# Near MAP Dynamical Delay Estimator and Bayesian CRB for Coded QAM Signals

Imen Nasr, Leïla Najjar Atallah, Sofiane Cherif, Benoit Geller

# Near MAP Dynamical Delay Estimator and Bayesian CRB for Coded QAM Signals

Imen Nasr [1,2], Leïla Najjar Atallah [1], Sofiane Cherif [1] and Benoît Geller [2]

[1] *Ecole Supérieure des Communications de Tunis SUP'COM, University of Carthage, Tunisia*

[2] *Ecole Nationale Supérieure de Techniques Avancées ENSTA ParisTech, Université de Paris Saclay, France*

[1] {nasr.imen,leila.najjar,sofiane.cherif}@supcom.rnu.tn

[2] {imen.nasr,benoit.geller}@ensta-paristech.fr

***Abstract*—This paper presents an off-line algorithm for dynamical time delay recovery for which the whole observation block is used. The time offset varies over the observation interval following a random walk model. The proposed synchronizer applies to data-aided (DA), non-data-aided (NDA) and code-aided (CA) modes. Theoretical performance of the off-line technique is derived and compared to simulation results. The Bayesian Cramer-Rao Bound (BCRB) is also evaluated for DA, NDA and CA modes and for both the off-line and on-line scenarios. Simulation results show the improvement brought by the off-line and the CA schemes. The presented algorithm outperforms the conventional on-line estimator, which only takes into account the current and previous observations, and its MSE approaches the BCRB.**

## I. Introduction

Time synchronization is performed at the receiver front end by adjusting the receiver clock to the transmitter clock in order to mitigate inter-symbol interference (ISI). To do so, one can implement a classical on-line timing error detector (TED) such as the Mueller and Muller Detector (MMD) [1], the Gardner Detector (GD) [2], the Zero Crossing Detector (ZCD) [3] and the Early Late Detector (ELD) [4], for which, the estimated time delay only depends on the previous observations. These TEDs are low cost algorithms and were previously applied to estimate a constant time delay. However, they have limited tracking performance especially if the true delay is time varying, e.g. time delay in a high mobility wireless device. In this case, some ameliorations can be brought by implementing off-line algorithms which operate on the whole observation block taking into account both past and future observations. Such smoothing technique was applied in [5] for phase recovery in an iterative way. In fact, the authors have proposed a MAP estimator for a phase offset following a Brownian model with a linear drift in the case of a non-coded BPSK signal.

In this paper, we derive from the Maximum A Posteriori (MAP) estimation theory an off-line time delay recovery algorithm in the case of a possibly coded QAM signal. Moreover, the performance analysis in terms of MSE and bias differs from the phase offset recovery performance. Another point is that timing synchronizers were mainly developed for DA and NDA scenarios. In the DA mode, pilot signals are inserted into the transmitted frame to aid the time synchronization process. Despite the fact that the best synchronizer performance is achieved with the DA mode, this technique leads to both loss of spectral efficiency and increase of the power budget due to the transmission of pilot signals in the data frame. To deal with this problem, NDA techniques perform synchronization from the sole use of the received signal. It consists of using hard decisions based on the received signal instead of the pilot signals, or in only exploring received signal samples shifted by a half-symbol-period in the case of the Gardner TED. This is why, the system performance becomes unsatisfactory especially in poor channel conditions [6, Chapter 7]. To find a compromise between the spectral efficiency and the synchronizer reliability, CA techniques have been proposed. These techniques take advantage of the decoder soft output to reduce the estimator error in the timing recovery [7]–[15] as well as in the carrier frequency and the phase synchronization process [16], [17]. In [7], [8], the authors deal with the time synchronization problem for a constant time delay and they suggest to use a CA TED derived from the Maximum Likelihood (ML) estimator. The theoretical performance analysis is rather difficult for CA estimation techniques and has rarely been evaluated in the literature so that CA timing recovery techniques are often evaluated by simulations. A comparison between different timing techniques in terms of BER has been proposed in [14] using the extrinsic information transfer chart (EXIT chart) [18]. Nevertheless, some statistical properties of the estimators such as the bias and the mean square error (MSE) cannot be provided by the EXIT chart analysis. The authors of [12] have derived semi-analytical expressions of the estimator mean and variance, as function of the timing offset, for a CA decision-directed (DD) timing synchronizer based on MMD. However, the performance evaluation has been made only at low SNR regime, based on the assumption that inter-symbol interference (ISI) could be approximated by an additive Gaussian noise as in [9]. To evaluate one estimator relevance, its MSE is traditionaly compared to lower bounds such as the Cramer-Rao Bounds (CRB) [19] as the one derived for the unknown random phase offset problem in [16], [20]. Closed form expressions of the CRB have been derived in [21] for CA carrier frequency and phase offset estimation for turbo-coded Square-QAM modulated signals. However, to the best of our knowledge, for the timing recovery problem, the CRB has been evaluated for DA [22], NDA [23] and CA [7], [8] only for a constant delay.

In this paper, an off-line dynamic timing estimator is proposed and its MSE theoretical performance is evaluated as well as the corresponding Bayesian CRB (BCRB). The proposed scheme performance is compared to that of the on-line estimator in DA, CA and NDA modes. We also evaluate the DA, NDA and CA BCRB for dynamical time delay recovery in the case of BPSK and QAM modulated signals and we show the relevance of combining the CA mode and the off-line approach.

This paper is organized as follows. In section II, the system model and the off-line time delay estimation based on the MAP approach are presented. In section III, the derivation of the MSE relative to the off-line technique is given for various scenarios. Then, section IV develops closed form expressions of BCRBs for DA, NDA and CA modes. Simulation results are provided in section V and validate our analysis. The last section concludes our work.

## GLOSSARY OF PRINCIPAL NOTATIONS

- $\Re\{z\}$ and $\Im\{z\}$: the real and imaginary parts of $z$
- $z^*$: the conjugate of $z$
- $\dot{f}(t)$ (resp. $\ddot{f}(t)$): the first (resp. the second) derivative of $f$, for any function $f$ with respect to $t$
- $\otimes$: the convolution operation
- $P(X)$: the probability mass function (pmf) of any discrete random variable $X$
- $p(x)$: the probability density function (pdf) of any continuous random variable $x$
- $E(X)$: the mean value of the random variable $X$
- $\nabla_{\boldsymbol{\tau}} f(\boldsymbol{\tau}) = \left(\frac{\partial f(\boldsymbol{\tau})}{\partial \tau_1} \quad \frac{\partial f(\boldsymbol{\tau})}{\partial \tau_2} \quad \cdots \quad \frac{\partial f(\boldsymbol{\tau})}{\partial \tau_N}\right)^T$: the gradient operator with respect to the vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_N]^T$
- $\Delta_{\boldsymbol{\tau}} f(\boldsymbol{\tau}) = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_1^2} & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_1 \partial \tau_2} & \cdots & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_1 \partial \tau_N} \\ \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_2 \partial \tau_1} & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_2 \partial \tau_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_N \partial \tau_1} & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_N \partial \tau_2} & \cdots & \frac{\partial^2 f(\boldsymbol{\tau})}{\partial \tau_N^2} \end{pmatrix}$: the Hessian matrix with respect to the vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_N]^T$

## II. Off-Line Forward-Backward Timing Recovery

Let us consider the linearly modulated transmitted signal $s(t)$ written as:

$$s(t) = \sum_i a_i h(t - iT), \tag{1}$$

where $a_i$ denotes the zero mean independent transmitted symbols drawn from a given finite size constellation, $h(t)$ is the impulse response of the transmission filter and $T$ is the symbol period.

The received signal is:

$$r(t) = s(t - \tau(t)) + n(t), \tag{2}$$

where $n(t)$ is a zero mean additive white Gaussian noise (AWGN) with a known variance $\sigma_n^2$ and $\tau(t)$ is a time varying delay.

We consider a sampled version of the previously defined continuous signals in equations (1) and (2):

$$r_k = s_k(\tau_k) + n_k, \tag{3}$$

where $r_k$, $\tau_k$, $s_k(\tau_k)$ and $n_k$ are the $k^{th}$ sample of respectively the received signal, the time delay, the transmitted signal $s(t-\tau(t))$ and the noise $n(t)$ taken at time $kT/F$, where $F$ is the oversampling factor. We assume that the delay $\tau_k$ follows a brownian evolution model [10], [11] according to:

$$\tau_k = \tau_{k-1} + w_k, \tag{4}$$

where $w_k$ is a non stationary zero mean AWGN with a known variance $\sigma_w^2$ which determines the severity of the timing jitter [10].

### A. Time Delay Maximum A Posteriori Estimator

Let us define $\boldsymbol{a}$ the vector of the transmitted symbols and $\mathbf{r}$ and $\boldsymbol{\tau}$ the vectors containing, respectively, the $N$ random observations $r_k$ and the $N$ delays $\tau_k$ as $\boldsymbol{r} = [r_1, r_2, \cdots, r_N]$ and $\boldsymbol{\tau} = [\tau_1, \tau_2, \cdots, \tau_N]$. The estimate in the MAP sense of $\boldsymbol{\tau}$ is the vector $\mathbf{u}$ with the highest probability given the observations samples. The time delay estimator is then written as:

$$\begin{aligned} \hat{\boldsymbol{\tau}} &= \arg\max_{\mathbf{u}} \{p(\mathbf{u}|\mathbf{r}, \mathbf{a})\} \\ &= \arg\max_{\mathbf{u}} \{p(\mathbf{r}|\mathbf{u}, \mathbf{a}) p(\mathbf{u})\} \\ &= \arg\max_{\mathbf{u}} \{\log(p(\mathbf{r}|\mathbf{u}, \mathbf{a})) + \log(p(\mathbf{u}))\}, \end{aligned} \tag{5}$$

where $\hat{\boldsymbol{\tau}} = [\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_N]$ and $\mathbf{u} = [u_1, u_2, \cdots, u_N]$.
For given $\boldsymbol{u}$ and $\boldsymbol{a}$, the independence of model (2) samples leads to:

$$p(\mathbf{r}|\mathbf{u}, \mathbf{a}) = \prod_{i=1}^{N} p(r_i|u_i, \mathbf{a}),$$

where:

$$p(r_i|u_i, \mathbf{a}) = \frac{1}{2\pi\sigma_n^2} \exp\left(-\frac{|r_i - s_i(u_i)|^2}{2\sigma_n^2}\right). \tag{6}$$

Thus:

$$p(\mathbf{r}|\mathbf{u}, \mathbf{a}) = \left(\frac{1}{2\pi\sigma_n^2}\right)^N \exp\left(-\frac{\sum_{i=1}^{N} |r_i - s_i(u_i)|^2}{2\sigma_n^2}\right). \tag{7}$$

By letting the observation period, $T_0$, of the signal $r(t)$ cover a large number $N$ of symbol periods, we can assume that [6, Chapter 7]:

$$\sum_{i=1}^{N} |r_i - s_i(u_i)|^2 \simeq \int_{T_0} |r(t) - s(t - u(t))|^2 dt \tag{8}$$

$$= \int_{T_0} |r(t)|^2 dt + \int_{T_0} |s(t - u(t))|^2 dt - 2\int_{T_0} \Re\{r(t) s(t - u(t))^*\} dt. \tag{9}$$

For sufficiently large $T_0$, the delay $u(t)$ can be neglected in the following integral [24, Chapter 5] [25, Chapter 6]:

$$\int_{T_0} |s(t - u(t))|^2 dt \simeq \int_{T_0} |s(t)|^2 dt = \sum_i |a_i|^2. \tag{10}$$

After some basic mathematical derivations and considering that $\tau(t) = \tau_i$ and $u(t) = u_i$ during the $i^{th}$ symbol period $T$, we obtain [6, Chapter 7]:

$$\int_{T_0} \Re\left\{r(t)s(t-u(t))^*\right\} dt \simeq \sum_i \Re\left\{a_i^* x_i(u_i)\right\}, \quad (11)$$

where $x_i(u_i)$ is the received signal after matched filtering:

$$\begin{aligned} x_i(u_i) &= y_i(u_i) + v_i(u_i), & (12)\\ y_i(u_i) &= \sum_l a_l g\left((i-l)T - (\tau_i - u_i)\right), & (13)\\ v_i(u_i) &= \int_{T_0} h(t - iT - u_i) n(t) dt, & (14)\\ g(t) &= h(t) \otimes h^*(-t). & (15) \end{aligned}$$

We choose a root Nyquist filter $h(t)$ such that the global filter $g(t)$ given by (15) is a Nyquist filter. Observing that $\int_{T_0} |r(t)|^2 dt$ does not depend on $u$, it can be dropped from the likelihood function. Having a finite observation samples number $N$, the maximization problem in (5) involves:

$$p(\mathbf{r}|\mathbf{u},\mathbf{a}) = \left(\frac{C}{2\pi\sigma_n^2}\right)^N \exp\left(\sum_{i=1}^N \left(\frac{\Re\left\{a_i^* x_i(u_i)\right\}}{\sigma_n^2} - \frac{|a_i|^2}{2\sigma_n^2}\right)\right), \quad (16)$$

where $C$ is a constant term. From the delay brownian model (4) we obtain:

$$\begin{aligned} p(\mathbf{u}) &= p(u_1) \prod_{i=2}^N p(u_i|u_{i-1}) & (17)\\ &= p(u_1) \prod_{i=2}^N \frac{1}{\sigma_w\sqrt{2\pi}} \exp\left(-\frac{(u_i - u_{i-1})^2}{2\sigma_w^2}\right) & (18) \end{aligned}$$

Thus, without loss of generality and by eliminating the constant terms, the maximization problem (5) can be written as:

$$\hat{\boldsymbol{\tau}} = \arg\max_{\mathbf{u}} \left\{\Lambda_L(\mathbf{a},\mathbf{u})\right\}, \quad (19)$$

where:

$$\Lambda_L(\mathbf{a},\mathbf{u}) = \frac{1}{\sigma_n^2}\sum_{i=1}^N \Re\left\{a_i^* x_i(u_i)\right\} + \log\left(p(u_1)\right) - \sum_{i=2}^N \frac{(u_i - u_{i-1})^2}{2\sigma_w^2}. \quad (20)$$

If we assume that the initial timing error is uniformly distributed between $-T/2$ and $T/2$, then the gradient of $\Lambda_L(\mathbf{a},\mathbf{u})$ with respect to $\mathbf{u}$ is given by:

$$\nabla_u \Lambda_L(\mathbf{a},\mathbf{u}) = \begin{cases} \Re\left\{\frac{a_k^*}{\sigma_n^2}\frac{\partial x_k(u_k)}{\partial u_k}\right\} + \frac{u_{k+1} - u_k}{\sigma_w^2}, \; k = 1; \\ \Re\left\{\frac{a_k^*}{\sigma_n^2}\frac{\partial x_k(u_k)}{\partial u_k}\right\} + \frac{u_{k+1} - 2u_k + u_{k-1}}{\sigma_w^2}, \; k = 2, \cdots, N-1; \\ \Re\left\{\frac{a_k^*}{\sigma_n^2}\frac{\partial x_k(u_k)}{\partial u_k}\right\} + \frac{u_{k-1} - u_k}{\sigma_w^2}, \; k = N. \end{cases}$$

The steepest descent gradient algorithm is often used to solve non linear equations. The estimated time delay $\hat{\boldsymbol{\tau}}$ that maximizes $\Lambda_L(\mathbf{a},\mathbf{u})$ with respect to $\mathbf{u}$, involving $\nabla_u \Lambda_L(\mathbf{a},\mathbf{u}) = 0$, is then given by:

$$\begin{cases} \hat{\tau}_1 = \hat{\tau}_2 + \frac{\sigma_w^2}{\sigma_n^2}\Re\left\{a_1^*\frac{\partial x_1(\hat{\tau}_1)}{\partial \hat{\tau}_1}\right\}; \\ \hat{\tau}_k = \frac{\hat{\tau}_{k+1} + \hat{\tau}_{k-1}}{2} + \frac{1}{2}\frac{\sigma_w^2}{\sigma_n^2}\Re\left\{a_k^*\frac{\partial x_k(\hat{\tau}_k)}{\partial \hat{\tau}_k}\right\}, \; k = 2, \cdots, N-1; \\ \hat{\tau}_N = \hat{\tau}_{N-1} + \frac{\sigma_w^2}{\sigma_n^2}\Re\left\{a_N^*\frac{\partial x_N(\hat{\tau}_N)}{\partial \hat{\tau}_N}\right\}. \end{cases} \quad (21)$$

The above adaptation equations of the time error detector involve the evaluation of the matched filter output derivative $\frac{\partial x_i(\hat{\tau}_i)}{\partial \hat{\tau}_i}$ at the estimated time delays values. This can be obtained by interpolation of an up-sampled version of the signal.

### *B. Off-line Delay Estimation Smoothing Algorithm*

According to the previous subsection, the estimated time delay at time instant $k$ is a function of the time delay at the sample $k-1$ and at the sample $k+1$ which is actually unknown for an on-line filtering procedure. The idea is then to estimate the time delay $\tau_k$ with a smoothing off-line procedure.

Let us define:

$$\begin{aligned} z_k(\hat{\tau}_i, a_k) &\triangleq \mu\Re\left\{a_k^*\frac{\partial x_k(\hat{\tau}_i)}{\partial \hat{\tau}_i}\right\}, & (22)\\ \hat{\tau}_k^{(F)} &\triangleq \hat{\tau}_{k-1}^{(F)} + z_k(\hat{\tau}_{k-1}^{(F)}, a_k), & (23)\\ \hat{\tau}_k^{(B)} &\triangleq \hat{\tau}_{k+1}^{(B)} + z_k(\hat{\tau}_{k+1}^{(B)}, a_k), & (24) \end{aligned}$$

where $z_k(\hat{\tau}_i, a_k)$ is the updating term, $\mu$ is the step size and $i$ is in $\{k-1, k+1\}$. We recognize in (23) the classical recursive timing recovery loop [1], [3], [4], [6]. Equation (24) is similar to (23) but proceeds terms posterior to $k$, the time index of relevance. From (21) and the previous expressions, we propose the following Forward-Backward (FB) algorithm which combines the classical forward (F) timing loop (23) with a backward (B) timing loop (24) proceeding similarly in the opposite direction from the end of the block ($k = N$) towards its beginning ($k = 1$):

$$\hat{\tau}_k^{(FB)} = \begin{cases} \hat{\tau}_2^{(B)} + z_1(\hat{\tau}_2^{(B)}, a_1), \text{ for } k = 1; \\ \frac{1}{2}\left(\hat{\tau}_k^{(B)} + \hat{\tau}_k^{(F)}\right), \text{ for } k = 2, \cdots, N-1; \\ \hat{\tau}_{N-1}^{(F)} + z_N(\hat{\tau}_{N-1}^{(F)}, a_N), \text{ for } k = N. \end{cases} \quad (25)$$

In an off-line context, the estimation of $\tau_k$, $k = 1, \cdots, N$, can even be improved by applying multiple iterations of the forward and backward loops using the same observation block. The term "off-line" means that the timing recovery is performed after the reception of the whole block of signal samples at the receiving end. Such approach can be considered for applications with low constraints on the processing time (latency), as any iterative algorithm. This is achieved by initializing the first forward estimate of the next iteration at the value of the last backward estimate of the previous iteration. It is thus possible to carry out several iterations (in practice 2) and by averaging the values taken by the forward loop and the backward loop at the instant $k$, one obtains the forward-backward estimate.

We note that the updating term $z_k(\hat{\tau}_i, a_k)$ depends on the transmitted symbols. These symbols can be known by

the receiver when some pilot signals are inserted within the transmitted frame. In this case, the proposed algorithms (on-line or off-line) operate in a DA mode. However, in order to enhance the spectral efficiency and increase the data rate, these reference signals can be removed. Then, in practice, hard estimates $\hat{a}_k$ of the transmitted symbols are computed at the receiver and replace $a_k$ in (25); this is called the NDA mode. In a coded system, we can take advantage of the decoder soft output to provide smoother and more accurate estimate of $a_k$ as it was proposed in [7], [8] for the on-line timing loop. We can similarly replace $a_k$ by the soft symbol $\tilde{a}_k$ in (25) for the smoothing off-line procedure and obtain an off-line CA mode. In section V we will firstly assess the improvement obtained by the off-line approach in the DA, NDA and CA modes and then compare the improvement to that obtained using only the CA mode (in an on-line scheme).

Hereafter, at each time index, the analytical performance after one and multiple iterations of the forward, the backward and the forward-backward algorithms are derived in terms of estimation bias and MSE.

## III. Performance Analysis of the forward-backward algorithm

We now approximate at any time index $k$, the updating term of each algorithm as the sum of a linear function of the estimation error, $\hat{\tau}_i - \tau_k$ and an additive self noise, $b_k(\hat{\tau}_i)$, where $i = k-1$ (resp. $i = k+1$) when the forward (resp. backward) algorithm is used. Let us consider $g_{m,n}(u) = g(mT-(\tau_n-u))$ where $\tau_n$ is the true delay at the $n^{th}$ sample. It is worthy to note that $g_{m,n}(\tau_n) = g(mT)$, $\dot{g}_{m,n}(\tau_n) = \dot{g}(mT)$ and $\ddot{g}_{m,n}(\tau_n) = \ddot{g}(mT)$.

Based on equations (22) and (12)-(15), we have:

$$\begin{aligned} z_k(\hat{\tau}_i, a_k) &= \mu\Re\left\{a_k^* \frac{\partial x_k(\hat{\tau}_i)}{\partial \hat{\tau}_i}\right\} \\ &= \mu\Re\left\{\sum_j a_k^* a_j \dot{g}_{k-j,k}(\hat{\tau}_i) + \mu a_k^* \frac{\partial v_k(u)}{\partial u}|_{u=\hat{\tau}_i}\right\} \\ &= S(\hat{\tau}_i - \tau_k) + \mu b_k(\hat{\tau}_i), \end{aligned} \quad (26)$$

where

$$\begin{aligned} S(\hat{\tau}_i - \tau_k) &= E\left[z_k(\hat{\tau}_i, a_k)\right] \\ &= \mu \sum_j \Re\left\{E[a_k^* a_j]\right\} \dot{g}_{k-j,k}(\hat{\tau}_i), \end{aligned}$$

and $b_k(\hat{\tau}_i)$ is the loop noise. Given that the transmitted symbols, $a_k$, are uncorrelated and assuming that $\hat{\tau}_i - \tau_k$ is close to 0, we find:

$$S(\hat{\tau}_i - \tau_k) = \mu E[|a_k|^2]\dot{g}(\hat{\tau}_i - \tau_k) \simeq \mu E[|a_k|^2]\ddot{g}(0)(\hat{\tau}_i - \tau_k). \quad (27)$$

The loop noise $b_k(\hat{\tau}_i)$ is:

$$b_k(\hat{\tau}_i) = \Re\left\{\sum_{j,j\neq k} a_k^* a_j \dot{g}_{k-j,k}(\hat{\tau}_i) + a_k^* \frac{\partial v_k(u)}{\partial u}|_{u=\hat{\tau}_i}\right\}. \quad (28)$$

Its variance is given by:

$$\sigma_b^2 = E\left[\sum_{j,j\neq k} |a_k|^2 |a_j|^2 \dot{g}^2_{k-j,k}(\hat{\tau}_i)\right] + E[|a_k|^2]\sigma_n^2,$$

where for a Nyquist filter, we can neglect $\dot{g}(t)$ values for arguments $|t| \geq 3T$ [6].

Using the following notations:

$$\gamma = \mu E[|a_k|^2]\ddot{g}(0) \text{ and } q_k = \mu b_k(\hat{\tau}_i), \quad (29)$$

and based on equations (26) and (27), we obtain:

$$z_k(\hat{\tau}_i, a_k) \simeq \gamma(\hat{\tau}_i - \tau_k) + q_k. \quad (30)$$

In what follows, we note $\sigma_q^2 = \mu^2\sigma_b^2$ the variance of $q_k$. From this linear expression, we obtain in the sequel, the expression of the estimation bias and of the MSE for the forward, the backward and the forward-backward algorithms after any arbitrary number of iterations, starting with the first iteration.

From Appendix A, the bias of the estimated time delay for only one forward loop is then given by:

$$B_k^{(F)} = (1+\gamma)^{k-1} B_1^{(F)}, \quad (31)$$

and the corresponding MSE is:

$$\begin{aligned} M_k^{(F)} &= E\left[\left(\tau_k - \hat{\tau}_k^{(F)}\right)^2\right] \quad (32) \\ &= (1+\gamma)^{2k-2} M_1^{(F)} + \frac{1-[(1+\gamma)^2]^{k-1}}{1-(1+\gamma)^2}\left[(1+\gamma)^2\sigma_w^2 + \sigma_q^2\right]. \end{aligned}$$

Similarly to the forward loop, we obtain for the backward algorithm:

$$\begin{aligned} \tau_k - \hat{\tau}_k^{(B)} &= (1+\gamma)^{N-k}(\tau_N - \hat{\tau}_N^{(B)}) - \sum_{i=k+1}^{N} w_i (1+\gamma)^{i-k} \\ &- \sum_{i=k}^{N-1} q_i (1+\gamma)^{i-k}. \end{aligned} \quad (33)$$

Thus the bias of the estimated time delay using one backward iteration is:

$$B_k^{(B)} = (1+\gamma)^{N-k} B_N^{(B)}, \quad (34)$$

and the corresponding MSE is:

$$M_k^{(B)} = (1+\gamma)^{2N-2k} M_N^{(B)} + \frac{1-[(1+\gamma)^2]^{N-k}}{1-(1+\gamma)^2}\left[(1+\gamma)^2\sigma_w^2 + \sigma_q^2\right]. \quad (35)$$

We also know that at the end of the observation block,

$$\tau_N - \hat{\tau}_N^{(B)} = \tau_N - \hat{\tau}_N^{(F)}, \quad (36)$$

so that from (31):

$$B_N^{(B)} = (1+\gamma)^{N-1} B_1^{(F)}, \quad (37)$$

and from (32):

$$M_N^{(B)} = (1+\gamma)^{2N-2} M_1^{(F)} + \frac{1-[(1+\gamma)^2]^{N-1}}{1-(1+\gamma)^2}\left[(1+\gamma)^2\sigma_w^2 + \sigma_q^2\right]. \quad (38)$$

From the expression of $\tau_k - \hat{\tau}_k^{(FB)}$ given in (25) for $k = 2, .., N-1$, and using (31) and (37), the bias of the estimated delay using a forward/backward algorithm is given by:

$$B_k^{(FB)} = \frac{1}{2}\left[(1+\gamma)^{k-1} + (1+\gamma)^{2N-k-1}\right] B_1^{(F)}, \quad (39)$$

and the MSE expression after one iteration is given by (40). The derivation details are provided the Appendix B.

$$M_k^{(FB)}=\frac{1}{4}M_k^{(B)}+\frac{1}{4}M_k^{(F)}+\frac{1}{2}\left[(1+\gamma)^{2N-2}M_1^{(F)}+(1+\gamma)^{2N-2k-4}\left((1+\gamma)^2\sigma_w^2+\sigma_q^2\right)\left(\frac{1-(1+\gamma)^{2k+2}}{1-(1+\gamma)^2}\right)+\sigma_q^2\right] \quad (40)$$

We now derive the analytical expressions of the bias, $B_k^m$, and the MSE, $M_k^m$, for respectively the forward, the backward and the forward-backward algorithms, after $m$ iterations. Let us denote by $\hat{\tau}_k^{(I),m}$, the estimate of $\tau_k$ at iteration $m$ using the algorithm $I$ where $I$ is in $\{F, B, FB\}$ and respectively denotes Forward, Backward and Forward-Backward.

We have shown in Appendix C that, at iteration $m$, the forward bias is:

$$B_k^{(F),m}=(1+\gamma)^{(2N-2)\times(m-1)+k-1}B_1^{(F),1}, \quad (41)$$

and the corresponding mean square error is given by (42). where:

$$\Omega(m)=\frac{1-(1+\gamma)^{(2N-2)\times(m-1)}}{1-(1+\gamma)^{2N-2}}. \quad (43)$$

We have also shown in Appendix C that at iteration $m$ and for the $k^{th}$ observation the backward bias is:

$$B_k^{(B),m}=(1+\gamma)^{(2N-2)\times(m-1)+2N-k-1}B_1^{(F),1}, \quad (44)$$

and the corresponding MSE expression is given by (45).

Based on (25), we have at iteration $m$:

$$\tau_k-\hat{\tau}_k^{(FB),m}=\frac{1}{2}\left(\tau_k-\hat{\tau}_k^{(F),m}\right)+\frac{1}{2}\left(\tau_k-\hat{\tau}_k^{(B),m}\right) \quad (46)$$

Replacing $\tau_k-\hat{\tau}_k^{(F),m}$ and $\tau_k-\hat{\tau}_k^{(B),m}$ by their corresponding expressions respectively from (118) and (119) provided in the Appendix C, the expression of the forward-backward bias at iteration $m$ for the $k^{th}$ observation is:

$$B_k^{(FB),m}=\frac{1}{2}(1+\gamma)^{(2N-2)(m-1)}\left[(1+\gamma)^{k-1}+(1+\gamma)^{2N-k-1}\right]B_1^{(F),1} \quad (47)$$

and the corresponding MSE expression is given by (48). We validate these analytical expressions in section V. Some conclusions can directly be drawn from the numerical evaluation of the obtained analytical expressions, as for instance the fact that with an adequate choice of the step size, 2 Forward-Backward iterations are sufficient for the proposed algorithm to reach to steady state in terms of MSE. To provide an absolute reference to evaluate both theoretical and experimental scheme performance, the next section is devoted to the derivation of the Bayesian CRB.

## IV. Bayesian Cramer Rao Bound

The standard and the modified CRBs [19, Chapter 3] [26, Chapter 2] are not suited to time-varying parameter estimation. In fact, these bounds do not take into account the statistical dependence between successive time delays. This dependence is naturally considered within the Bayesian framework: on one hand, the prior distribution $p(\boldsymbol{\tau})$ implicitly models the time dependence between stochastic time delays; on the other hand, the Bayesian framework is intrinsically based on the knowledge of a vector prior distribution $p(\boldsymbol{\tau})$. Within this context, a BCRB is derived by inverting the following Bayesian information matrix [26, Chapter 2]:

$$\boldsymbol{G_N}=E_{\boldsymbol{\tau}}[\boldsymbol{F(\tau)}]+E_{\boldsymbol{\tau}}[-\Delta_{\boldsymbol{\tau}}\log(p(\boldsymbol{\tau}))], \quad (49)$$

where $E_{\boldsymbol{\tau}}$ and $\Delta_{\boldsymbol{\tau}}$ are respectively the expectation with respect to the vector $\boldsymbol{\tau}$ and the Hessian matrix; $\boldsymbol{F(\tau)}$ is similar to the conventional Fisher information matrix [19]:

$$\boldsymbol{F(\tau)}=E_{\mathbf{r}|\boldsymbol{\tau}}[-\Delta_{\boldsymbol{\tau}}\log(p(\mathbf{r}|\boldsymbol{\tau}))]. \quad (50)$$

The first term of (49) considers the mean with respect to $\boldsymbol{\tau}$ of the information brought by the observation $\mathbf{r}$. The second term depends on the *a priori* information on $\boldsymbol{\tau}$. The diagonal elements of the inverse of $\boldsymbol{G_N}$ give the BCRB with respect to $\boldsymbol{\tau}$.

In the following, we distinguish two types of BCRB, namely, the off-line and the on-line BCRB for time synchronization. In an off-line context, the receiver waits until the whole observation block $\mathbf{r}=[r_1,...,r_N]$ is received. Then, the time delays $\tau_k$ $(k=1,..,N)$ are estimated using all the received samples. In an on-line context, $\tau_k$ is estimated using only the current and the previous observations $[r_1,...,r_k]$.

### A. Off-line BCRB

According to (50), and based on the expression of $p(\boldsymbol{r}|\boldsymbol{\tau})$ given by (16), we can deduce that $\boldsymbol{F(\tau)}$ is a diagonal matrix, independent of $\tau_k$. Thus, the first term of (49) can be written as:

$$E_{\boldsymbol{\tau}}[\boldsymbol{F(\tau)}]=J_D\boldsymbol{I_N}, \quad (51)$$

where $\boldsymbol{I_N}$ is the $(N\times N)$ identity matrix and for any index $k$:

$$J_D = E_{r_k|\tau_k}\left[-\frac{\partial^2\log(p(r_k|\tau_k))}{\partial\tau_k^2}\right] \quad (52)$$

$$= E_{r_k|\tau_k}\left[\left(\frac{\partial\log(p(r_k|\tau_k))}{\partial\tau_k}\right)^2\right]. \quad (53)$$

These expressions will be evaluated later for various transmission modes. From (18), the last term of (49) is:

$$E_{\boldsymbol{\tau}}[-\Delta_\tau\log(p(\tau))]=\begin{pmatrix}\frac{1}{\sigma_w^2}-D & -\frac{1}{\sigma_w^2} & 0 & \cdots & & 0\\ -\frac{1}{\sigma_w^2} & \frac{2}{\sigma_w^2} & -\frac{1}{\sigma_w^2} & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & \ddots & 0\\ \vdots & \ddots & 0 & -\frac{1}{\sigma_w^2} & \frac{2}{\sigma_w^2} & -\frac{1}{\sigma_w^2}\\ 0 & \cdots & 0 & 0 & -\frac{1}{\sigma_w^2} & \frac{1}{\sigma_w^2}\end{pmatrix}, \quad (54)$$

where $D=E_{\tau_1}[\frac{\partial^2\log(p(\tau_1))}{\partial\tau_1^2}]$. If we assume that the initial delay $\tau_1$, is uniformly distributed between $-T/2$ and $T/2$, then $D=0$.

$$M_k^{(F),m} = (1+\gamma)^{(2N-2)\times 2(m-1)+2k-2} M_1^{(F),m} + \Bigg( \sum_{i=2}^{k} \Big[ (1+\gamma)^{2N-1+k-i}\Omega(m) - (1+\gamma)^{k+i-2}\Omega(m) + (1+\gamma)^{k-i+1} \Big]^2 + \sum_{i=k+1}^{N} \Omega(m)^2 \Big[ (1+\gamma)^{2N-1+k-i} - (1+\gamma)^{k+i-2} \Big]^2 \Bigg) \sigma_w^2 + \Bigg( \sum_{i=k+1}^{N} \Omega(m)^2 \Big[ (1+\gamma)^{2N-2+k-i} - (1+\gamma)^{k+i-2} \Big]^2 + \sum_{i=2}^{k} \Big[ (1+\gamma)^{2N-2+k-i}\Omega(m) - (1+\gamma)^{k+i-2}\Omega(m) + (1+\gamma)^{k-i} \Big]^2 + \Omega(m)^2 \Big[ (1+\gamma)^{2N+2k-2} + (1+\gamma)^{2k-2} \Big] \Bigg) \sigma_q^2. \quad (42)$$

$$M_k^{(B),m} = (1+\gamma)^{(2N-2)\times 2(m-1)+4N-2k-2} M_1^{(F),m} + \Bigg( \sum_{i=2}^{k} \Big[ (1+\gamma)^{4N-1-i-k}\Omega(m) - (1+\gamma)^{2N-k+i-2}\Omega(m) + (1+\gamma)^{2N-k-i+1} \Big]^2 + \sum_{i=k+1}^{N} \Big[ (1+\gamma)^{4N-1-i-k}\Omega(m) - (1+\gamma)^{2N-k+i-2}\Omega(m) + (1+\gamma)^{2N-k-i+1} - (1+\gamma)^{i-k} \Big]^2 \Bigg) \sigma_w^2 + \Bigg( (1+\gamma)^{4N-2k} \sum_{i=2}^{k-1} \Big[ (1+\gamma)^{2N-2-i}\Omega(m) - (1+\gamma)^{i-2}\Omega(m) + (1+\gamma)^{-i} \Big]^2 + (1+\gamma)^{2N-2k} \Big[ 1 + \Omega(m)(1+\gamma)^{2N-2} \Big]^2 + \sum_{i=k}^{N-1} \Big[ (1+\gamma)^{4N-2-i-k}\Omega(m) - (1+\gamma)^{2N-k+i-2}\Omega(m) + (1+\gamma)^{2N-k-i} + (1+\gamma)^{i-k} \Big]^2 \Bigg) \sigma_q^2. \quad (45)$$

$$M_k^{(FB),m} = \frac{1}{4}\Big\{ M_k^{(F),m} + M_k^{(B),m} \Big\} + \frac{1}{2}\Big\{ (1+\gamma)^{2(2N-2)(m-1)+2N-2} M_1^{(F),1} \quad (48)$$
$$+ \Bigg( \sum_{i=2}^{k} (1+\gamma)^{2N} \Big[ (1+\gamma)^{-i+1} + \Omega(m)(1+\gamma)^{2N-i-1} - \Omega(m)(1+\gamma)^{i-2} \Big]^2 - \sum_{i=k+1}^{N} \Omega(m) \Big[ (1+\gamma)^{2N-i-1} + (1+\gamma)^{i-2} \Big] \Big[ \Omega(m)(1+\gamma)^{4N-i-1} - \Omega(m)(1+\gamma)^{2N+i-2} + (1+\gamma)^{2N-i+1} \Big] \Bigg) \sigma_w^2$$
$$- \Bigg( \sum_{i=2}^{k-1} (1+\gamma)^{2N} \Big[ \Omega(m)(1+\gamma)^{2N-i-2} + \Omega(m)(1+\gamma)^{i-2} + (1+\gamma)^{-i} \Big]^2 + \Big[ (1+\gamma)^{N-2} - \Omega(m)(1+\gamma)^{3N-2} \Big] (1+\gamma)^{N-2} - \sum_{i=k}^{N-1} \Omega(m) \Big[ (1+\gamma)^{2N-i-2} - (1+\gamma)^{i-2} \Big] \Big[ \Omega(m)(1+\gamma)^{4N-i-2} - \Omega(m)(1+\gamma)^{2N+i-2} - (1+\gamma)^{2N-i} - (1+\gamma)^{i} \Big] \Bigg) \sigma_q^2 \Big\}.$$

Consequently, from (49), (51) and (54):

$$\boldsymbol{G_N} = \beta \begin{pmatrix} A+1 & 1 & 0 & \cdots & & 0 \\ 1 & A & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & A & 1 \\ 0 & \cdots & 0 & 0 & 1 & A+1 \end{pmatrix}, \quad (55)$$

where $A = -\sigma_w^2 J_D - 2$ and $\beta = -\frac{1}{\sigma_w^2}$.

A similar study was carried on the BCRB relative to the estimate of the phase offset for a non-coded BPSK signal in [20]. The authors of [20] have presented the general expression of the inverse of the BIM which has the same form given by equation (55). They obtained the following expression expression:

$$[\boldsymbol{G_N}^{-1}]_{k,k} = \frac{1}{|G_N|} \Big[ \rho_1^2(\beta+\nu_1)^2 \nu_1^{N-3} + \rho_2^2(\beta+\nu_2)^2 \nu_2^{N-3} - \frac{\beta^2}{A-2} (\nu_1^{k-2}\nu_2^{N-k-1} + \nu_1^{N-k-1}\nu_2^{k-2}) \Big], \quad (56)$$

where $|\boldsymbol{G_N}|$ is the determinant of $\boldsymbol{G_N}$ given by:

$$|\boldsymbol{G_N}| = (A+2)\beta \left( \rho_1 \nu_1^{N-1} + \rho_2 \nu_2^{N-1} \right), \quad (57)$$

and for $m = 1, 2$:

$$\nu_m = \frac{1}{\sigma_w^2} + \frac{J_D}{2} \left( 1 + (-1)^m \times \sqrt{1 + \frac{4}{J_D \sigma_w^2}} \right), \quad (58)$$

$$\rho_m = \frac{\sqrt{1 + \frac{4}{\sigma_w^2 J_D}} + (-1)^m \times (1 + \frac{2}{\sigma_w^2 J_D})}{2\sqrt{1 + \frac{4}{\sigma_w^2 J_D}}}. \quad (59)$$

The BCRB of time delay estimation differs from the BCRB of phase offset estimation with the expression of $J_D$ which is going to be explicitly analyzed in the following subsections

for BPSK and QAM signals in various scenarios (DA, NDA and CA).

### B. On-line BCRB

In the on-line mode, the observations $r_k$ are received in order to update the estimated value of $\tau_k$. Only past and current observations are available. Tichavsky *et al.* [27] have already provided a method for updating the Bayesian information matrix in (49) from the time index $k-1$ to the time index $k$ according to the following recursive sequence:

$$C_k = \frac{\sigma_w^2 + C_{k-1}}{\sigma_w^2 J_D + 1 + J_D C_{k-1}}, \tag{60}$$

where $C_1 = \frac{1}{J_D}$. From (56), we also have that:

$$\begin{aligned} C_N &= [G_N^{-1}]_{N,N} && (61)\\ &= \frac{1}{|G_N|}\Bigg[\rho_1^2(\beta+\nu_1)^2\nu_1^{N-3} + \rho_2^2(\beta+\nu_2)^2\nu_2^{N-3} \\ &- \frac{b^2}{A-2}(\nu_1^{N-2}\nu_2^{-1} + \nu_1^{-1}\nu_2^{N-2})\Bigg]. && (62)\end{aligned}$$

This equation reflects that in the last position, the same amount of information is available both with the on-line and the off-line techniques. We also note that this bound decreases in time and converges to:

$$\lim_{N\to\infty}[G_N^{-1}]_{N,N} = \frac{-\sigma_w^2 + \sqrt{\sigma_w^4 + 4\frac{\sigma_w^2}{J_D}}}{2}. \tag{63}$$

### C. Derivation of $J_D$ for a BPSK signal

The analytical expressions of $J_D$ in the case of BPSK and square-QAM modulated signals is hereafter derived. Let us begin by considering the following equalities: $x_{k,1}(\tau_k) = \Re\{x_k(\tau_k)\}$, $x_{k,2}(\tau_k) = \Im\{x_k(\tau_k)\}$, $\dot{x}_{k,i}(\tau_k) = \frac{\partial x_{k,i}(u_k)}{\partial u_k}|_{u_k=\tau_k}$ and $\ddot{x}_{k,i}(\tau_k) = \frac{\partial^2 x_{k,i}(u_k)}{\partial u_k^2}|_{u_k=\tau_k}$, $i \in \{1,2\}$ with:

$$\begin{aligned} x_k(\tau_k) &= \sum_m a_{k-m} g(mT) + v_k(\tau_k) \\ &= a_k + v_k(\tau_k), && (64)\end{aligned}$$

$$\frac{\partial x_k(u_k)}{\partial u_k}|_{u_k=\tau_k} = \sum_m a_{k-m}\dot{g}(mT) + \frac{\partial v_k(u)}{\partial u}|_{u=\tau_k}, \tag{65}$$

and:

$$\begin{aligned} \frac{\partial^2 x_k(u_k)}{\partial u_k^2}|_{u_k=\tau_k} &= \sum_m a_{k-m}\ddot{g}(mT) + \frac{\partial^2 v_k(u)}{\partial u^2}|_{u=\tau_k} \\ &= a_k\ddot{g}(0) + \frac{\partial^2 v_k(u)}{\partial u^2}|_{u=\tau_k}. && (66)\end{aligned}$$

The analytical expression of $J_D$ depends on the selected synchronization mode. In the following subsections, we present an analytical development of $J_D$, as function of the SNR, respectively, in the case of DA, NDA and CA timing recovery.

*1) DA case:* Based on (16), (52) and (66), we obtain:

$$J_D = -\frac{|a_k|^2}{\sigma_n^2}\ddot{g}(0), \tag{67}$$

with $|a_k|^2 = 1$ for BPSK signals.

*2) NDA case:* From (6) and (16) and without loss of generality:

$$p(r_k|\tau_k, \mathbf{a}) = \frac{C}{2\pi\sigma_n^2}\exp\left(\frac{\Re\{a_k^* x_k(\tau_k)\}}{\sigma_n^2} - \frac{|a_k|^2}{2\sigma_n^2}\right). \tag{68}$$

By averaging over the equiprobable BPSK transmitted symbols **a**, we obtain:

$$\log\left(P\left(r_k|\tau_k\right)\right) = \log\left(\cosh\left(\frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)\right) + K, \tag{69}$$

where $K$ is a constant term independent of $\tau_k$. Thus, according to (53):

$$J_D = \frac{1}{\sigma_n^4}E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\dot{x}_{k,1}(\tau_k)^2\right]. \tag{70}$$

Given that $\dot{g}(mT)$ is equal to 0 for $m=0$, that the symbols $a_i$ are assumed to be independent and that the noise $\Re\{n(t)\}$ is independent of the symbols $a_i$ for any index $i$, we can conclude that $x_{k,1}(\tau_k)$ and $\dot{x}_{k,1}(\tau_k)$ are uncorrelated and consequently $\tanh\left(\frac{1}{\sigma_n^2}x_{k,1}(\tau_k)\right)$ and $\dot{x}_{k,1}(\tau_k)$ are uncorrelated. As a result:

$$J_D = \frac{1}{\sigma_n^4}E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\right]E\left[\dot{x}_{k,1}(\tau_k)^2\right], \tag{71}$$

where according to Appendix D:

$$E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\right] = 1 - \frac{\exp\left(-\frac{1}{2\sigma_n^2}\right)}{\sqrt{2\pi}\sigma_n}\int_{-\infty}^{+\infty}\frac{\exp\left(-\frac{x^2}{2\sigma_n^2}\right)}{\cosh\left(\frac{x}{\sigma_n^2}\right)}dx, \tag{72}$$

and:

$$E\left[\dot{x}_{k,1}(\tau_k)^2\right] = \sum_m \dot{g}(mT)^2 + \frac{\sigma_n^2}{2}\ddot{g}(0). \tag{73}$$

*3) CA case:* Let us consider $\lambda_k$ the log-likelihood ratio (LLR) output of the decoder at time instant $k$. We have that:

$$P(a_k = \pm 1) = \frac{\exp\left(\pm\frac{\lambda_k}{2}\right)}{2\cosh\left(\frac{\lambda_k}{2}\right)}. \tag{74}$$

Then, based on (68) and by averaging over the *a priori* probability of the BPSK transmitted symbols **a**, one obtains:

$$\log\left(P\left(r_k|\tau_k\right)\right) = \log\left(\frac{\cosh\left(\frac{x_{k,1}(\tau_k)}{\sigma_n^2} + \frac{\lambda_k}{2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)}\right) + K. \tag{75}$$

Thus, according to (53):

$$J_D = \frac{1}{\sigma_n^4}E\left[\tanh\left(\frac{\lambda_k}{2} + \frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\dot{x}_{k,1}(\tau_k)^2\right]. \tag{76}$$

Given that $\tanh\left(\frac{\lambda_k}{2} + \frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)$ and $\dot{x}_{k,1}(\tau_k)$ can be considered as uncorrelated if the coded symbols are transmitted with a large size interleaver [28], [29], we have:

$$J_D = \frac{1}{\sigma_n^4}E\left[\tanh\left(\frac{\lambda_k}{2} + \frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\right]E\left[\dot{x}_{k,1}(\tau_k)^2\right], \tag{77}$$

where $E\left[\dot{x}_{k,1}(\tau_k)^2\right]$ is given by (73) and according to Appendix D:

$$E\left[\left(\left(\frac{\lambda_k}{2}+\frac{x_{k,1}(\tau_k)}{\sigma_n^2}\right)^2\right]= 1-\frac{\exp\left(-\frac{1}{2\sigma_n^2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma_n}\int_{-\infty}^{+\infty}\frac{\exp\left(-\frac{x^2}{2\sigma_n^2}\right)}{\cosh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma_n^2}\right)}dx. \quad (78)$$

### *D. Derivation of $J_D$ for a square-QAM signal*

Similarly to the BPSK case, the analytical expressions of $J_D$ for square-QAM modulated signals are derived for the DA, NDA and CA transmission modes.

*1) DA case:* For DA mode the expression of $J_D$ is the same as that obtained for BPSK signals given by (67).

*2) NDA case:* Let us consider the case of equiprobable transmitted symbols. In order to compute the expression of $J_D$ using (53), we need to develop the expression of $p(r_k|\tau_k)$.

From (68) and assuming that the transmitted symbols $\boldsymbol{a}$ are equally likely, the probability likelihood function becomes:

$$\begin{aligned} p(r_k|\tau_k) &= \frac{C}{M}\sum_{v_i\in\mathcal{V}}\frac{1}{2\pi\sigma_n^2}\exp\left(\frac{\Re\{v_i^* x_k(\tau_k)\}}{\sigma_n^2}-\frac{|v_i|^2}{2\sigma_n^2}\right) \\ &\triangleq CF_k(\tau_k), \end{aligned} \quad (79)$$

where $\mathcal{V}=\{v_1,v_1,...,v_M\}$ is the constellation alphabet set.

Considering only square QAM modulated signals, the constellation size can be written as an even power of two, $M=2^{2p}$ and we can exploit the symmetry of the constellation.

Let us consider $\tilde{\mathcal{V}}$ the subset of the constellation alphabet with positive real and imaginary parts. If $\tilde{v}$ belongs to $\tilde{\mathcal{V}}$ then it can be written as:

$$\tilde{v}=(2m-1)d_p+j(2n-1)d_p, \quad (80)$$

where $\{m,n\}\in\{1,...,2^{p-1}\}$ and $d_p$ is the inter-symbol distance which has the following expression under the assumption of normalized energy symbols:

$$d_p=\frac{2^{p-1}}{\sqrt{2^p\sum_{k=1}^{2^{p-1}}(2k-1)^2}}. \quad (81)$$

We note that $\mathcal{V}=\tilde{\mathcal{V}}\bigcup\tilde{\mathcal{V}}^*\bigcup(-\tilde{\mathcal{V}})\bigcup(-\tilde{\mathcal{V}}^*)$, so that:

$$\begin{aligned} F_k(\tau_k) &= \sum_{\tilde{v}\in\tilde{\mathcal{V}}}\frac{\exp\left(-\frac{|\tilde{v}|^2}{2\sigma_n^2}\right)}{2\pi\sigma_n^2 2^{2p-2}}\cosh\left(\frac{\Re\{\tilde{v}\}x_{k,1}(\tau_k)}{\sigma_n^2}\right) \\ &\times \cosh\left(\frac{\Im\{\tilde{v}\}x_{k,2}(\tau_k)}{\sigma_n^2}\right). \end{aligned} \quad (82)$$

Using (80), we obtain:

$$\begin{aligned} F_k(\tau_k) &= \frac{1}{2\pi\sigma_n^2 2^{2p-2}}\sum_{m=1}^{2^{p-1}}\sum_{n=1}^{2^{p-1}}\exp\left(-\frac{(2m-1)^2+(2n-1)^2}{2\sigma_n^2}d_p^2\right) \\ &\times\cosh\left(\frac{(2m-1)d_p x_{k,1}(\tau_k)}{\sigma_n^2}\right)\cosh\left(\frac{(2n-1)d_p x_{k,2}(\tau_k)}{\sigma_n^2}\right) \\ &= H_k\Big(x_{k,1}(\tau_k)\Big)H_k\Big(x_{k,2}(\tau_k)\Big), \end{aligned} \quad (83)$$

where:

$$H_k(x)=\sum_{m=1}^{2^{p-1}}\frac{2\exp\left(-\frac{(2m-1)^2}{2\sigma_n^2}d_p^2\right)}{\sqrt{2\pi\sigma_n^2 2^p}}\cosh\left(\frac{(2m-1)d_p x}{\sigma_n^2}\right). \quad (84)$$

A similar factorization of the likelihood function, in the NDA case, using the full symmetry of the square QAM constellation has been proposed in [30] and [31] respectively for SNR and for frequency and phase NDA estimation.

As a result, from (79) and (83), we have:

$$\begin{aligned} \frac{\partial\log\left(p(x_{k,1}k|\tau_k)\right)}{\partial\tau_k} &= \frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\dot{x}_{k,1}(\tau_k) \\ &+ \frac{\dot{H}_k\left(x_{k,2}(\tau_k)\right)}{H_k\left(x_{k,2}(\tau_k)\right)}\dot{x}_{k,2}(\tau_k), \end{aligned} \quad (85)$$

where:

$$\begin{aligned} \dot{H}_k(x) &= \frac{\partial H_k(x)}{\partial x} \\ &= \frac{2}{\sqrt{2\pi\sigma_n^2}2^p}\sum_{m=1}^{2^{p-1}}\exp\left(-\frac{(2m-1)^2}{2\sigma_n^2}d_p^2\right) \\ &\times \frac{(2m-1)d_p}{\sigma_n^2}\sinh\left(\frac{(2m-1)d_p x}{\sigma_n^2}\right). \end{aligned} \quad (86)$$

Then using (53) and (85), we find:

$$\begin{aligned} J_D &= E\left[\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2\dot{x}_{k,1}(\tau_k)^2\right] \\ &+ E\left[\left(\frac{\dot{H}_k\left(x_{k,2}(\tau_k)\right)}{H_k\left(x_{k,2}(\tau_k)\right)}\right)^2\dot{x}_{k,2}(\tau_k)^2\right] \\ &+ 2E\left[\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\frac{\dot{H}_k\left(x_{k,2}(\tau_k)\right)}{H_k\left(x_{k,2}(\tau_k)\right)}\dot{x}_{k,1}(\tau_k)\dot{x}_{k,2}(\tau_k)\right]. \end{aligned} \quad (87)$$

Given that $x_{k,1}(\tau_k)$ (resp. $x_{k,2}(\tau_k)$) and $\dot{x}_{k,1}(\tau_k)$ (resp. $\dot{x}_{k,2}(\tau_k)$) are uncorrelated, then the third term of (87) is equal to zero. Furthermore, $x_{k,1}(\tau_k)$ and $x_{k,2}(\tau_k)$ have the same statistical properties. Thus:

$$\begin{aligned} J_D &= 2E\left[\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2\dot{x}_{k,1}(\tau_k)^2\right] \quad (88) \\ &= 2E\left[\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2\right]E\left[\dot{x}_{k,1}(\tau_k)^2\right], \quad (89) \end{aligned}$$

with:

$$E\left[\dot{x}_{k,1}(\tau_k)^2\right]=\sum_i E[a_i^2]\dot{g}^2\left((k-i)T\right)-\frac{\sigma_n^2}{2}\ddot{g}(0). \quad (90)$$

We now develop the first term of (89), which by definition is:

$$E\left[\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2\right]=\int_{-\infty}^{+\infty}\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2 p(x_{k,1}(\tau_k))dx_{k,1}(\tau_k). \quad (91)$$

$g$ being a Nyquist filter, the received matched filtered signal can be written as:

$$x_k(\tau_k)=a_k+v_k(\tau_k), \quad (92)$$

so that:

$$p(x_k(\tau_k)) = \sum_{v\in\mathcal{V}} \frac{p(a_k=v)}{2\pi\sigma_n^2}\exp\left(-\frac{|x_k(\tau_k)-v|^2}{2\sigma_n^2}\right) = H_k(x_{k,1}(\tau_k))H_k(x_{k,2}(\tau_k))\times \exp\left(-\frac{x_{k,1}^2(\tau_k)+x_{k,2}^2(\tau_k)}{2\sigma_n^2}\right). \quad (93)$$

Given that $x_{k,1}(\tau_k)$ and $x_{k,2}(\tau_k)$ are respectively the real and imaginary parts of $x_k(\tau_k)$, which are two independent random variables identically distributed, we have:

$$p(x_k(\tau_k)) = p(x_{k,1}(\tau_k))p(x_{k,2}(\tau_k)), \quad (94)$$

where:

$$p(x_{k,i}(\tau_k)) = H_k(x_{k,i}(\tau_k))\exp\left(-\frac{x_{k,i}^2(\tau_k)}{2\sigma_n^2}\right), \quad (95)$$

with $i\in\{1,2\}$. As a result (91) becomes:

$$E\left[\left(\frac{\dot{H}_k\left(x_{k,1}(\tau_k)\right)}{H_k\left(x_{k,1}(\tau_k)\right)}\right)^2\right] = \int_{-\infty}^{+\infty}\frac{\dot{H}_k(x)^2}{H_k(x)}\exp\left(-\frac{x^2}{2\sigma_n^2}\right)dx. \quad (96)$$

By replacing (96) and (90) into (89), we obtain the expression of $J_D$.

*3) CA case:* Let us consider the case of Gray-Mapped square-QAM symbols $a_k$. The results can also be transposed to any other mapping technique. We assume the constellation to be Gray-coded taking values in an alphabet set of size $M$. The chosen $k^{th}$ transmitted symbol $a_k$ is denoted as:

$$a_k \Leftrightarrow b_1^k b_2^k ... b_{log_2(M)}^k, \quad (97)$$

where $b_j^k$ corresponds to the $j^{th}$ binary information of $a_k$.

We can assume that the coded bits in a symbol are statistically independent by using a large-size interleaver [28], [29]. Let us consider, $\lambda_m^k$ the soft output of the decoder at any time index $k$ such that:

$$\lambda_m^k = \ln\left(\frac{P[b_m^k=1]}{P[b_m^k=0]}\right),\ m=1,\cdots,log_2(M). \quad (98)$$

The soft-information $\lambda_m^k$ is generally obtained after several decoding iterations by a soft decoder.

In [8], we developed the expression of the likelihood probability for a constant time delay $\tau$ by incorporating the LLRs of the coded bits, into the likelihood function for Gray-coded square QAM constellations. This was based on results proposed by [32]. The obtained result can be directly adapted to the variable time delay model which allows us to obtain the following expression:

$$p(r_k|\tau_k) = H_k^{2p}\left(x_{k,1}(\tau_k)\right)H_k^{2p-1}\left(x_{k,2}(\tau_k)\right), \quad (99)$$

where:

$$H_k^l(x) = \frac{2w_{k,l}}{\sqrt{2\pi\sigma_n^2}}\sum_{i=1}^{2^{p-1}}\theta_{k,l}(i)\exp\left(-\frac{(2i-1)^2d_p^2}{2\sigma_n^2}\right)\times \cosh\left(\frac{(2i-1)d_px}{\sigma_n^2}+(-1)^l\frac{\lambda_l^k}{2}\right), \quad (100)$$

$$\omega_{k,2p} = \prod_{l=1}^{p}\frac{1}{2\cosh\left(\frac{\lambda_{2l}^k}{2}\right)}, \quad (101)$$

$$\omega_{k,2p-1} = \prod_{l=1}^{p}\frac{1}{2\cosh\left(\frac{\lambda_{2l-1}^k}{2}\right)}, \quad (102)$$

and $\theta_{k,2p}(i)$ and $\theta_{k,2p-1}(i)$ are recursively obtained according to the following equations:

$$\theta_{k,2p}(i) = \theta_{k,2p-2}\left(\frac{|2i-1-2^{p-1}|+1}{2}\right)\times \exp\left(\left(2\lfloor\frac{i-1}{2^{p-2}}\rfloor-1\right)\frac{\lambda_{2p-2}^k}{2}\right), \quad (103)$$

$$\theta_{k,2p-1}(i) = \theta_{k,2p-3}\left(\frac{|2i-1-2^{p-1}|+1}{2}\right)\times \exp\left(\left(2\lfloor\frac{i-1}{2^{p-2}}\rfloor-1\right)\frac{\lambda_{2p-3}^k}{2}\right), \quad (104)$$

where $\lfloor x\rfloor$ is the integer part of $x$, $\theta_{k,1}(1)=1$ and $\theta_{k,2}(1)=1$.

Thus similarly to the NDA case and based on the fact that $x_{k,1}(\tau_k)$ (resp. $x_{k,2}(\tau_k)$) and $\dot{x}_{k,1}(\tau_k)$ (resp. $\dot{x}_{k,2}(\tau_k)$) are uncorrelated, the expression of $J_D$ is given by:

$$J_D = E\left[\left(\frac{\dot{H}_k^{2p}\left(x_{k,1}(\tau_k)\right)}{H_k^{2p}\left(x_{k,1}(\tau_k)\right)}\right)^2\right]E\left[\dot{x}_{k,1}(\tau_k)^2\right] + E\left[\left(\frac{\dot{H}_k^{2p-1}\left(x_{k,2}(\tau_k)\right)}{H_k^{2p-1}\left(x_{k,2}(\tau_k)\right)}\right)^2\right]E\left[\dot{x}_{k,2}(\tau_k)^2\right] \quad (105)$$

where $\dot{H}_k^l\left(x_{k,1}(\tau_k)\right) = \partial H_k^l(x)/\partial x$ for $l\in\{2p,2p-1\}$. The expression of both $E\left[\dot{x}_{k,1}(\tau_k)^2\right]$ and $E\left[\dot{x}_{k,2}(\tau_k)^2\right]$ is given by (90). Following the steps used in the NDA case we obtain:

$$p(x_{k,i}(\tau_k)) = H_k^{2p+1-i}(x_{k,i}(\tau_k))\exp\left(-\frac{x_{k,i}^2(\tau_k)}{2\sigma_n^2}\right), \quad (106)$$

with $i\in\{1,2\}$ and thus:

$$E\left[\left(\frac{\dot{H}_k^{2p+1-i}(x_{k,i}(\tau_k))}{H_k^{2p}(x_{k,i}(\tau_k))}\right)^2\right] = \int_{-\infty}^{+\infty}\frac{\dot{H}_k^{2p+1-i}(x)^2}{H_k^{2p+1-i}(x)}\exp\left(-\frac{x^2}{2\sigma_n^2}\right)dx. \quad (107)$$

By replacing (107) and (90) into (105) we finally obtain the expression of $J_D$ in the CA case.

We mention that the integrand functions involved in (96), (107), respectively, decrease rapidly as $x$ increases. Therefore, the integrals over $]-\infty,+\infty[$ can be accurately approximated by a finite integral over an interval $[-L,+L]$ and the Riemann integration method can be adequately used. The evaluation of the BCRB is thus possible as all implied expressions in equation (49) have been derived.

In this section, we derived the BCRB for code-aided delay estimation in the case of BPSK, QPSK and QAM modulated signals. It is worthy to note that the herein derived BCRB is conditioned to the LLR values and thus depends on the coding technique. A further averaging over the possible soft output values should be then carried to evaluate the BCRB. This is here processed by averaging over Monte Carlo trials. Since the

BCRB depends on the coding technique, thus for the Monte Carlo simulation, we consider the same encoder as that used by the selected synchronizer. For the BCRB evaluation in the CA mode, the LLRs are computed in the case where the time delay is perfectly recovered, however, for the MSE evaluation, the LLRs are computed by considering the estimated time delay given by the CA TED. Further work can be carried to develop an analytical BCRB expression by analytically accounting for the LLRs variation as done in [12] for the problem of MSE computation.

The closed form expressions of the previous BCRB for the different transmission modes will be compared to the estimator performance in the next section.

## V. Simulation Results

In this paragraph, we simulate in the first four figures, the performance of the studied synchronization algorithm for small blocks of non-coded symbols transmitted over a Gaussian channel. The signal, with an up-sampling factor equal to $8$, is passed through a raised cosine filter, with a roll-off factor $\alpha = 0.3$, before being sent. The channel introduces a variable time delay $\tau_k$ following a brownian model (see (4)). The signal is matched filtered at the receiver and retrieved by the synchronizer. We finally evaluate the MSE over $1000$ Monte Carlo trials. The above adaptation equations of the time error detector involve the evaluation of the matched filter output derivative $(\partial x_k(\hat{\tau}_i)/\partial\hat{\tau}_i)$ at the estimated time delays values. The values of $x_k(\hat{\tau}_{k-1})$ are obtained via a cubic $4-$sample Lagrange interpolator. This interpolator can be implemented efficiently using the Farrow structure [33]. The LLR values are initialized by the soft demapper's outputs. The turbo-decoder's outputs are reinjected only twice to the synchronizer's input. $\hat{\tau}_k$ is initialized to $0$ and its estimation is depicted after two forward-backward synchronization iterations when the steady state is achieved. Simulation results are evaluated for an initial time delay uniformly distributed in $[-T/2, T/2]$ and randomly generated delay samples following the brownian evolution model for a given $\sigma_w^2$. The step-size $\mu$ is chosen so as to insure a global convergence of the algorithm to the optimum delay. An adaptive step-size [34] could be implemented in order to have both a faster convergence and low MSE at the steady-state. However, for simplicity reasons, the step size value is chosen as constant during our synchronization process.

We use the following notations in the figures of the present paragraph. "One Forward" means that the MSE is measured after one (on-line) forward estimation without any backward estimation. "Multiple Forward" (resp. "Multiple F/B") means that the MSE of the Forward (resp. Forward/Backward) estimation is measured after three (off-line) F/B iterations. Finally "Simu" is related to simulation results whereas "Theo" refers to the theoretical MSE results obtained with the previous linear model analysis of section III.

Fig. 1 illustrates the normalized estimated time delay $\hat{\tau}_k/T$ using the forward, the backward and the forward-backward algorithms after three iterations in a DA context for a BPSK modulated signal. Along the observation block, the off-line algorithm provides a better estimate of the time delay than the forward only or the backward only as it averages at any time the two trajectories. At the extremities of the observation block, the three algorithms have equivalent performance. This is a logical result since in practice the first backward loop estimate of the time delay uses the last estimate of the forward loop and vice versa.

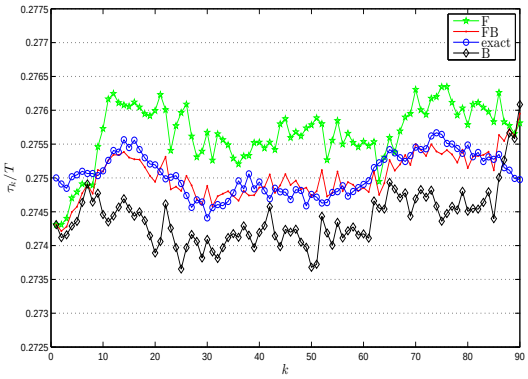


Fig. 1. $\tau_k/T$ and its estimate along the observation block, $\sigma_w/T = 10^{-4}$ in DA mode for a BPSK modulation

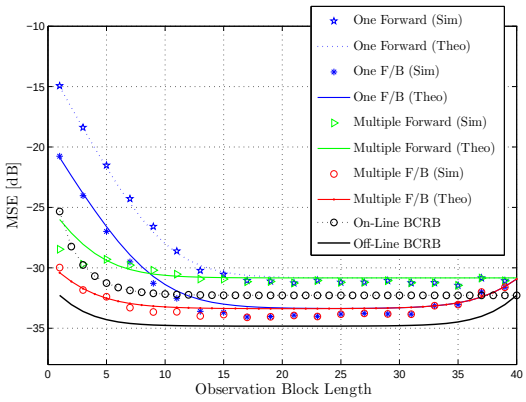


Fig. 2. Comparison between analytical and simulated MSE in various positions of the block, SNR=10 dB, $\sigma_w/T = 10^{-4}$ , BPSK in DA mode.

Fig. 2 presents versus time index, both the DA MSE and BCRBs (on-line and off-line) for a SNR$= 10$ dB. By increasing the number of observations, the on-line performance is improved. The on-line bound thus decreases and converges to its asymptote given by (63). The one forward MSE also decreases and saturates near the corresponding on-line bound. This saturation is both due to the time varying delay and the self noise (see (48)). The same figure shows the FB recursion MSE curves and their corresponding lower limit given by the off-line BCRB. One notes that there is a gain of $3$ dB in terms of MSE in the center of the block by using the FB algorithm and that, similarly, the on-line BCRB is higher than the off-line BCRB. It is also noted that, the performance of the proposed off-line FB is better than that of one on-line forward iteration and even than that of the multiple-forward scheme. We would also like to point out that we only need two forward-backward iterations to reach the steady state. Finally, there is a good

agreement between the simulated and the theoretical results of section III which validates our linear analytical model.

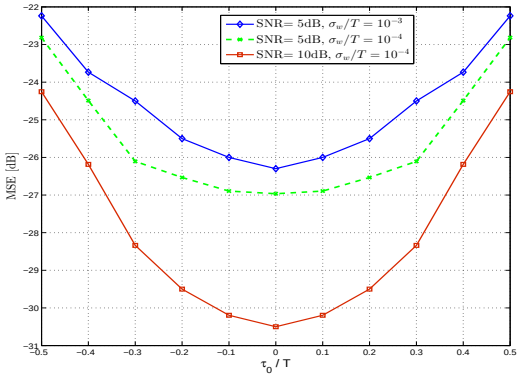


Fig. 3. MSE versus $\frac{\tau_0}{T}$ in the center of the observation block for various SNRs and $\sigma_w/T$ values, 16QAM in DA mode.

Fig. 3 displays the mean square error of the time delay estimation versus the mean value of the time delay in the center of the observation block for various SNRs and time delay standard deviation values $\sigma_w$ when using a 16QAM modulation. It illustrates that the MSE decreases for lower $\sigma_w$ and higher SNR values.

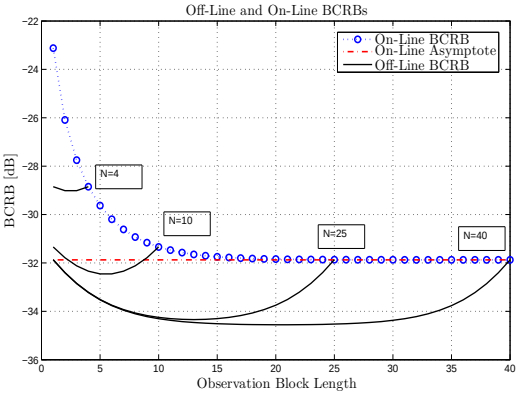


Fig. 4. BCRBs for various positions and different block lengths, SNR=10dB, $\sigma_w/T = 10^{-4}$, BPSK in DA mode.

Fig. 4 depicts the off-line and the on-line BCRB at each time index for different block lengths as well as the infinite block length on-line asymptote (given by (63)). Once again, one can conclude that the off-line bound is always lower than the on-line bound independently of the observation block size. Also, the number of observations processed at the synchronizer has a great impact on the system performance. It is noted that the best time estimate in an off-line context, can be expected at the center of the observation block. In fact, at this position, both past and future observations, which are strongly correlated, are used for the timing recovery. On the contrary, less information is fed to the synchronizer at the borders of the block where mostly only past (resp. future) observation is used by the forward (resp. backward) recursion.

In Fig. 5.a (resp. Fig. 5.b), we compare, versus time index,

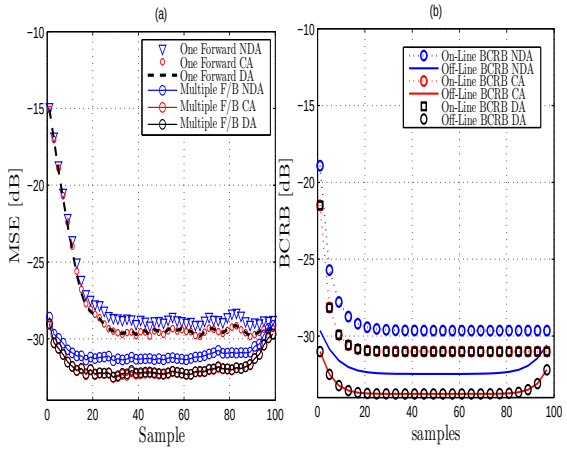


Fig. 5. MSE and BCRB in various positions in the block, SNR=10dB, $\sigma_w = 10^{-4}$, BPSK signals.

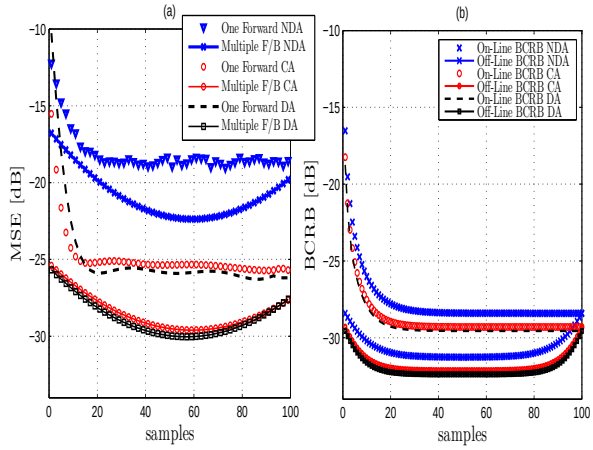


Fig. 6. MSE and BCRB in various positions in the block, SNR=10dB, $\sigma_w = 10^{-4}$, 16QAM signals.

the MSE (resp. the on-line and off-line BCRB) for respectively the DA, NDA and CA synchronization mode, for BPSK modulated signals. Similar curves are drawn in Fig. 6 for 16QAM modulated signals. For the CA mode, the encoder is composed of two identical RSCs concatenated in parallel with systematic rate $r = \frac{1}{2}$ and with generator polynomials $(1, 0, 1, 1)$ and $(1, 1, 0, 1)$. The LLR values are evaluated after 6 turbo decoding iterations for a BPSK modulation and after 4 turbo decoding iterations for a 16QAM modulation. A large interleaver is placed between the two RSCs. Similar results can be obtained with soft decoders corresponding to LDPC codes, Block Turbo Codes or coded modulations. The SNR is fixed to 10 dB. It is clear that the off-line techniques outperform the on-line techniques (for any mode DA/NDA/CA). We note that, logically, the CA mode outperforms the NDA mode and approaches the DA mode. As expected, similar conclusions to those drawn in Fig. 2 can be obtained for the corresponding BCRBs in the various modes (DA/NDA/CA).

Fig. 7 (resp. Fig. 8) displays the MSE and BCRB curves versus the SNR at the center position of the observation block of both the on-line and off-line scenarios for the DA, NDA and CA modes for a BPSK (resp. 16QAM) modulated signal.

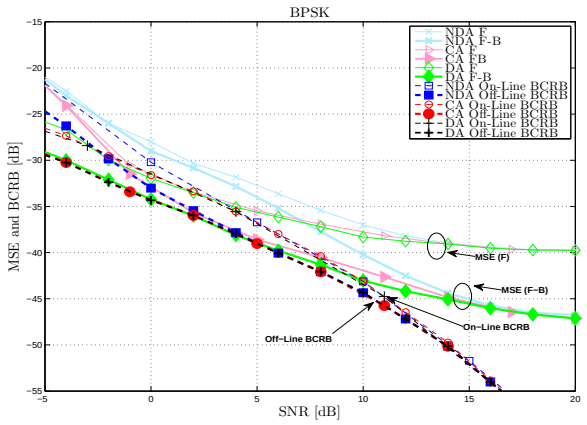


Fig. 7. On-Line and Off-Line BCRB versus SNR, $\sigma_w/T = 10^{-4}$, Block length $= 2000$, BPSK signal.

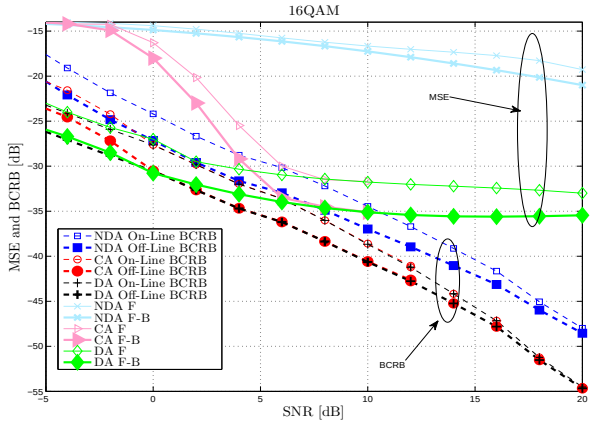


Fig. 8. On-Line and Off-Line BCRB versus SNR, $\sigma_w/T = 10^{-4}$, Block length $= 2000$, 16QAM signal.

For both the on-line and off-line approaches, the CA mode outperforms the NDA mode and approaches the DA mode over a large interval of SNRs. However, better performance can sometimes be achieved with the off-line NDA scenario compared to the on-line CA scenario for BPSK signals. According to Fig. 7 when the SNR $\geq -2$dB (resp. SNR $\geq$ 7dB) the off-line BCRB of the NDA mode is lower than the on-line BCRB in the CA mode (resp. the MSE of the NDA forward-backward algorithm is lower than the MSE of CA forward algorithm). Equivalent performance is expected with the off-line NDA scenario and the on-line CA scenario using a 16QAM modulation for a SNR inferior to $-2$dB (see Fig. 8). For such a case, the NDA off-line loop is easier to implement than the on-line CA approach. For high SNRs, the off-line and the on-line BCRBs merge. In other words, the theoretical Bayesian time recovery problem becomes equivalent to a deterministic time delay estimation. However, as illustrated by the MSE curves of Fig. 7 and Fig. 8, in practice there is an advantage for the off-line approach as the self-noise is less critical for this mode than for the on-line mode. For mid-range SNRs, the off-line scenario is more advantageous than the on-line scenario: as one can expect from the BCRB study, a gain of 3 dB is brought by the use of the *a priori* information. At low SNR, the CA and the NDA modes merge and the DA mode achieves lower errors. Actually, in modern systems, receivers are often constrained to work at low to medium SNRs so that one can benefit from the advantages brought by the off-line and even the CA approaches.

## VI. Conclusion

In this paper, an off-line smoothing algorithm for dynamical time delay recovery is proposed. Theoretical performance of the off-line technique is derived and fits well the simulation results. The Bayesian CRB is also evaluated for DA, NDA and CA estimators for both off-line and on-line scenarios. The presented algorithm reaches a MSE performance close to the Bayesian CRB and outperforms the conventional on-line timing detectors. One forward-backward iteration outperforms several one way processing iterations over a large range of observation block lengths. Simulation results show the improvement brought by respectively the off-line and the CA schemes. The NDA off- line approach is simple to implement as it just averages two gradient descent trajectories. Therefore, trading off implementation complexity and performance, the NDA off-line approach can be preferable to the CA on-line scheme.

## Appendix A

According to (4), (23) and (30), for the forward algorithm we have:

$$\begin{cases} \hat{\tau}_k^{(F)} = \hat{\tau}_{k-1}^{(F)} + \gamma(\hat{\tau}_{k-1}^{(F)} - \tau_k) + q_k \\ \tau_k \quad = \tau_{k-1} + w_k. \end{cases} \tag{108}$$

Based on (108), we have:

$$\begin{aligned} \tau_k - \hat{\tau}_k^{(F)} &= (1+\gamma)(\tau_k - \hat{\tau}_{k-1}^{(F)}) - q_k \\ &= (1+\gamma)^{k-1}(\tau_1 - \hat{\tau}_1^{(F)}) + \sum_{i=2}^{k} w_i (1+\gamma)^{k-i+1} \\ &- \sum_{i=2}^{k} q_i (1+\gamma)^{k-i}. \end{aligned} \tag{109}$$

This leads to the estimation bias of the time delay after only one forward loop given by (31).

## Appendix B

The MSE expression is given by:

$$M_k^{(FB)} = \frac{1}{4} M_k^{(B)} + \frac{1}{4} M_k^{(F)} + \frac{1}{2} E\left[ \left( \tau_k - \hat{\tau}_k^{(B)} \right) \left( \tau_k - \hat{\tau}_k^{(F)} \right) \right]. \tag{110}$$

Given that $\tau_N - \hat{\tau}_N^{(B)} = \tau_N - \hat{\tau}_N^{(F)}$, thus using (109) for $k = N$ and (33), we find that:

$$\begin{aligned} \tau_k - \hat{\tau}_k^{(B)} = &(1+\gamma)^{2N-k-1}(\tau_1 - \hat{\tau}_1^{(F)}) + \sum_{i=2}^{N} w_i (1+\gamma)^{2N-k-i+1} \\ &- \sum_{i=2}^{N} q_i (1+\gamma)^{2N-k-i} - \sum_{i=k+1}^{N} w_i (1+\gamma)^{i-k} \\ &- \sum_{i=k+1}^{N-1} q_i (1+\gamma)^{i-k} - q_k. \end{aligned} \tag{111}$$

Thus, from (110), the MSE expression after one forward/backward iteration is given by (40).

## APPENDIX C

We initialize (at iteration $m \geq 2$) the first estimate of the forward algorithm at the last estimate of the backward algorithm of the previous iteration.

According to the iterative processing of the forward algorithm specified in equation (111), the estimation error at the end of the backward loop is equal to the initial error of the second iteration of the forward loop. Thus:

$$\tau_1 - \hat{\tau}_1^{(F),2} = \tau_1 - \hat{\tau}_1^{(B),1}. \tag{112}$$

The expression of $\tau_k - \hat{\tau}_k^{(B),1}$ is provided by (111). For $k = 1$:

$$\begin{aligned}\tau_1 - \hat{\tau}_1^{(F),2} &= \tau_1 - \hat{\tau}_1^{(B),1} \\ &= (1+\gamma)^{2N-2}(\tau_1 - \hat{\tau}_1^{(F),1}) \\ &+ \sum_{i=2}^{N} w_i(1+\gamma)^{2N-i} - \sum_{i=2}^{N} q_i(1+\gamma)^{2N-1-i} \\ &- \sum_{i=2}^{N} w_i(1+\gamma)^{i-1} - \sum_{i=1}^{N-1} q_i(1+\gamma)^{i-1}\end{aligned} \tag{113}$$

Similarly, we use for the following iteration:

$$\tau_1 - \hat{\tau}_1^{(F),3} = \tau_1 - \hat{\tau}_1^{(B),2}. \tag{114}$$

Based on (**??**) and for $k = 1$:

$$\begin{aligned}\tau_1 - \hat{\tau}_1^{(B),2} &= (1+\gamma)^{2N-2}(\tau_1 - \hat{\tau}_1^{(F),2}) + \sum_{i=2}^{N} w_i(1+\gamma)^{2N-i} \\ &- \sum_{i=2}^{N} q_i(1+\gamma)^{2N-i-1} - \sum_{i=2}^{N} w_i(1+\gamma)^{i-1} \\ &- \sum_{i=1}^{N-1} q_i(1+\gamma)^{i-1}\end{aligned} \tag{115}$$

Introducing (113) into (115), we obtain:

$$\tau_1 - \hat{\tau}_1^{(B),2} = \tau_1 - \hat{\tau}_1^{(F),3}. \tag{116}$$

By induction, at iteration $m$ we obtain:

$$\begin{aligned}\tau_1 - \hat{\tau}_1^{(F),m} &= (1+\gamma)^{(2N-2)\times(m-1)}(\tau_1 - \hat{\tau}_1^{(F),1}) \\ &+ \Omega(m)\left[\sum_{i=2}^{N} w_i(1+\gamma)^{2N-i} - q_i(1+\gamma)^{2N-1-i}\right] \\ &- \Omega(m)\left[\sum_{i=2}^{N} w_i(1+\gamma)^{i-1} + \sum_{i=1}^{N-1} q_i(1+\gamma)^{i-1}\right],\end{aligned} \tag{117}$$

where $\Omega(m)$ is given by (43). According to expressions of $\tau_k - \hat{\tau}_k^{(F)}$ and $\tau_k - \hat{\tau}_k^{(B)}$ respectively in (109) and (111) and by replacing $\tau_1 - \hat{\tau}_1^{(F)}$ with $\tau_1 - \hat{\tau}_1^{(F),m}$ from (117), at iteration $m$ and the $k^{th}$ observation, we obtain the expression of $\tau_k - \hat{\tau}_k^{(F),m}$ (resp. $\tau_k - \hat{\tau}_k^{(B),m}$) given by (118) (resp. (119)):

$$\begin{aligned}\tau_k - \hat{\tau}_k^{(F),m} &= (1+\gamma)^{(2N-2)\times(m-1)+k-1}(\tau_1 - \hat{\tau}_1^{(F),1}) \\ &+ \Omega(m)\sum_{i=2}^{N}\left[w_i\left((1+\gamma)^{2N-1+k-i} - (1+\gamma)^{k+i-2}\right)\right] \\ &- \Omega(m)\left[\sum_{i=2}^{N} q_i(1+\gamma)^{2N-2+k-i} + \sum_{i=1}^{N-1} q_i(1+\gamma)^{k+i-2}\right] \\ &- \sum_{i=2}^{k}(1+\gamma)^{k-i}\left(q_i + w_i(1+\gamma)\right),\end{aligned} \tag{118}$$

$$\begin{aligned}\tau_k - \hat{\tau}_k^{(B),m} &= (1+\gamma)^{(2N-2)\times(m-1)+2N-k-1}(\tau_1 - \hat{\tau}_1^{(F),1}) \\ &+\Omega(m)\left[\sum_{i=2}^{N} w_i(1+\gamma)^{4N-k-1-i} - q_i(1+\gamma)^{4N-k-2-i}\right] \\ &-\Omega(m)\left[\sum_{i=2}^{N} w_i(1+\gamma)^{2N-k+i-2} + \sum_{i=1}^{N-1} q_i(1+\gamma)^{2N-k+i-2}\right] \\ &+\sum_{i=2}^{N} w_i(1+\gamma)^{2N-k-i+1} - \sum_{i=2}^{N} q_i(1+\gamma)^{2N-k-i} \\ &- \sum_{i=k+1}^{N} w_i(1+\gamma)^{i-k} - \sum_{i=k}^{N-1} q_i(1+\gamma)^{i-k}.\end{aligned} \tag{119}$$

The previous expressions lead to the estimation bias after $m$ iterations using the forward (resp. backward) algorithm given by (41) (resp. (44) ) and the MSE expressions given by (42) and (45).

## APPENDIX D

- **Derivation of** $E\left[\dot{x}_{k,1}(\tau_k)^2\right]$:

We have that:

$$\begin{aligned}E\left[\dot{x}_{k,1}(\tau_k)^2\right] &= E\left[\left[\sum_m a_{k-m}\dot{g}(mT) + \Re\{\dot{v}_k(\tau_k)\}\right]\right. \\ &\quad \left.\times\left[\sum_n a_{k-n}\dot{g}(nT) + \Re\{\dot{v}_k(\tau_k)\}\right]\right] \\ &= E\left[\sum_m a_{k-m}^2\dot{g}(mT)^2\right. \\ &+ \left.\sum_{i=1}^{N}\Re\{n_i\}^2\dot{h}(iT - kT - \tau_k)^2\right] \\ &= \sum_m \dot{g}(mT)^2 + \frac{\sigma^2}{2}\ddot{g}(0).\end{aligned} \tag{120}$$

- **Derivation of** $E\left[\tanh\left(\frac{\lambda_k}{2} + \frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right]$:

Based on (64) and given the LLR, $\lambda_k$, the pdf of $x_{k,1}(\tau_k)$ is:

$$\begin{aligned}p_{x_{k,1}(\tau_k)}(x) &= \frac{1}{2\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma}\left[\exp\left(\frac{\lambda_k}{2}\right)\exp\left(-\frac{(x-1)^2}{2\sigma^2}\right)\right. \\ &+ \left.\exp\left(-\frac{\lambda_k}{2}\right)\exp\left(-\frac{(x+1)^2}{2\sigma^2}\right)\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2+1}{2\sigma^2}\right)\cosh\left(\frac{\lambda_k}{2} + \frac{x}{\sigma^2}\right).\end{aligned} \tag{121}$$

Consequently:

$$\begin{aligned} & E\left[\tanh\left(\frac{\lambda_k}{2}+\frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right] \\ = & \int_{-\infty}^{+\infty}\tanh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)^2\frac{\exp\left(-\frac{x^2+1}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma}\cosh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)dx \\ = & \frac{\exp\left(-\frac{1}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\frac{\sinh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)^2}{\cosh\left(\frac{x}{\sigma^2}\right)}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx \end{aligned}$$

Using again $\sinh(x)^2=\cosh(x)^2-1$, we find:

$$\begin{aligned} & E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right] \\ = & \frac{\exp\left(-\frac{1}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma}\left[I_k-\int_{-\infty}^{+\infty}\frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)}dx\right] \end{aligned} \quad (122)$$

where

$$I_k = \int_{-\infty}^{+\infty}\cosh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right)dx.$$

As $\cosh(a+b)=\cosh(a)\cosh(b)+\sinh(a)\sinh(b)$ and given that the integral over $R$ of an odd function is equal to 0, thus using (123),

$$\int_0^{+\infty}\exp(-\beta x^2)\cosh(ax)dx=\frac{1}{2}\sqrt{\frac{\pi}{\beta}}\exp\left(\frac{a^2}{4\beta}\right),\ \text{for } \beta>0, \quad (123)$$

we obtain:

$$I_k = \cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma\exp\left(\frac{1}{2\sigma^2}\right). \quad (124)$$

Thus:

$$\begin{aligned} & E\left[\left(\frac{\lambda_k}{2}+\frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right] \\ = & 1-\frac{\exp\left(-\frac{1}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}\right)\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\cosh\left(\frac{\lambda_k}{2}+\frac{x}{\sigma^2}\right)}dx \end{aligned} \quad (125)$$

- **Derivation of** $E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right]$:

$E\left[\tanh\left(\frac{x_{k,1}(\tau_k)}{\sigma^2}\right)^2\right]$ is a special case of the previous result and can be obtained by taking $\lambda_k=0$.

## REFERENCES


[1] K. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers," vol. 24, no. 5, pp. 516–531, May 1976.

[2] F. M. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," *IEEE Trans. on Communications*, vol. 34, no. 5, pp. 423–429, May 1986.

[3] F. Gardner, *Demodulator Reference Recovery Techniques Suited for Digital Implementation*. Gardner Research Comp., 1988.

[4] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*. Dover Publications, Incorporated, 1991.

[5] J. Yang and B. Geller, "Near-optimum low-complexity smoothing loops for dynamical phase estimation - Application to BPSK modulated signals," *IEEE Trans. on Signal Processing*, vol. 57, no. 9, pp. 3704–3711, Sep. 2009.

[6] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. Plenum Press, New York and London, 1997.

[7] I. Nasr, B. Geller, L. Najjar, and S. Cherif, "Performance study of a near maximum likelihood code-aided timing recovery technique," *IEEE Trans. on Signal Processing*, vol. 64, no. 3, pp. 799–811, Feb. 2016.

[8] I. Nasr, L. Atallah, B. Geller, and S. Cherif, "CRB derivation and new code-aided timing recovery technique for QAM modulated signals," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 4901–4906.

[9] B. Mielczarek and A. Svensson, "Timing error recovery in turbo-coded systems on AWGN channels," *IEEE Trans. on Communications*, vol. 50, no. 10, pp. 1584–1592, Oct. 2002.

[10] A. Nayak, J. Barry, and S. McLaughlin, "Joint timing recovery and turbo equalization for coded partial response channels," *IEEE Trans. on Magnetics*, vol. 38, no. 5, pp. 2295–2297, Sep. 2002.

[11] R. Barry, A. Kavcic, S. McLaughlin, A. Nayak, and W. Zeng, "Iterative timing recovery," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 89–102, Jan. 2004.

[12] N. Wu, H. Wang, J. Kuang, and C. Yan, "Performance analysis of code-aided symbol timing recovery on AWGN channels," *IEEE Trans. on Communications*, vol. 59, no. 7, pp. 1975–1984, July 2011.

[13] L. Zhang and A. Burr, "APPA symbol timing recovery scheme for turbo codes," *PIMRC*, vol. 1, pp. 44–48 vol.1, Sep. 2002.

[14] P. Kovintavewat and J. Barry, "Exit chart analysis for iterative timing recovery," *IEEE Global Telecommunications Conference, GLOBECOM*, pp. 2435–2439, Nov. 2004.

[15] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "Turbo synchronization: an EM algorithm interpretation," *IEEE International Conference on Communications, ICC*, vol. 4, pp. 2933–2937, 2003.

[16] J. Yang, B. Geller, and S. Bay, "Bayesian and Hybrid Cramer-Rao bounds for the carrier recovery under dynamic phase uncertain channels," *IEEE Trans. on Signal Processing*, vol. 59, no. 2, pp. 667–680, Feb. 2011.

[17] F. Simoens and M. Moeneclaey, "Reduced complexity data-aided and code-aided frequency offset estimation for flat-fading MIMO channels," *IEEE Trans. on Wireless Communications*, vol. 5, no. 6, pp. 1558–1567, June 2006.

[18] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Communications*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[20] S. Bay, C. Herzet, J. M. Brossier, J. P. Barbot, and B. Geller, "Analytic and asymptotic analysis of Bayesian Cramer-Rao bound for dynamical phase offset estimation," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 61–70, Jan. 2008.

[21] F. Bellili, A. Methenni, and S. Affes, "Closed-form CRLBs for CFO and phase estimation from turbo-coded square-qam-modulated transmissions," *IEEE Trans. Wireless Communications*, vol. 14, no. 5, pp. 2513–2531, May 2015.

[22] T. Jesupret, M. Moeneclaey, and G. Ascheid, *Digital demodulator synchronization - Performance analysis*, ser. Final report ESTEC contract nr. 8437/89/NL/RE, June 1991, 1992.

[23] A. Masmoudi, F. Bellili, S. Affes, and A. Stephenne, "Closed-form expressions for the exact Cramer-Rao bounds of timing recovery estimators from BPSK, MSK and square-QAM transmissions," *IEEE Trans. on Signal Processing*, vol. 59, no. 6, pp. 2474–2484, June 2011.

[24] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*, J. G. Proakis Series, Ed. Wiley Series in Telecommunications and Signal Processing, 1998.

[25] J. Proakis, *Digital Communications*, ser. McGraw-Hill series in electrical and computer engineering : communications and signal processing. McGraw-Hill, 2001. [Online]. Available: https://books.google.tn/books?id=aUp2QgAACAAJ

[26] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Melbourne, FL, USA: Krieger Publishing Co., Inc., 1992.

[27] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[28] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. on Communications*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[29] B. Vucetic and J. Yuan, *Turbo Codes: Principles and Applications*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.

[30] F. Bellili, A. Stephenne, and S. Affes, "Cramer-Rao Lower Bounds for NDA SNR estimates of square QAM modulated transmissions," *IEEE Trans. on Communications*, vol. 58, no. 11, pp. 3211–3218, Nov. 2010.

[31] F. Bellili, N. Atitallah, S. Affes, and A. Stephenne, "Cramer-Rao Lower Bounds for Frequency and Phase NDA Estimation From Arbitrary

Square QAM-Modulated Signals," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4517–4525, Sep. 2010.
[32] F. Bellili, A. Methenni, and S. Affes, "Closed-form Cramer Rao lower bound for SNR estimation from turbo-coded BPSK-, MSK-, and square-QAM-modulated signals," *IEEE Trans. on Signal Processing*, vol. 62, no. 15, pp. 4018–4033, Aug. 2014.
[33] C. K. S. Pun, Y. C. Wu, S. C. Chan, and K. L. Ho, "On the design and efficient implementation of the Farrow structure," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 189–192, July 2003.
[34] B. Geller, V. Capellano, J. M. Brossier, A. Essebbar, and G. Jourdain, "Equalizer for video rate transmission in multipath underwater communications," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 2, pp. 150–156, Apr. 1996.

**Imen Nasr** received an Engineering and a Master Degree in Telecommunication from the Ecole Supérieure des Communications de Tunis - SUP'COM, University of Carthage, Tunisia, in 2011 and a Ph.D. degree in Information and Communications Technologies from both SUP'COM and the Ecole Nationale Supérieure de Techniques Avancées (ENSTA) ParisTech - Université de Paris Saclay, France in 2017. Her research activities include digital communication and signal processing for Wireless Body Area Networks (WBAN).

**Leila Najjar Atallah** received the engineering degree from Polytechnic School of Tunisia in 1997, a Master Degree in Automatic and Signal Processing from Ecole Supérieure d'Electricité Supélec, France in 1998, and a PhD in Sciences from University Paris XI in 2002. she joined the Higher School of Communications of Tunis (SUP'COM), University of Carthage, as an Assistant Professor in 2006, where she is currently an Associate Professor. She carries her research activities in the research laboratory Communications, Signal and Image (COSIM) in Sup'Com. Her current research interests are in the field of statistical signal processing and in signal processing for wireless communications. They include channel estimation, synchronization and localization. She is also interested in sparse regularization problems with application to energy efficient Wireless Sensor Networks.

**Sofiane Cherif** is professor in telecommunication engineering at the Higher School of Communications of Tunis (SUP'COM), University of Carthage. He received engineering, MS degrees, andPh.D. in Electrical Engineering from the "National Engineering School of Tunis (ENIT)", University of Tunis-El Manar, in 1990 and 1998, respectively and the "Habilitation Universitaire" in Telecommunication from SUP'COM, in 2007. From 2011 to 2014, he was head of the doctoral school in ICT, and presently, he is the head of SUP'COM and COSIM research Lab at SUP'COM. His current research interests are signal processing for communications, resource allocation and interference mitigation in wireless networks, wireless sensor networks and cognitive radio.

**Benoit Geller** (SM) received the Telecommunications engineering master degree from ENSEIRB and the PhD in Telecommunications from INP Grenoble in 1992. He was the head of the Multisensor and Information Team (TIM, 17 permanent researchers) at SATIE lab - ENS Cachan until he joined ENSTA ParisTech - Université de Paris Saclay, in 2007 where he is currently a Full Professor. He works on wireless networks and on the Internet of Things, with a special focus on iterative methods applied to digital communications. He has been involved in many European and industrial projects and has published about 100 international publications.