{6}$$

In short, this equation describes the structure of the OFDM symbol collected at the output of the FFT device during the training phase.

This paper will focus on the estimation issue of the parameter δ and also of the channel. In practice we do not need the knowledge of the channel $\mathbf{g}$ but rather the knowledge of its Fourier transform. Moreover, during the training and data transmission modes, the OFDM symbols are often subjected to a certain frequency mask constraint. Therefore only the knowledge of the frequency response of the channel $\mathbf{g}$ at the FFT frequency n/N (with $n = 0, \ldots, N-1$) weighted by the mask is actually necessary at the receiver. More precisely, let $\{P_{N,n}\}_{n=0,\ldots,N-1}$ be the sequence of positive real numbers representing the mask profile and let $\mathbf{F}_{N,L}$ be the matrix extracted from $\mathbf{F}_{N,N}$ by keeping its first L columns. Assuming that the frequency clock offset is perfectly compensated for during the data phase, it is easy to check that the FFT output vector for an OFDM symbol during the data phase writes $\mathbf{Y}_N^{(\mathrm{data})} = \sqrt{N}\mathbf{D}_N^{(\mathrm{data})}\mathbf{P}_N\mathbf{F}_{N,L}\mathbf{g} + \mathbf{v}_N$ where $\mathbf{D}_N^{(\mathrm{data})}$ is a $N \times N$ diagonal matrix that bears on its diagonal the random information symbols, and $\mathbf{P}_N = \mathrm{diag}\left([P_{N,0}, \ldots, P_{N,N-1}]\right)$. Notice that $\mathbf{F}_{N,L}\mathbf{g}$ stands for the Fourier transform of vector $\mathbf{g}$. Because the receiver will have to compensate for the channel distorsions in the frequency domain during the data transmission phase, an estimate of the vector $\mathbf{h}_N = \mathbf{P}_N\mathbf{F}_{N,L}\mathbf{g}$ should therefore be available at its site. Our purpose is therefore to derive the Cramér-Rao Bound on the column vector $\left[\tilde{\mathbf{h}}_N^{\mathrm{T}}, \delta\right]^{\mathrm{T}}$ where $\tilde{\mathbf{h}}_N = \left[\mathfrak{Re}\left(\mathbf{h}_N^{\mathrm{T}}\right), \mathfrak{Im}\left(\mathbf{h}_N^{\mathrm{T}}\right)\right]^{\mathrm{T}}$, and where $\mathfrak{Re}(.)$ and $\mathfrak{Im}(.)$ denote the real and the imaginary parts respectively.

III. Exact Cramér-Rao Bound

Considering the general model (6), the vector $\mathbf{Y}_N$ is circular Gaussian with the unknown mean $\boldsymbol{\mu}_N = \mathbf{F}_{N,N}\mathbf{R}_N(\delta)\mathbf{g}$ and the known covariance matrix $\sigma^2\mathbf{I}_N$. Additionally, the complex $N \times L$ matrix function $\mathbf{R}_N(\delta)$ is differentiable. Consequently, according to [10], the Fisher Information Matrix (FIM) associated with the parameter vector $\boldsymbol{\theta} = \left[\widetilde{\mathbf{g}}^{\mathrm{T}}, \delta\right]^{\mathrm{T}}$ with $\widetilde{\mathbf{g}} = \left[\mathfrak{Re}\left(\mathbf{g}^{\mathrm{T}}\right), \mathfrak{Im}\left(\mathbf{g}^{\mathrm{T}}\right)\right]^{\mathrm{T}}$ expresses as follows

$$\mathbf{J}_N = \frac{2}{\sigma^2}\mathfrak{Re}\left[\frac{\partial\boldsymbol{\mu}_N^{\mathrm{H}}}{\partial\boldsymbol{\theta}}.\frac{\partial\boldsymbol{\mu}_N}{\partial\boldsymbol{\theta}}\right]$$

where the superscript $^{\mathrm{H}}$ stands for the transpose-conjugate and $\frac{\partial\boldsymbol{\mu}_N}{\partial\boldsymbol{\theta}} = [\frac{\partial\boldsymbol{\mu}_N}{\partial\theta_0}, \cdots, \frac{\partial\boldsymbol{\mu}_N}{\partial\theta_{2(L+1)}}]$. More precisely, as $\boldsymbol{\mu}_N = \mathbf{F}_{N,N}\mathbf{R}_N(\delta)\mathbf{g}$ and $\mathbf{F}_{N,N}^H\mathbf{F}_{N,N} = \mathbf{I}_N$, one can then easily show that

$$\mathbf{J}_N = \frac{2}{\sigma^2}\begin{bmatrix} N\mathfrak{Re}(\mathbf{U}_N) & -N\mathfrak{Im}(\mathbf{U}_N) & N^2\mathfrak{Re}(\mathbf{V}_N\mathbf{g}) \\ N\mathfrak{Im}(\mathbf{U}_N) & N\mathfrak{Re}(\mathbf{U}_N) & N^2\mathfrak{Im}(\mathbf{V}_N\mathbf{g}) \\ N^2\mathfrak{Re}(\mathbf{g}^{\mathrm{H}}\mathbf{V}_N^{\mathrm{H}}) & -N^2\mathfrak{Im}(\mathbf{g}^{\mathrm{H}}\mathbf{V}_N^{\mathrm{H}}) & N^3\mathbf{g}^{\mathrm{H}}\mathbf{W}_N\mathbf{g} \end{bmatrix} \tag{7}$$

where

$$\begin{aligned}\mathbf{U}_N &= \frac{1}{N}\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{R}_N(\delta)\\ \mathbf{V}_N &= \frac{1}{N^2}\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{Q}_N(\delta)\\ \mathbf{W}_N &= \frac{1}{N^3}\mathbf{Q}_N^{\mathrm{H}}(\delta)\mathbf{Q}_N(\delta)\end{aligned}$$

and $\mathbf{Q}_N(\delta) = d\mathbf{R}_N(\delta)/d\delta$. The reason for introducing the factors $1/N$, $1/N^2$, and $1/N^3$ will become apparent later. By applying the well known formulas for the inversion of block partitioned matrices, we obtain :

$$\mathbf{J}_N^{-1} = \left[\begin{array}{cc}\mathbf{A}_N & \mathbf{b}_N\\ \mathbf{b}_N^{\mathrm{T}} & c_N\end{array}\right] \tag{8}$$

where

$$\mathbf{A}_N = \frac{\sigma^2}{2N}\left(\left[\begin{array}{cc}\Re\mathrm{e}(\mathbf{U}_N^{-1}) & -\Im\mathrm{m}(\mathbf{U}_N^{-1})\\ \Im\mathrm{m}(\mathbf{U}_N^{-1}) & \Re\mathrm{e}(\mathbf{U}_N^{-1})\end{array}\right] + \frac{1}{\gamma_N}\left[\begin{array}{c}\Re\mathrm{e}(\boldsymbol{\beta}_N)\\ \Im\mathrm{m}(\boldsymbol{\beta}_N)\end{array}\right]\left[\begin{array}{cc}\Re\mathrm{e}(\boldsymbol{\beta}_N^{\mathrm{T}}) & \Im\mathrm{m}(\boldsymbol{\beta}_N^{\mathrm{T}})\end{array}\right]\right) \tag{9}$$

$$\mathbf{b}_N = -\frac{\sigma^2}{2N^2\gamma_N}\left[\begin{array}{c}\Re\mathrm{e}(\boldsymbol{\beta}_N)\\ \Im\mathrm{m}(\boldsymbol{\beta}_N)\end{array}\right] \tag{10}$$

$$c_N = \frac{\sigma^2}{2N^3\gamma_N} \tag{11}$$

with

$$\boldsymbol{\beta}_N = \mathbf{U}_N^{-1}\mathbf{V}_N\mathbf{g} \tag{12}$$

$$\gamma_N = \mathbf{g}^{\mathrm{H}}\left(\mathbf{W}_N - \mathbf{V}_N^{\mathrm{H}}\mathbf{U}_N^{-1}\mathbf{V}_N\right)\mathbf{g}. \tag{13}$$

As a consequence, we can find the following inequalities

$$\mathbb{E}\left[\|\hat{\mathbf{g}}_N - \mathbf{g}\|^2\right] \geq \mathrm{tr}\left(\mathbf{A}_N\right) = \frac{\sigma^2}{2N}\left(2\mathrm{tr}\left(\mathbf{U}_N^{-1}\right) + \frac{1}{\gamma_N}\|\boldsymbol{\beta}_N\|^2\right) \tag{14}$$

and

$$\mathbb{E}\left[\left(\hat{\delta}_N - \delta\right)^2\right] \geq c_N = \frac{\sigma^2}{2N^3\gamma_N} \tag{15}$$

where $\hat{\mathbf{g}}_N$ and $\hat{\delta}_N$ are estimates of $\mathbf{g}$ and δ obtained from the observation $\mathbf{Y}_N$.

As $\mathbf{h}_N = \mathbf{P}_N\mathbf{F}_{N,L}\mathbf{g}$, the CRB associated with the parameters of interest $\left[\tilde{\mathbf{h}}_N^{\mathrm{T}}, \delta\right]^{\mathrm{T}}$ can be written as follows

$$\left[\begin{array}{cc}\mathbf{CRB}_N^{(\mathbf{h},\mathbf{h})} & \mathbf{CRB}_N^{(\mathbf{h},\delta)}\\ \mathbf{CRB}_N^{(\delta,\mathbf{h})} & \mathrm{CRB}_N^{(\delta,\delta)}\end{array}\right] = \left[\begin{array}{cc}\widetilde{\mathbf{E}}_N\mathbf{A}_N\widetilde{\mathbf{E}}_N^{\mathrm{T}} & \widetilde{\mathbf{E}}_N\mathbf{b}_N\\ \mathbf{b}_N^{\mathrm{T}}\widetilde{\mathbf{E}}_N^{\mathrm{T}} & c_N\end{array}\right]$$

where $\widetilde{\mathbf{E}}_N = \begin{bmatrix} \mathfrak{Re}\left(\mathbf{E}_N\right) & -\mathfrak{Im}\left(\mathbf{E}_N\right) \\ \mathfrak{Im}\left(\mathbf{E}_N\right) & \mathfrak{Re}\left(\mathbf{E}_N\right) \end{bmatrix}$ and $\mathbf{E}_N = \mathbf{P}_N\mathbf{F}_{N,L}$. It yields that

$$\mathbb{E}\left[\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|^2\right] \geq \frac{\sigma^2}{2N}\left(2\mathrm{tr}\left(\mathbf{U}_N^{-1}\mathbf{T}_N\right) + \frac{\boldsymbol{\beta}_N^{\mathrm{H}}\mathbf{T}_N\boldsymbol{\beta}_N}{\gamma_N}\right) \tag{16}$$

where $\hat{\mathbf{h}}_N$ is any unbiased estimate of $\mathbf{h}_N$ and $\mathbf{T}_N = \mathbf{E}_N^{\mathrm{H}}\mathbf{E}_N$.

Expressions (15) and (16) provide the Cramér-Rao lower bound for the parameters of interest. Nevertheless these expressions are difficult to interpret. To overcome this drawback and to obtain simpler expressions, we will analyze the asymptotic behavior of these expressions. By "asymptotic", we mean that a large number of subcarriers, then a long channel impulse response are considered.

IV. Asymptotic Cramér-Rao Bound

By "asymptotic" one frequently means that the number of observed samples or equivalently the number N of subcarriers in our context, grows towards infinity, the channel length L being held fixed. With some assumptions, it will appear that in this regime, matrices $\mathbf{U}_N$, $\mathbf{V}_N$, and $\mathbf{W}_N$ converge elementwise to deterministic matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ respectively. This will be the first part of this section. However, in our particular situation, it is interesting to go further and to assume in a second step that $L \to \infty$. As said in the introduction, this assumption has a practical interest in wireline communications. Notice that we assume $N \to \infty$ *then* $L \to \infty$. In practice, our study will be relevant in situations where N and L are large but $L \ll N$.

As a first step, we thus focus on the asymptotic analysis of the CRB as $N \to \infty$. Most of the training sequence structures encountered in the literature ([8], [9], [11]) can be encompassed within an unique framework by writing the elements of the training sequence $(D_{N,0}, \ldots, D_{N,N-1})$ as follows

$$D_{N,l} = X_{N,l}P(l/N) \tag{17}$$

where $f \mapsto P(f)$ is a bounded real function defined on the interval $[0,1]$ and integrable in the Riemann sense. Note that $P(f)$ refers to the frequency mask constraint. Let Q be any integer. The variable $X_{N,l}$ is a random variable of zero-mean and variance $\mathbb{E}[|X_{N,l}|^2] = Q$ if Q divides l, and $\mathbb{E}[|X_{N,l}|^2] = 0$ otherwise. The random variables $(X_{N,l})_{l=0,\ldots,N-1}$ are furthermore assumed to be independent and their 8^{th} moments exist and are uniformly bounded, *i.e.*, there exists a constant C such that

$$\sup_N \max_{l=0,\ldots,N-1} \mathbb{E}\left[|X_{N,l}|^8\right] < C\,. \tag{18}$$

With this definition of $X_{N,l}$, the energy consumed by the OFDM symbol associated with $(D_{N,l})_{l=0,\ldots,N-1}$ does not depend on Q.

The asymptotic behavior of the CRB is driven by the asymptotic behavior of the matrices $\mathbf{U}_N$, $\mathbf{V}_N$ and $\mathbf{W}_N$ when $N \to \infty$, the channel length L being fixed. The following lemma will help us to do this asymptotic study, since the elements of matrices $\mathbf{U}_N$, $\mathbf{V}_N$ and $\mathbf{W}_N$ can be decomposed according to the expression of ξ_N.

Lemma 1: Let $\alpha > 0$ and let $\phi_N : \mathbb{N} \to \mathbb{C}$ be a function such that for every integer k with $1 \leq |k| \leq \lfloor \alpha N \rfloor$, $|\phi_N(k)| \leq \frac{C}{k}$ where C is a constant that does not depend on N. Then, for every real number r,

$$\xi_N = \frac{1}{N} \sum_{k=1}^{\lfloor \alpha N \rfloor} \sum_{l=k}^{N-1} D_{N,l} D_{N,l-k}^* e^{-\frac{2i\pi}{N} lr} \phi_N(k)$$

converges almost surely to 0 as $N \to \infty$.

Proof: See Appendix A. ■

Let us begin with the evaluation of $\mathbf{U}_N$. The (p,q) element of this matrix for $\{p,q\} \in \{0,\ldots,L-1\}$ writes

$$[\mathbf{U}_N]_{p,q} = \frac{1}{N} \sum_{l_1,l_2=0}^{N-1} D_{N,l_2} D_{N,l_1}^* e^{-\frac{2i\pi}{N}(ql_2 - pl_1)} \psi_N^{(0)}\left((l_1 - l_2)(1+\delta)\right) \tag{19}$$

where

$$\psi_N^{(0)}(x) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-\frac{2i\pi}{N} nx} = \begin{cases} e^{-i\pi\frac{N-1}{N}x} \frac{1}{N} \frac{\sin \pi x}{\sin \frac{\pi x}{N}} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases} \tag{20}$$

Equation (19) can be rewritten

$$[\mathbf{U}_N]_{p,q} = U_{N,0} + U_{N,1} + U_{N,2}$$

where

$$\begin{aligned} U_{N,0} &= \frac{1}{N} \sum_{l=0}^{N-1} |D_{N,l}|^2 e^{-\frac{2i\pi}{N}(q-p)l} \\ U_{N,1} &= \frac{1}{N} \sum_{k=1}^{N-1} \sum_{l=k}^{N-1} D_{N,l} D_{N,l-k}^* e^{-\frac{2i\pi}{N}(q-p)l} e^{-\frac{2i\pi}{N}pk} \psi_N^{(0)}\left(k(1+\delta)\right) \\ U_{N,2} &= \frac{1}{N} \sum_{k=1}^{N-1} \sum_{l=0}^{N-1-k} D_{N,l} D_{N,l+k}^* e^{-\frac{2i\pi}{N}(q-p)l} e^{\frac{2i\pi}{N}pk} \psi_N^{(0)}\left(-k(1+\delta)\right) . \end{aligned}$$

Let us prove that $U_{N,1}$ converges to 0 almost surely. The proof of the convergence of $U_{N,2}$ towards 0 can be done similarly. $U_{N,1}$ can be written

$$U_{N,1} = U_{N,1,1} + U_{N,1,2} + U_{N,1,3}$$

where

$$U_{N,1,1} = \frac{1}{N} \sum_{k=1}^{\lfloor (\frac{N}{2}-1)\frac{1}{1+\delta} \rfloor} S_k \ , \ U_{N,1,2} = \frac{1}{N} \sum_{\lfloor (\frac{N}{2}-1)\frac{1}{1+\delta} \rfloor + 1}^{\lfloor \frac{N-1}{1+\delta} \rfloor} S_k \text{ and } U_{N,1,3} = \frac{1}{N} \sum_{\lfloor \frac{N-1}{1+\delta} \rfloor + 1}^{N-1} S_k$$

and S_k is the inner sum in the expression of $U_{N,1}$. It is easy to check that $|\psi_N^{(0)}(x)| \leq \frac{1}{2|x|}$ for $|x| \leq N/2$. Lemma 1 can therefore be used after identifying the function $\phi_N(k)$ in its statement with $e^{-\frac{2i\pi}{N}pk}\psi_N^{(0)}(k(1+\delta))$, the number r with $q-p$ and α with $\frac{1}{2(1+\delta)}$. By consequence, $U_{N,1,1} \to 0$ a.s. when $N \to \infty$. The terms $U_{N,1,2}$ and $U_{N,1,3}$ can be treated similarly after noticing that $\left|\psi_N^{(0)}(x)\right|$ is periodic with period N.

It leads to

$$[\mathbf{U}_N]_{p,q} - \frac{1}{N} \sum_{l=0}^{N-1} |D_{N,l}|^2 e^{-\frac{2i\pi}{N}(q-p)l} \to 0 \text{ a.s.} \tag{21}$$

According to the definition of the sequence $(D_{N,l})_{l=0,\ldots,N-1}$ introduced in (17), and by using tools similar to those of the proof of lemma 1, it can be seen that

$$\frac{1}{N} \sum_{l=0}^{N-1} |D_{N,l}|^2 e^{-\frac{2i\pi}{N}(q-p)l} - \frac{Q}{N} \sum_{l=0}^{\frac{N}{Q}-1} P\left(\frac{lQ}{N}\right)^2 e^{-\frac{2i\pi Q}{N}(q-p)l} \ \to \ 0 \text{ a.s.} \tag{22}$$

Moreover, as $f \mapsto P(f)$ is assumed Riemann-integrable, we get

$$\lim_{N\to\infty} \frac{Q}{N} \sum_{l=0}^{\frac{N}{Q}-1} P\left(\frac{lQ}{N}\right)^2 e^{-\frac{2i\pi Q}{N}(q-p)l} \ = \ \int_0^1 P(f)^2 \mathbf{d}_L(e^{2i\pi f}) \mathbf{d}_L^{\mathrm{H}}(e^{2i\pi f}) df \tag{23}$$

where $\mathbf{d}_L(e^{2i\pi f}) = [1, \cdots, e^{2i\pi f(L-1)}]^{\mathrm{T}}$.

From (21), (22), and (23), we deduce that the matrix $\mathbf{U}_N$ converges elementwise almost surely towards

$$\mathbf{U} = \int_0^1 P(f)^2 \mathbf{d}_L(e^{2i\pi f}) \mathbf{d}_L^{\mathrm{H}}(e^{2i\pi f}) df \ . \tag{24}$$

One notices that $\mathbf{U}$ coincides with the covariance matrix of a stationary process having $P(f)^2$ as a spectral density.

Let us now consider the asymptotic behavior of $\mathbf{V}_N$. The element (p,q) of this matrix writes

$$[\mathbf{V}_N]_{p,q} = \frac{2i\pi}{N^2} \sum_{l_1,l_2=0}^{N-1} D_{N,l_2} D_{N,l_1}^* e^{-\frac{2i\pi}{N}(ql_2 - pl_1)} l_2 \psi_N^{(1)}((l_1-l_2)(1+\delta))$$

where [12]

$$\psi_N^{(1)}(x) = \frac{1}{N^2} \sum_{n=0}^{N-1} n e^{-\frac{2i\pi}{N}nx} = \begin{cases} \frac{(N-1)e^{-2i\pi x(N+1)/N} - Ne^{-2i\pi x} + e^{-2i\pi x/N}}{N^2(e^{-2i\pi x/N}-1)^2} & \text{if } x \neq 0 \\ (N-1)/2N & \text{if } x = 0 \end{cases} .$$

Once again, it is possible to verify that $|\psi_N^{(1)}(k)| < C/k$ for $1 \leq |k| \leq N/2$. After some computations similar to those of $[\mathbf{U}_N]_{p,q}$ that lead to (21), we obtain

$$[\mathbf{V}_N]_{p,q} - i\pi \frac{1}{N} \sum_{l=0}^{N-1} \frac{l}{N} |D_{N,l}|^2 e^{-\frac{2i\pi}{N}(q-p)l} \rightarrow 0 \text{ a.s.}$$

Therefore the almost sure limit $\mathbf{V}$ of $\mathbf{V}_N$ expresses as

$$\mathbf{V} = i\pi \int_0^1 f P(f)^2 \mathbf{d}_L(e^{2i\pi f}) \mathbf{d}_L^{\mathrm{H}}(e^{2i\pi f}) df \ . \tag{25}$$

Similar derivations lead to the almost sure limit $\mathbf{W}$ of $\mathbf{W}_N$ which expresses as follows

$$\mathbf{W} = \frac{4}{3}\pi^2 \int_0^1 f^2 P(f)^2 \mathbf{d}_L(e^{2i\pi f}) \mathbf{d}_L^{\mathrm{H}}(e^{2i\pi f}) df \ . \tag{26}$$

In order to analyze the right hand side of (16) in the asymptotic regime, we also need to study the asymptotic behavior of $\mathbf{T}_N$ as $N \rightarrow \infty$. In parallel with the model (17) relative to the training sequence symbols, we will assume that diagonal elements of $\mathbf{P}_N$ (which represent the frequency mask for the data transmission phase) satisfy $P_{N,l} = P(l/N)$ for $l = 0, \ldots, N-1$. The element (p,q) of $\mathbf{T}_N$ for $p, q = 0, \ldots, L-1$ writes then

$$[\mathbf{T}_N]_{p,q} = \frac{1}{N} \sum_{l=0}^{N-1} P\left(\frac{l}{N}\right)^2 e^{-\frac{2i\pi}{N}(q-p)l} \ .$$

Therefore, this matrix clearly converges to the Toeplitz matrix $\mathbf{T}$ defined as

$$\mathbf{T} = \lim_{N\rightarrow\infty} \mathbf{E}_N^{\mathrm{H}} \mathbf{E}_N = \int_0^1 P(f)^2 \mathbf{d}_L(e^{2i\pi f}) \mathbf{d}_L^{\mathrm{H}}(e^{2i\pi f}) df \ = \ \mathbf{U} \ . \tag{27}$$

In order to obtain more compact CRB expressions, we put into profit the Toeplitz structure of matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ obtained above and study the asymptotic regime where the channel length L is large (*i.e.*, $L \rightarrow \infty$). The rest of the study will be practically relevant in the situations where N and L are large but $L \ll N$.

In practice the mask profile $P(f)$ is band limited. Indeed, some frequencies are forbidden in order to mitigate the interference with systems operating at adjacent frequencies or with systems using narrow frequency bands within our band of interest. A typical example of such system is the radio amateur system which is known to use frequencies that lie in the interval used by VDSL or PLT systems. Because of this band-limited nature of $P(f)$, the integration in (24), (25) and (26) may be done only over a subset of $[0;1]$ having a Lebesgue measure less than one. Consequently, due to well-known results provided in

[13], these matrices are rank-deficient as soon as L becomes large, that is to say that, they admit some negligible eigenvalues which prevent a standard inversion. By inspecting (7), one notices that the limit of the Fisher Information Matrix is also singular.

We shall neglect the eigenvalues of $\mathbf{U}$ less than a given $\varepsilon > 0$. Let $\Delta_P^{(\varepsilon)} = \{f \in [0,1], P(f)^2 \geq \varepsilon\}$ and define $F_\varepsilon(x)$ as $F_\varepsilon(x) = 0$ if $x < \varepsilon$ and $F_\varepsilon(x) = 1/x$ if $x \geq \varepsilon$. With a small notation abuse, we denote by $\mathbf{U}^\#$ the "pseudo-inverse" matrix [1] $\mathbf{U}^\# = F_\varepsilon(\mathbf{U})$ where ε is chosen small enough so as to retain only the dominant eigenvalues of $\mathbf{U}$.

Using results introduced in [14] related to the CRB with singular Fisher Information Matrices, our CRB analysis remains valid by replacing the matrix $\mathbf{U}^{-1}$ with $\mathbf{U}^\#$ in the CRB expressions for $[\tilde{\mathbf{h}}^{\mathrm{T}}, \delta]^{\mathrm{T}}$. Then, (16) and (15) can be modified as follows

$$\frac{N}{L}\mathbb{E}\left[\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|^2\right] \geq \frac{\sigma^2}{L}\mathrm{tr}\left(\mathbf{U}^\#\mathbf{T}\right) + \mathcal{O}\left(\frac{1}{L}\right) . \tag{28}$$

$$\mathbb{E}\left[\left(\hat{\delta}_N - \delta\right)^2\right] \geq \frac{\sigma^2}{2N^3\mathbf{g}^{\mathrm{H}}\left(\mathbf{W} - \mathbf{V}^{\mathrm{H}}\mathbf{U}^\#\mathbf{V}\right)\mathbf{g}} \tag{29}$$

Using results introduced in [15] concerning the asymptotic behavior of Toeplitz matrices, it can be shown that, when L tends to infinity, (28) and (29) reduce to

$$\frac{N}{L}\mathbb{E}\left[\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|^2\right] \geq \sigma^2|\Delta_P^{(\varepsilon)}| \tag{30}$$

and

$$N^3\mathbb{E}\left[(\hat{\delta}_N - \delta)^2\right] \geq \frac{3\sigma^2}{2\pi^2\int_0^1 f^2P(f)^2|G(e^{2i\pi f})|^2df} , \tag{31}$$

$|\Delta_P^{(\varepsilon)}|$ being the Lebesgue measure of the useful frequency support $\Delta_P^{(\varepsilon)}$ and $G(e^{2i\pi f}) = \sum_{l=0}^{L-1} g_l e^{-2i\pi lf}$. In [16], using different mathematical derivations and a time domain approach, similar results were obtained in the context of single-carrier transmissions.

Equations (30) and (31) call for some observations. First, it is clear that the CRBs over the channel and the clock frequency offset decrease in $\mathcal{O}(1/N)$ and in $\mathcal{O}(1/N^3)$ respectively. Furthermore, the activation of one subcarrier over Q has no effect on the asymptotic CRBs. By inspecting (31), due to the term f^2 in the integral, it also appears that a frequency mask that is too constraining in the high frequencies region can be detrimental to the clock frequency offset estimation. It can also be noticed that in the asymptotic regime, the estimation of δ has no effect on the CRB over the channel. Indeed, if δ were

[1] Let $\mathbf{Q}.\mathrm{diag}([\lambda_0, \ldots, \lambda_L]).\mathbf{Q}^{\mathrm{H}}$ be the eigenvalue decomposition of $\mathbf{U}$ then $F_\varepsilon(\mathbf{U})$ is defined as $F_\varepsilon(\mathbf{U}) = \mathbf{Q}.\mathrm{diag}([F_\varepsilon(\lambda_0), \ldots, F_\varepsilon(\lambda_L)]).\mathbf{Q}^{\mathrm{H}}$

perfectly known, then (16) would be replaced by $\mathbb{E}\left[\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|^2\right] \geq \sigma^2 \mathrm{tr}\left(\mathbf{U}_N^{-1}\mathbf{T}_N\right)/N$. Getting to (28), we only need to remove the term $\mathcal{O}(1/L)$ in its right hand member if we want to suppress the effect of the estimation of δ. Now, if the channel were perfectly known, then γ_N given by (13) would have to be replaced by $\gamma_N = \mathbf{g}^{\mathrm{H}}\mathbf{W}_N\mathbf{g}$, resulting in $N^3\mathbb{E}\left[(\hat{\delta}_N - \delta)^2\right] \geq 3\sigma^2 / \left(8\pi^2 \int_0^1 f^2 P(f)^2 |G(e^{2i\pi f})|^2 df\right)$ in the asymptotic regime. Therefore, by comparing this expression with (31), we notice that the absence of knowledge of the channel impulse response leads to a 6 dB loss on the CRB over the clock frequency offset.

V. Estimation Algorithms

A. ML like Algorithms

Getting back to the received signal model (6) in the frequency domain, the Log-Likelihood function to be minimized $\mathcal{L}(\boldsymbol{\theta})$ is

$$\mathcal{L}(\boldsymbol{\theta}) = \|\mathbf{Y}_N - \mathbf{F}_{N,N}\mathbf{R}_N(\delta)\mathbf{g}\|^2 \ .$$

The minimization of $\mathcal{L}(\boldsymbol{\theta})$ leads to the following ML based estimates of δ and $\mathbf{h}_N$ (see [17],[10] for more explanations) :

$$\left\{\begin{array}{lcl} \hat{\delta}_N & = & \arg\max_\delta \mathbf{Y}_N^{\mathrm{H}}\mathbf{\Pi}_N(\delta)\mathbf{Y}_N \\ \hat{\mathbf{g}}_N & = & \left(\mathbf{R}_N^{\mathrm{H}}(\hat{\delta}_N)\mathbf{R}_N(\hat{\delta}_N)\right)^{-1}\mathbf{R}_N^{\mathrm{H}}(\hat{\delta}_N)\mathbf{F}_{N,N}^{\mathrm{H}}\mathbf{Y}_N \\ \hat{\mathbf{h}}_N & = & \mathbf{E}_N\hat{\mathbf{g}}_N \end{array}\right.$$

where

$$\mathbf{\Pi}_N(\delta) = \mathbf{F}_{N,N}\mathbf{R}_N(\delta)\left(\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{R}_N(\delta)\right)^{-1}\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{F}_{N,N}^{\mathrm{H}}$$

is the orthogonal projection matrix onto the subspace of $\mathbb{C}^N$ spanned by the columns of $\mathbf{F}_{N,N}\mathbf{R}_N(\delta)$. For estimating the sampling clock offset, each try of a value of δ requires the inversion of $\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{R}_N(\delta)$. The implementation of this algorithm is therefore impractical. However, it can be simplified in the asymptotic regime described at the end of the previous section. In this regime, $\mathbf{\Pi}_N(\delta)$ can indeed be replaced with

$$\underline{\mathbf{\Pi}}_N(\delta) = \frac{1}{N}\mathbf{F}_{N,N}\mathbf{R}_N(\delta)\mathbf{U}^{\#}\mathbf{R}_N^{\mathrm{H}}(\delta)\mathbf{F}_{N,N}^{\mathrm{H}} \ .$$

Notice that $\mathbf{U}^{\#}$ is independent of δ and so this matrix is computed only once. Notice also that, because we are only able to consider the significant eigenvalues of $\mathbf{U}$, we can only estimate $G(e^{2i\pi f})$ for $f \in \Delta_P^{(\epsilon)}$. Nevertheless as the parameter of interest is $\mathbf{h}_N$, values of $G(e^{2i\pi f})$ out of this set are not needed, and

therefore the estimate $\hat{\mathbf{h}}_N$ remains accurate. Here, the estimation algorithm becomes

$$\hat{\delta}_N = \arg\max_{\delta} \mathbf{Y}_N^{\mathrm{H}} \underline{\mathbf{\Pi}}_N(\delta) \mathbf{Y}_N \quad (32)$$

$$\hat{\mathbf{h}}_N = \frac{1}{N} \mathbf{E}_N \mathbf{U}^{\#} \mathbf{R}_N^{\mathrm{H}}(\hat{\delta}_N) \mathbf{F}_{N,N}^{\mathrm{H}} \mathbf{Y}_N \quad (33)$$

B. Sub-Optimal Algorithms

The complexity of the ML algorithm presented in the previous subsection prevents its implementation in most practical situations even if one resorts to the simplification (32). It appears that the estimation problem can be largely simplified by endowing the OFDM training symbol with a particular structure. The principle of the approach is the following. Neglecting the additive noise, let us assume that the received sequence $(y_N(0), \ldots, y_N(N-1))$ consists of two identical parts of length $N/2$ each, *i.e.*, $y_N(n) = y_N(n+N/2)$ for $n = 0, \ldots, N/2-1$. This comes down to setting the $N/2$ symbols at the odd subcarriers to zero in the transmitted OFDM symbol, or in other words, the training sequence $(D_{N,0}, \ldots, D_{N,N-1})$ in the frequency domain is asserted to satisfy $D_{N,2l+1} = 0$ for $l = 0, \ldots, N/2-1$. At the receiver side, two consecutive FFTs with length $N/2$ each are performed. If δ were equal to 0, then the outputs of these FFTs would be identical. When $\delta \neq 0$, if we neglect the so-called Inter-Carrier Interference (ICI) created by this mis-synchronization, then the m^{th} output of the second FFT is equal to the m^{th} output of the first FFT rotated by the angle $2\pi m\delta$. The delay δ can thus be estimated from these rotations. With this new model for the training sequence, the asymptotic analysis of section IV remains obviously true as we have simply chosen $Q = 2$ (see (17)).

The idea of transmitting two identical signal halves and exploiting this structure for synchronization is not new. It appeared for the first time in [11] in the context of Doppler shift estimation. Notice that when a Doppler shift $\delta_{\mathrm{Doppler}}/NT$ exists, the m^{th} output of the second FFT is equal to the m^{th} output of the first FFT rotated by a constant angle $\pi\delta_{\mathrm{Doppler}}$ (instead of $2\pi m\delta$ in the Sampling Clock Offset estimation context). Consequently the approach of [11] has to be modified. This modification is reported in subsection V-B.1.

In the context of sampling clock offset estimation, a close idea, that consists in transmitting two whole identical OFDM symbols, has already been exploited in [8] and [9]. A brief description of these other two algorithms will be given at the end of this section.

B.1 An approach, based on a standard structured symbol

Assume that $N \geq 2L$ and let $M = N/2$, $y_{N,1}(n) = y_N(n)$ and $y_{N,2}(n) = y_N(n+M)$ for $n = 0, \ldots, M-1$. We have from (4)

$$\begin{aligned} y_{N,1}(n) &= \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} \frac{D_{N,2m}}{\sqrt{2}} g_l e^{\frac{2i\pi}{M} m(n(1+\delta)-l)} + v(n) \\ y_{N,2}(n) &= \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} \frac{D_{N,2m}}{\sqrt{2}} g_l e^{\frac{2i\pi}{M} m(n(1+\delta)-l)} e^{2i\pi m\delta} + v(n+M)\,. \end{aligned} \tag{34}$$

Let $\mathbf{y}_{N,1} = [y_N(0), \ldots, y_N(M-1)]^{\mathrm{T}}$ and $\mathbf{y}_{N,2} = [y_N(M), \ldots, y_N(N-1)]^{\mathrm{T}}$ be the vectors that represent the two OFDM symbols of size M received successively before the FFT operation. Denote by $\mathbf{Y}_{N,1}$ and $\mathbf{Y}_{N,2}$ the corresponding Fourier transformed vectors, and let $\widetilde{\mathbf{Y}}_N = \left[\mathbf{Y}_{N,1}^{\mathrm{T}}, \mathbf{Y}_{N,2}^{\mathrm{T}}\right]^{\mathrm{T}}$. Let $\mathbf{\Psi}_M(\delta)$ be the $M \times M$ matrix which element (k,m) is $[\mathbf{\Psi}_M(\delta)]_{k,m} = \psi_M^{(0)}(k - m(1+\delta))$ for $k, m = 0, \ldots, M-1$ where $\psi_M^{(0)}(x)$ is given by (20). Finally, let $\mathbf{\Theta}_M(\delta) = \mathrm{diag}\left(\left[1, e^{2i\pi\delta}, \ldots, e^{2i\pi\delta(M-1)}\right]\right)$. From (34), we get after some simple computations

$$\widetilde{\mathbf{Y}}_N = \sqrt{M}\underline{\mathbf{\Psi}}_M(\delta)\underline{\mathbf{\Theta}}_M(\delta)\mathbf{S}_M + \mathbf{V}_N \tag{35}$$

where

$$\underline{\mathbf{\Psi}}_M(\delta) = \begin{bmatrix} \mathbf{\Psi}_M(\delta) & \\ & \mathbf{\Psi}_M(\delta) \end{bmatrix}, \; \underline{\mathbf{\Theta}}_M(\delta) = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{\Theta}_M(\delta) \end{bmatrix},$$

$$\mathbf{S}_M = \frac{1}{\sqrt{2}}\mathbf{D}_M\mathbf{F}_{M,L}\mathbf{g}$$

where $\mathbf{D}_M = \mathrm{diag}([D_{N,0}, D_{N,2}, \ldots, D_{N,M-2}])$, and $\mathbf{V}_N$ represents the Gaussian additive noise term after Fourier transformation. Because $\mathbf{\Psi}_M(0) = \mathbf{\Theta}_M(0) = \mathbf{I}_M$, the vector $\mathbf{S}_M$ would be the output of any of the two FFT operations if we had no noise and if we had $\delta = 0$. When $\delta \neq 0$, an Inter-Carrier Interference term, accounted for by the non-diagonal terms of the matrix $\mathbf{\Psi}_M(\delta)$, appears at the outputs of both FFT operations. Additionally, the second FFT operates on rotated versions of the elements of $\mathbf{S}_M$, element m being rotated by the angle $2\pi m\delta$.

From (35), we notice that the noiseless part of the received signal $\widetilde{\mathbf{Y}}_N$ belongs to the subspace of $\mathbb{C}^N$ generated by the columns of the $N \times M$ matrix $\underline{\mathbf{\Psi}}_M(\delta)\underline{\mathbf{\Theta}}_M(\delta)$. It is therefore possible to look for the estimate $\hat{\delta}$ that maximizes the norm of the projection of $\widetilde{\mathbf{Y}}_N$ over this subspace, in other words,

$$\hat{\delta} = \arg\max_{\delta} \tilde{\mathbf{Y}}_N^{\mathrm{H}} \underline{\mathbf{\Psi}}_M(\delta)\underline{\mathbf{\Theta}}_M(\delta) \left(\underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\delta)\underline{\mathbf{\Psi}}_M^{\mathrm{H}}(\delta)\underline{\mathbf{\Psi}}(\delta)\underline{\mathbf{\Theta}}(\delta)\right)^{-1} \underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\delta)\underline{\mathbf{\Psi}}_M^{\mathrm{H}}(\delta)\tilde{\mathbf{Y}}_N\,.$$

To gain in simplicity, we approximate $\mathbf{\Psi}_M^{\mathrm{H}}(\delta)$ by a diagonal matrix such as $\mathbf{\Psi}_M^{\mathrm{H}}(\delta)\mathbf{\Psi}_M(\delta) = \mathbf{I}_M$, thus neglecting the ICI term. In this case, after some calculations, the last equation reduces to

$$\hat{\delta} = \arg\max_{\delta} \left(\sum_{l=0}^{M-1} \mathfrak{Re}\left(Y_{2,l} Y_{1,l}^* e^{-2i\pi l\delta} \right) \right) \tag{36}$$

where we have written $\mathbf{Y}_{N,1} = [Y_{1,0}, \ldots, Y_{1,M-1}]^{\mathrm{T}}$ and $\mathbf{Y}_{N,2} = [Y_{2,0}, \ldots, Y_{2,M-1}]^{\mathrm{T}}$. Notice that, in the context of Doppler shift estimation, the previous equation is simpler since $e^{-2i\pi l\delta}$ has to be replaced with $e^{-i\pi\delta_{\mathrm{Doppler}}}$ as in [11]. In practice, denoting by $\chi(\delta)$ the term to be maximized in (36), its derivative with respect to δ writes

$$\frac{d\chi}{d\delta} = \sum_{l=0}^{M-1} 2\pi l \left(\mathfrak{Im}\left(Y_{2,l} Y_{1,l}^* \cos 2\pi l\delta \right) - \mathfrak{Re}\left(Y_{2,l} Y_{1,l}^* \sin 2\pi l\delta \right) \right)$$

By cancelling out this derivative, and by using the approximations $\cos 2\pi l\delta \approx 1$ et $\sin 2\pi l\delta \approx 2\pi l\delta$ which are valid for the most common values of $M\delta/2$, we obtain

$$\hat{\delta} = \frac{\sum_{l=0}^{M-1} l\mathfrak{Im}\left(Y_{2,l} Y_{1,l}^* \right)}{2\pi \sum_{l=0}^{M-1} l^2 \mathfrak{Re}\left(Y_{2,l} Y_{1,l}^* \right)} . \tag{37}$$

Let us now turn to the estimation of the channel $\mathbf{h}_N$. Merging (35) and (17) leads to the following model

$$\tilde{\mathbf{Y}}_N = \sqrt{\frac{M}{2}} \underline{\mathbf{\Psi}}_M(\delta) \underline{\mathbf{\Theta}}_M(\delta) \underline{\mathbf{X}}_M \underline{\mathbf{E}}_M \mathbf{g} + \mathbf{V}_N \tag{38}$$

where $\underline{\mathbf{X}}_M = \mathrm{diag}([X_{N,0}, X_{N,2}, \ldots, X_{N,N-2}])$ and $\underline{\mathbf{E}}_M = \underline{\mathbf{P}}_M \mathbf{F}_{M,L}$ with $\underline{\mathbf{P}}_M = \mathrm{diag}([P(0), P(2/N), \ldots, P((N-2)/N)])$. The Least-Square (LS) estimate of the channel writes

$$\hat{\mathbf{g}} = \sqrt{\frac{2}{M}} \left(\underline{\mathbf{E}}_M^{\mathrm{H}} \underline{\mathbf{X}}_M^{\mathrm{H}} \underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\hat{\delta}) \underline{\mathbf{\Psi}}_M^{\mathrm{H}}(\hat{\delta}) \underline{\mathbf{\Psi}}_M(\hat{\delta}) \underline{\mathbf{\Theta}}_M(\hat{\delta}) \underline{\mathbf{X}}_M \underline{\mathbf{E}}_M \right)^{-1} \underline{\mathbf{E}}_M^{\mathrm{H}} \underline{\mathbf{X}}_M^{\mathrm{H}} \underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\hat{\delta}) \underline{\mathbf{\Psi}}_M^{\mathrm{H}}(\hat{\delta}) \tilde{\mathbf{Y}}_N . \tag{39}$$

The inversion operation in this equation increases dramatically the implementation complexity. For this reason the simpler estimate

$$\hat{\mathbf{g}} = \frac{1}{2\sqrt{2M}} \underline{\mathbf{E}}_M^{\mathrm{H}} \underline{\mathbf{X}}_M^{\mathrm{H}} \underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\hat{\delta}) \underline{\mathbf{\Psi}}^{\mathrm{H}}(\hat{\delta}) \tilde{\mathbf{Y}}_N \tag{40}$$

can be used instead. Notice that (40) refers to the standard and simple correlation estimator. We can further simplify (40) by neglecting the ICI represented by the non-diagonal terms and by the slight attenuation on the diagonal terms in the matrix $\underline{\mathbf{\Psi}}_M(\hat{\delta})$. This simplification results in

$$\hat{\mathbf{g}} = \frac{1}{2\sqrt{2M}} \underline{\mathbf{E}}_M^{\mathrm{H}} \underline{\mathbf{X}}_M^{\mathrm{H}} \underline{\mathbf{\Theta}}_M^{\mathrm{H}}(\hat{\delta}) \underline{\mathbf{\Xi}}_M^{\mathrm{H}}(\hat{\delta}) \tilde{\mathbf{Y}}_N \tag{41}$$

where $\underline{\mathbf{\Xi}}_M(\delta) = \mathbf{I}_2 \otimes \mathbf{\Xi}_M(\delta)$ with $\mathbf{\Xi}_M(\delta) = \mathrm{diag}([1, e^{\frac{i\pi(M-1)}{M}\delta}, \cdots, e^{\frac{i\pi(M-1)^2}{M}\delta}])$.

Even if we work in the frequency domain (*i.e.*, after the FFT), we do estimate first the temporal impulse response of the channel $\mathbf{g}$ (41) and not directly the frequency response of the channel $\mathbf{h}_N$ (42). Indeed a frequency estimation of the channel, subcarrier by subcarrier, would not use the coherence bandwidth of the channel and thus would not profit from the induced correlation between adjacent subcarriers. Moreover, if we plan to use 2048 subcarriers (as done in VDSL), we would need to estimate about 2000 parameters instead of only about 100 for the temporal impulse response.
Once this is done, the channel response which is needed in frequency domain for further operations such as equalization, can then be easily obtained through

$$\hat{\mathbf{h}}_N = \mathbf{P}_N \mathbf{F}_{N,L} \hat{\mathbf{g}} \, . \tag{42}$$

B.2 Other methods for the sampling clock offset estimation

Most of the algorithms met in the sampling clock offset estimation literature are based on the phase comparison between two known OFDM symbols. Nevertheless, with minor modifications, these algorithms can be adapted to the situation where one transmitted OFDM symbol consists of two identical halves. In this case, the received signal is described by (34).

Liu's algorithm [8]

Let

$$\Phi_N(k) = \angle Y_{2,k} - \angle Y_{1,k} = \angle (Y_{2,k} Y_{1,k}^*).$$

One can remark that

$$\Phi_N(k) = 2\pi k \delta + \nu_k \tag{43}$$

where ν_k refers to a noise which vanishes in absence of the additive noise $v(n)$ and when the ICI is neglected. By applying on (43) a least square approach, one can obtain, as done in [8], the following estimate for δ

$$\hat{\delta} = \frac{2\pi \sum_{k=0}^{N/2-1} k \Phi_N(k)}{\left(4\pi^2 \sum_{k=0}^{N/2-1} k^2\right)} = \frac{2\pi \sum_{k=0}^{N/2-1} k \angle (Y_{2,k} Y_{1,k}^*)}{\left(4\pi^2 \sum_{k=0}^{N/2-1} k^2\right)} \tag{44}$$

Speth's algorithm [9]

Let $\mathcal{C}_1 = \{0, \cdots, N/4-1\}$ and $\mathcal{C}_2 = \{N/4, \cdots, N/2-1\}$. Define ϕ_1 and ϕ_2 as

$$\phi_{1|2} = \angle \Big(\sum_{k \in \mathcal{C}_{1|2}} Y_{2,k} Y_{1,k}^* \Big) \, .$$

It can be shown that, without additive noise and ICI, δ is equal to the quantity $2(\phi_2 - \phi_1)/\pi N$. Therefore, [9] suggests the following estimate

$$\hat{\delta} = \frac{2}{\pi N}(\phi_2 - \phi_1) \; . \tag{45}$$

VI. SIMULATIONS

We consider a powerline OFDM system operating within the band $[1 \text{ MHz}, 20 \text{ MHz}]$. The magnitude of the channel transfer function used in this section is represented on Fig.1. The corresponding channel impulse response sampled at 20 MHz is made up of 80 complex coefficients. Via computer simulations, we compare the performance of the sub-optimal methods introduced in section V with existing methods as well as with the Cramér-Rao Bounds.

In Fig. 2, Mean Squares Errors (MSE) of the estimates of δ and $\mathbf{h}_N$ are plotted versus the length N of the known OFDM symbol. Here N varies from 256 to 4096, the Signal-to-Noise Ratio (SNR) is fixed to 20dB, and δ is equal to 7.10^{-5} (*i.e.*, 70 ppm). The MSE are averaged over 500 trials. At each trial, the training sequence made of QPSK symbols is different.

The figures show that the performance of the ML is very close to the CRB. Concerning the estimation of δ, the estimator (37) offers good performance. Its performance is even close to that of the ML estimator until $N = 2048$. Recall that the estimator (37) does not take into account the ICI. Unfortunately, from $N = 2048$, the ICI can not be neglected anymore, therefore the performance of the estimator (37) reaches a floor. We also notice that this estimator provides a better performance than each of the estimators (44) introduced in [8] and (45) introduced in [9]. Concerning the channel estimation, the LS estimator given by (42) and (39) provides a remarkable performance. On the contrary, the correlator estimator given by (42) and (41) shows bad performance as soon as N is greater than 512. Thus the correlator estimator is strongly sensitive to the presence of the ICI. Finally we remark that the asymptotic CRB fits well the exact CRB whatever is the value of N.

In Fig. 3, the MSE and asymptotic CRB are displayed versus N, for two different values of δ (10 ppm and 70 ppm) and two different values of the SNR (10 dB and 20 dB). Recall that the asymptotic Cramér-Rao Bounds do not depend on δ.

We notice that, for small values of N, the estimator (37) shows a MSE that does not depend much on the value of δ. However, when N is greater than 2048, then the ICI effect becomes important and affects the performance of this estimator. As for the channel estimation, the ICI effect occurs at smaller values of N and dramatically affects the performance. Actually, the ICI produces an effective noise that quickly dominates the additive noise because its variance grows with N and with δ. For this reason, the MSE of

the channel estimator at a SNR of 20 dB meets the MSE at 10 dB as N grows. The value of N where these two MSEs become equal decreases with δ.

The last figure (Fig. 4) shows the performance vs. the SNR, N being fixed to 2048 and δ to 70 ppm. The estimator (37) of δ shows a MSE close to the CRB in these conditions, and outperforms the estimators (44) and (45). The channel estimator given by (42) and (39) is also close to the CRB. Concerning the channel estimator by correlation, as the SNR grows, the Gaussian noise becomes negligible in comparison with the ICI, which results in a performance floor.

VII. Conclusion

In this paper, we considered the issue of joint sampling clock offset estimation and channel estimation for OFDM modulation used in wireline transmissions. The estimation performance has been studied through the CRB analysis. Simple expressions for the CRB have been obtained in the asymptotic regime, *i.e.*, when the number of subcarriers and the channel length are large. The ML joint algorithm has been derived. Sub-optimal approaches have also been proposed and compared to the existing literature.

Appendix

I. Proof of Lemma 1

For proving lemma 1, we show that $\mathbb{E}\left[|\xi_N|^4\right]$ decreases rapidly enough so that almost sure convergence is guaranteed by the Borel-Cantelli lemma. In this proof, C designates a constant that can change through the equations.

We have

$$\mathbb{E}\left[|\xi_N|^4\right] \leq \frac{C}{N^4} \sum_{k_1,k_2,k_3,k_4} \sum_{l_1,l_2,l_3,l_4} \mathbb{E}\left[|D_{N,l_1} D_{N,l_1-k_1} D_{N,l_2} D_{N,l_2-k_2} D_{N,l_3} D_{N,l_3-k_3} D_{N,l_4} D_{N,l_4-k_4}|\right] \frac{1}{k_1 k_2 k_3 k_4} \ . \quad (46)$$

Because of the independence of the random variables $\{D_{N,l}\}$ and the fact that they are zero mean, the expectation term in the right hand side member of this inequality is zero if there exists in its argument at least one term that appears only once. The situations where every term appears at least twice are described by a finite number of systems of four equations in the indices $k_1, \ldots, k_4, l_1, \ldots, l_4$. One such system is for instance $l_1 = l_2, l_3 = l_4, k_1 = k_2, k_3 = k_4$. Let

$$I_N = \frac{C}{N^4} \sum_{k,k'} \sum_{l,l'} \mathbb{E}\left[|D_{N,l} D_{N,l-k}|^2 \left|D_{N,l'} D_{N,l'-k'}\right|^2\right] \frac{1}{k^2} \frac{1}{k'^2}$$

be the corresponding term in the RHS member of (46). We have

$$\begin{aligned} I_N &\leq \left(\frac{C}{N^2} \sum_{k=1}^{\lfloor \alpha N \rfloor} \sum_{l=k}^{N-1} \mathbb{E}\left[|D_{N,l} D_{N,l-k}|^4 \right]^{1/2} \frac{1}{k^2} \right)^2 \\ &\leq \left(\frac{C}{N^2} \sum_{k=1}^{\lfloor \alpha N \rfloor} \sum_{l=k}^{N-1} \mathbb{E}\left[|D_{N,l}|^8 \right]^{1/4} \mathbb{E}\left[|D_{N,l-k}|^8 \right]^{1/4} \frac{1}{k^2} \right)^2 \\ &\leq \left(\frac{C}{N} \sum_{k=1}^{\lfloor \alpha N \rfloor} \frac{1}{k^2} \right)^2 \\ &\leq \frac{C}{N^2} \end{aligned}$$

where the first two inequalities result from Cauchy-Schwartz inequality and the third one is deduced from condition (18). Similar results can be obtained for the other systems of equations in the indices, leading finally to the inequality $\mathbb{E}\left[|\xi_N|^4 \right] \leq C/N^2$. Markov's inequality implies that $\forall\, \epsilon > 0,\ \mathbb{P}\left(|\xi_N| > \epsilon \right) \leq \frac{\mathbb{E}\left(|\xi_N|^4 \right)}{\epsilon^4} = O(N^{-2})$. Therefore, $\sum_{N=1}^{\infty} \mathbb{P}\left(|\xi_N| > \epsilon \right) < \infty$ and the result follows from the Borel-Cantelli lemma.

REFERENCES

[1] T. Pollet, P. Spruyt, and M. Moeneclaey, "The BER Performance of OFDM Systems Using Non-Synchronized Sampling," in *Proc. GLOBECOM*, 1994, pp. 253–257.

[2] ETSI, *Transmission and Multiplexing (TM); Access transmission systems on metallic access cables; Very high speed Digital Subscriber Line (VDSL); Part 2: Transceiver specification*, 2002–2002, TS 101 270-2.

[3] F.J. Cañete, J.A. Cortés, L. Díez, and J.T. Entrambasaguas, "Modeling and Evaluation of the Indoor Power Line Transmission Medium," *IEEE Comm. Mag.*, vol. 41, no. 4, pp. 41–47, Apr. 2003.

[4] K. Bucket and M. Moeneclaey, "Tracking performance of feedback timing synchronizer operating on interpolated signals," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 1996, pp. 67 –71.

[5] B. Yang, K. Letaief, R. Cheng, and Z. Cao, "An Improved Combined Symbol and Sampling Synchronization Method for OFDM Systems," in *Proc. WCNC*, 1999, vol. 3, pp. 1153–1157.

[6] B. Yang, Z. Ma, and Z. Cao, "ML-Oriented DA Sampling Clock Synchronization for OFDM Systems," in *Proc. WCC/ICCT*, 2000, pp. 781–784.

[7] R. Heaton, S. Duncan, and B. Hodson, "A fine frequency and fine sample clock estimation technique for OFDM systems," in *IEEE Vehicular Technology Conference (VTC)*, 2001, pp. 678–682.

[8] S.Y. Liu and J.W. Chong, "A Study of Joint Tracking Algorithms of Carrier Frequency Offset and Sampling Clock Offset for OFDM-Based WLANs," in *IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions*, 2002, vol. 1, pp. 109 –113.

[9] M. Speth, S. Fechtel, G. Fock, and H. Meyr, "Optimal Receiver Design for OFDM based Broadband Transmission. Part II : A Case Study," *IEEE Trans. on Communications*, vol. 49, no. 4, pp. 571–578, Apr. 2001.

[10] O. Besson and P. Stoica, "Training sequence selection for frequency offset estimation in frequency selective channels," *Digital Signal Processing*, vol. 13, pp. 106–127, 2003.

[11] T.M. Schmidl and D.C. Cox, "Robust Frequency and Timing Synchronization for OFDM," *IEEE Trans. on Communications*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.

[12] L.S. Gradsteyn and I.M. Ryzhik, *Table of integrals, series, and products*, Academic Press, 2000.

[13] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty," *Bell System Technical Journal*, vol. 57, no. 5, May 1978.

[14] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. on SP*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

[15] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, Univ. of California Press, Berkeley, 1958.

[16] S. Gault, W. Hachem, and P. Ciblat, "Cramér-Rao Bounds for data-aided sampling clock offset and channel estimation," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal (Canada), May 2004.

[17] M. Morelli and U. Mengali, "Carrier frequency estimation for transmissions over selective channels," *IEEE Trans. on Communications*, vol. 48, no. 9, pp. 1580–1589, Sept. 2000.

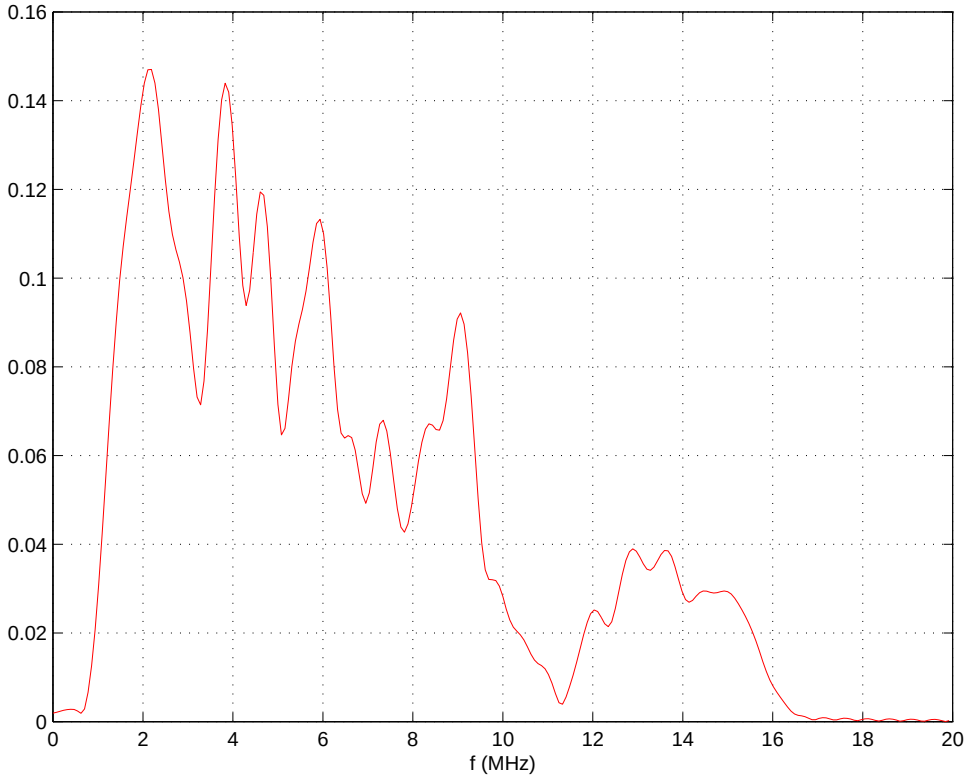

Fig. 1. Magnitude of the channel transfer function

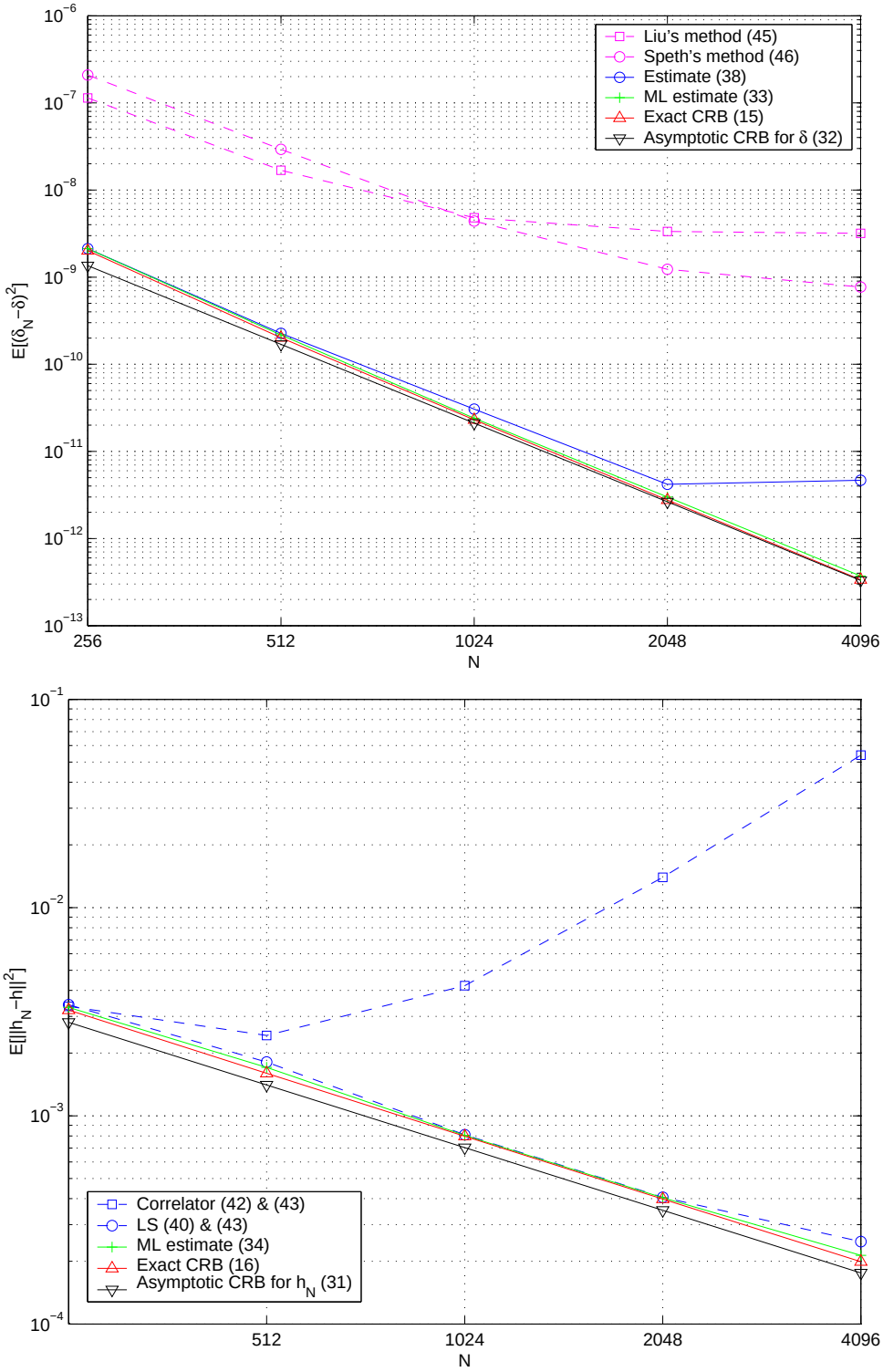

Fig. 2. MSE and CRB for δ (top) and $\mathbf{h}_N$ (bottom) vs. N ($\delta = 7.10^{-5}$)

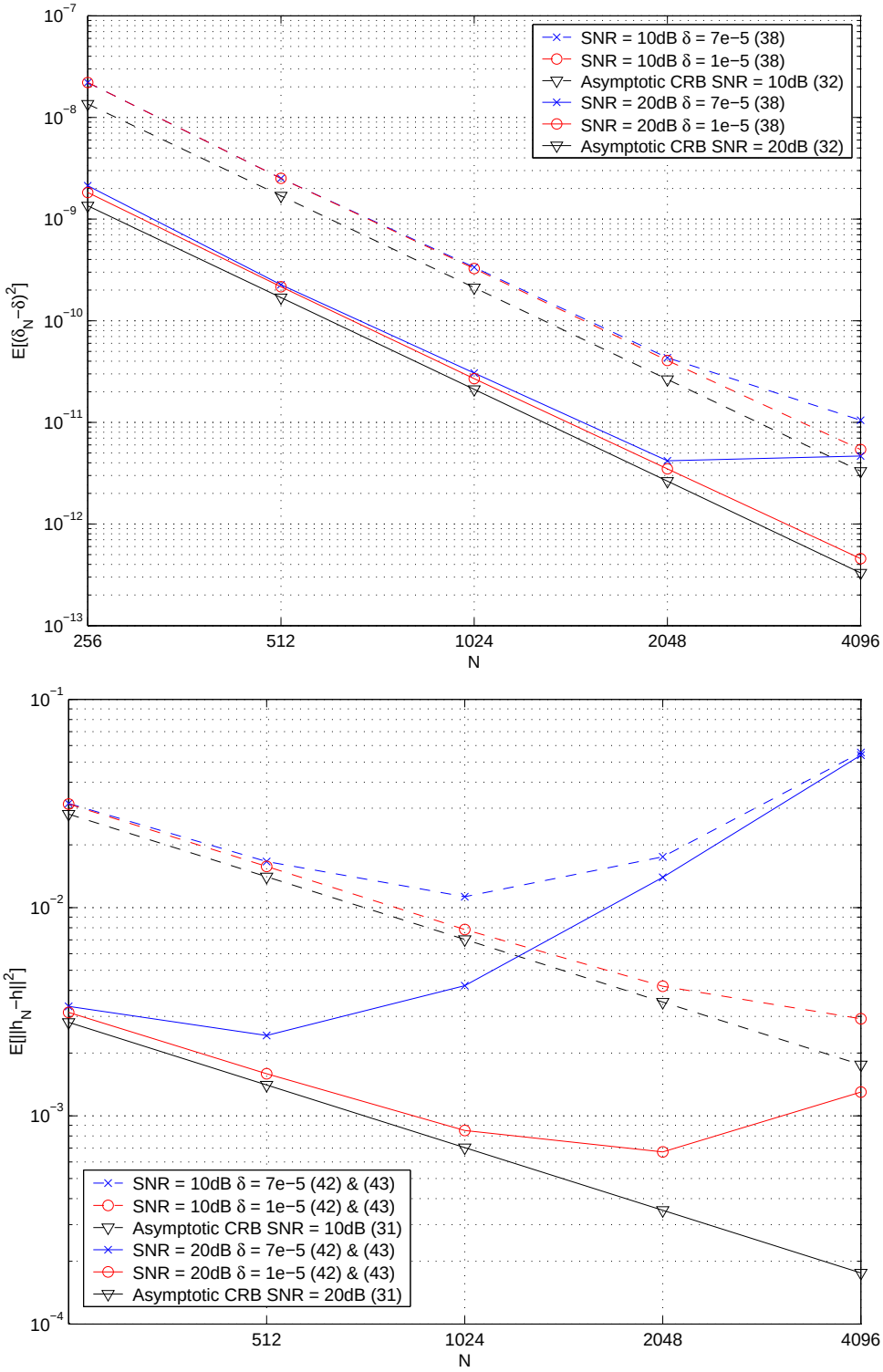

Fig. 3. MSE and CRB vs. N for different values of δ and SNR

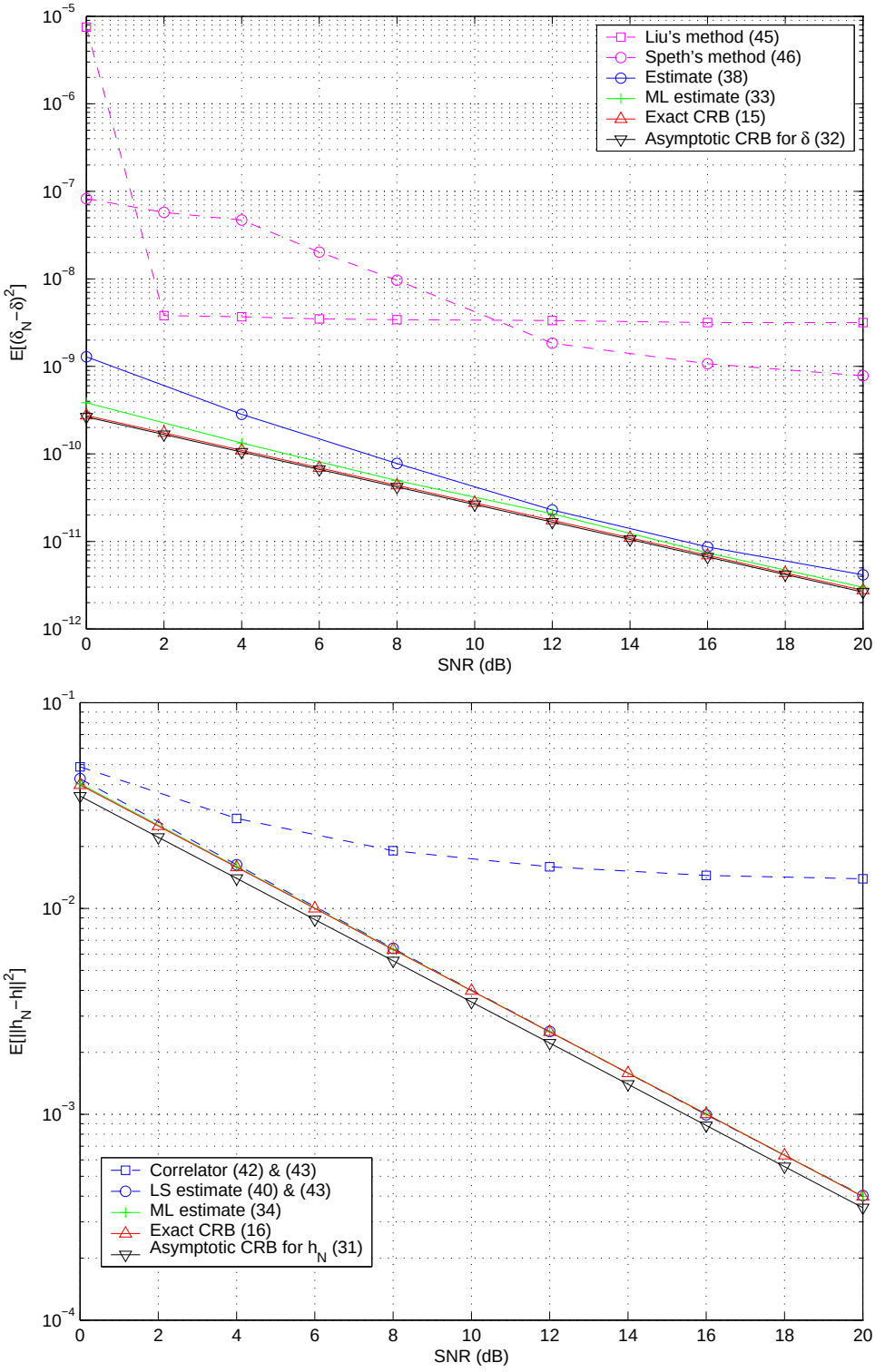

Fig. 4. MSE and CRB vs. SNR ($\delta = 7.10^{-5}$)