A Non-Data-Aided Maximum Likelihood Time Delay Estimator Using Importance Sampling

Ahmed Masmoudi, Faouzi Bellili, Sofiène Affes, *Senior Member, IEEE*, and Alex Stéphenne, *Senior Member, IEEE*

Abstract—In this paper, we present a new time delay maximum likelihood estimator based on importance sampling (IS). We show that a grid search and lack of convergence from which most iterative estimators suffer can be avoided. It is assumed that the transmitted data are completely unknown at the receiver. Moreover the carrier phase is considered as an unknown nuisance parameter. The time delay remains constant over the observation interval and the received signal is corrupted by additive white Gaussian noise (AWGN). We use importance sampling to find the global maximum of the compressed likelihood function. Based on a global optimization procedure, the main idea of the new estimator is to generate realizations of a random variable using an importance function, which approximates the actual compressed likelihood function. We will see that the algorithm parameters affect the estimation performance and that with an appropriate parameter choice, even over a small observation interval, the time delay can be accurately estimated at far lower computational cost than with classical iterative methods.

Index Terms—Cramér–Rao lower bound (CRLB), Monte Carlo methods, non-data-aided (NDA) estimation, optimization methods, symbol timing recovery.

I. INTRODUCTION

PARAMETER estimation is a crucial operation for any digital receiver; in particular the recovery of time delay introduced by the channel. Typically, in network communications, the time delay is usually assumed to be confined within the symbol duration [1]. Particularly, symbol timing recovery allows for sampling the signal at accurate time instants in order to achieve satisfactory performances. The key task of timing recovery consists in determining the time instants at which the received signal should be sampled in order to perform reliable data recovery. However, in many other applications such as radar or sonar systems [2], [3], where it can exceed the symbol duration, the time delay is used to localize targets.

During the last few decades, many time-delay estimators have been developed trying to achieve the well-known Cramér–Rao lower bound (CRLB). A key step in time recovery schemes is the determination of an objective function from the statistics of the received signal from which an estimate of the time delay can be extracted. According to the criterion under which this function is derived, time delay estimation techniques are classified as [4]: minimum mean-square error (MMSE) schemes, zero-forcing (ZF) schemes, early-late schemes and maximum-likelihood schemes. In this sense, it is known that the maximum-likelihood estimator is an asymptotically efficient estimator, and that it performs close to the CRLB at relatively high signal-to-noise ration (SNR) values [7], even for short data records. Therefore, it has been subject to intense research. In the case of data-aided transmissions, where the transmitted data are *a priori* completely known, an expression for the global maximum of the log-likelihood function is analytically tractable. However, when the transmitted data are completely unknown (i.e., the parameter of interest should be blindly estimated), the log-likelihood function becomes extremely nonlinear and it is difficult to analytically find its global maximum. In this case, maximum-likelihood (ML) solutions must be numerically tackled. The grid search technique is the most basic alternative to numerically find the maximum of the nonlinear likelihood function. Unfortunately, this technique can be used only if the range of the parameter is confined to a finite interval, otherwise, iterative maximization procedures must be envisaged. The most famous iterative procedures are the Newton–Raphson method [5] and the expectation-maximization algorithm [6]. However, these two prominent methods are known to converge to the ML solution only if the initial guess is close enough to the true unknown parameter value. If not, these iterative algorithms may converge to a local maximum of the likelihood function, or even diverge. To circumvent this problem, these algorithms may use many initial values to improve their performance, but this increases in counterpart their computational complexity without even ultimately warranting their convergence to the global maximum.

In this work, we resort to an entirely different approach for the estimation of the time delay parameter. The compressed likelihood function is derived considering the transmitted symbols as unknown but deterministic. Based on this function, an iterative algorithm earlier implemented in [8] performs better in the high SNR region than the low-SNR unconditional ML (UML) timing error detectors (TEDs) [1], but its performance still depends on the initialization value making it therefore prone to severe degradation due convergence uncertainty.

Motivated by these facts, we develop in this paper a new noniterative approach to find the time delay conditional maximum-likelihood (CML) estimates. We implement the CML

algorithm in a noniterative way. We avoid the grid search, essential in traditional iterative approaches, by using the importance sampling technique which has been shown to be a powerful tool in performing NDA ML estimation. In fact, this method was successfully applied to estimate other crucial parameters such as the direction of arrival (DOA) [9], the carrier frequency [10] or the joint DOA-Doppler frequency [11]. The importance sampling technique is used in this paper in the context of time delay estimation. Moreover, we adopt the discrete-time model widely used in the field of sensors array processing [12] and more recently formulated in the context of time-delay estimation [8]. The resulting IS-based estimator attains the modified CRLB (MCRLB) over both the medium and high SNR regions, whereas the traditional UML TED, being derived under the assumption of low SNR, does not approach the MCRLB at the high SNR region.

The remainder of this paper is organized as follows. In Section II, we present the discrete-time signal model that will be used throughout this paper. We derive the compressed likelihood function in Section III. In Section IV, we introduce the importance sampling method that will be used in this article to find the global maximum of the compressed likelihood function. Section V deals with the choice of the importance function and discusses the impact of some parameters on the estimator performance. The newly proposed algorithm is developed in Section VI. Simulation results are discussed in Section VII and, finally, some concluding remarks are drawn out in Section VIII.

II. DISCRETE-TIME SIGNAL MODEL

First, we present a list of notations and definitions that will be used in this paper.

$E_x\{.\}$	The expectation with respect to x.
$\|.\|$	Euclidean norm.
$(.)^T, (.)^H$	Transposition and conjugate transposition.
SNR	Signal-to-noise ratio.
IS	Importance sampling.
ML	Maximum likelihood.
CML	Conditional maximum likelihood.
MCRLB	Modified Cramér–Rao lower bound.
QAM	Quadrature amplitude modulation.
PAM	Pulse-amplitude modulation.

Consider a traditional communication system where on one hand the channel delays the transmitted signal and on the other hand an AWGN with an overall power of N_0 corrupts the received signal as follows:

$$y(t) = \sqrt{E_s}x(t-\tau^*)e^{j\theta} + w(t) \qquad (1)$$

where τ^* is the unknown time delay to be estimated, θ is the unknown but deterministic channel distortion phase, $w(t)$ is an additive white Gaussian noise (AWGN) with independent real and imaginary parts, each of variance $\frac{N_0}{2}$ and $\sqrt{E_s}$ is the signal

amplitude. The unknown transmitted signal $x(t)$ is modeled as follows:

$$x(t) = \sum_{i=0}^{K-1} c_i h(t-iT) \qquad (2)$$

where K is the number of transmitted symbols in the observation interval, $\{c_i\}_{i=0}^{K-1}$ are the unknown complex-valued symbols, $h(t)$ is the shaping pulse of energy E_h and T is the symbol's duration.

In the sequel, we outline the discrete-time signal model which was proposed for the first time in [8] to derive an iterative CML timing recovery algorithm. The received signal $y(t)$ is passed through an ideal lowpass filter of bandwidth $\frac{F_s}{2}$ and sampled at a frequency $F_s = \frac{1}{T_s} = \frac{k}{T}$, where k is a given integer which guarantees that F_s is above the Nyquist rate. Then, the received samples $\boldsymbol{y} = [y(0), y(T_s), y(2T_s), \ldots, y((M-1)T_s)]^T$ can be written in a matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{A}_{\tau^*}\boldsymbol{x} + \boldsymbol{w} \qquad (3)$$

where M is the number of samples of $y(t)$ and $\boldsymbol{w}$ and $\boldsymbol{A}_{\tau^*}$ are defined as follows:

$$\boldsymbol{w} = [w(0), \ w(T_s), \ \ldots, \ w((M-1)T_s)]^T \qquad (4)$$
$$\boldsymbol{A}_{\tau^*} = [\boldsymbol{a}_0(\tau^*), \ \boldsymbol{a}_1(\tau^*), \ \ldots, \ \boldsymbol{a}_{K-1}(\tau^*)] \qquad (5)$$

with

$$\boldsymbol{a}_i(\tau^*) = [h(-iT - \tau^*), \ h(T_s - iT - \tau^*), \ \ldots,$$
$$h((M-1)T_s - iT - \tau^*)]^T. \qquad (6)$$

In (3), $\boldsymbol{x}$ is the set of unknown data and signal phase, which is given by

$$\boldsymbol{x} = \boldsymbol{c}e^{j\theta} = [c_0, \ c_1 \ \ldots \ c_{K-1}]^T \ e^{j\theta}. \qquad (7)$$

Moreover, the covariance matrix of $\boldsymbol{w}$ is given by

$$\boldsymbol{C}_w = 2\sigma^2 \boldsymbol{I}_M = N_0 F_s \boldsymbol{I}_M \qquad (8)$$

where $\boldsymbol{I}_M$ refers to the $(M \times M)$ identity matrix and $2\sigma^2 = N_0 F_s$. The sampled data $\boldsymbol{y}$ is a linear function of the vector $\boldsymbol{x}$ but depends nonlinearly on the time delay τ^*. We mention that the model of (3) presented in [8] is inspired from the model widely used in array signal processing where each column of the transfer matrix is a function of a different parameter, usually the direction-of-arrival or the frequency of each incoming signal. In the context of time delay estimation, the entire matrix $\boldsymbol{A}_{\tau^*}$ depends on the same parameter τ^*.

III. LIKELIHOOD FUNCTION

The conditional likelihood function of the observed data $\boldsymbol{y}$ is given by

$$\Lambda(\boldsymbol{y}|\boldsymbol{x};\tau) \propto p(\boldsymbol{y}|\boldsymbol{x};\tau) = C \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}_\tau\boldsymbol{x}\|^2\right\}, \qquad (9)$$

where $p(\boldsymbol{y}|\boldsymbol{x};\tau)$ is the probability density function (pdf) of $\boldsymbol{y}$ conditioned on $\boldsymbol{x}$ and parameterized by τ, and C is a pos-

itive constant which does not depend on the time delay and therefore will be dropped, without loss of generality. Note here that τ is any possible value of the time delay parameter τ^* and that $\Lambda(\boldsymbol{y}|\boldsymbol{x};\tau)$ attains its maximum at $\tau = \tau^*$, i.e., $\tau^* = \arg\max_{\tau} \Lambda(\boldsymbol{y}|\boldsymbol{x};\tau)$.

Actually, one needs to maximize $\Lambda(\boldsymbol{y}|\boldsymbol{x};\tau)$ with respect to τ in order to find the ML solution $\widehat{\tau^*}$. However, (9) imposes a joint estimation of $\boldsymbol{x}$ and τ^*, which is very difficult to perform. Therefore, two principal approaches are developed in the literature in order to obtain a likelihood function that depends only on τ. On one side, the unconditional maximum-likelihood (UML) estimator introduced in [1] considers the data symbols as random and hence averages the joint likelihood function over $\boldsymbol{x}$ to obtain a function that depends only on the time delay as follows:

$$\Lambda(\boldsymbol{y}|\tau) = E_{\boldsymbol{x}}\{\Lambda(\boldsymbol{y}|\boldsymbol{x};\tau)\}. \quad (10)$$

On the other side, the data symbols are modeled as unknown but deterministic in the formulation of the conditional likelihood function. Therefore, $\widehat{\boldsymbol{x}}$, the solution that maximizes (9) with respect to $\boldsymbol{x}$, for a given τ is used in (9) as a substitute of $\boldsymbol{x}$. Actually, $\widehat{\boldsymbol{x}}$ which maximizes $\Lambda(\boldsymbol{y};\boldsymbol{x},\tau)$ also maximizes the log-likelihood function given by

$$L(\boldsymbol{y};\boldsymbol{x},\tau) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}_\tau\boldsymbol{x}\|^2. \quad (11)$$

Therefore, taking the gradient of $L(\boldsymbol{y}|\boldsymbol{x};\tau)$ with respect to $\boldsymbol{x}$ and setting it to zero,

$$\frac{\partial L(\boldsymbol{y};\boldsymbol{x},\tau)}{\partial \boldsymbol{x}} = -\frac{1}{\sigma^2}\left(\boldsymbol{A}_\tau^T\boldsymbol{y} - \boldsymbol{A}_\tau^T\boldsymbol{A}_\tau\boldsymbol{x}\right) = 0, \quad (12)$$

yields the following result:

$$\widehat{\boldsymbol{x}} = \left(\boldsymbol{A}_\tau^T\boldsymbol{A}_\tau\right)^{-1}\boldsymbol{A}_\tau^T\boldsymbol{y} = \boldsymbol{A}_\tau^{\#}\boldsymbol{y} \quad (13)$$

where $\boldsymbol{A}_\tau^{\#} = \left(\boldsymbol{A}_\tau^T\boldsymbol{A}_\tau\right)^{-1}\boldsymbol{A}_\tau^T$ is the pseudo-inverse of the matrix $\boldsymbol{A}_\tau$. Substituting $\widehat{\boldsymbol{x}}$ into (11), one obtains the so-called compressed likelihood function, that depends only on the unknown time delay parameter

$$L(\boldsymbol{y};\tau,\widehat{\boldsymbol{x}}) = -\frac{1}{2\sigma^2}\boldsymbol{y}^H\left(\boldsymbol{I}_K - \boldsymbol{A}_\tau\boldsymbol{A}_\tau^{\#}\right)\boldsymbol{y} \quad (14)$$

which can be further simplified by dropping the constant terms to obtain the useful compressed likelihood function denoted by $L_c(\boldsymbol{y};\tau)$ as follows:

$$L_c(\boldsymbol{y};\tau) = \boldsymbol{y}^H\boldsymbol{A}_\tau\left(\boldsymbol{A}_\tau^T\boldsymbol{A}_\tau\right)^{-1}\boldsymbol{A}_\tau^T\boldsymbol{y}. \quad (15)$$

Note that the expression in (15) represents the cross-energy between the pseudo-inverse filter $\boldsymbol{A}_\tau^{\#}$ and the sampled matched filter $\boldsymbol{A}_\tau^T$. For τ equal to the timing parameter to be estimated, the filter $\boldsymbol{A}_\tau^{\#}$ becomes a zero-forcing equalizer since the components of $\boldsymbol{A}_\tau^{\#}\boldsymbol{y}$ are intersymbol interference (ISI)-free (i.e., $\boldsymbol{A}_\tau^{\#}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{A}_\tau^{\#}\boldsymbol{w}$, see [8]).

IV. GLOBAL MAXIMIZATION OF THE COMPRESSED LIKELIHOOD FUNCTION

To perform maximum-likelihood estimation, we have to maximize (15) with respect to τ. Unfortunately, a closed-form expression for this optimization problem is not analytically

tractable since the objective function in (15) is extremely nonlinear with respect to τ. Therefore, many methods have been developed to numerically find the maximum, but most of them are iterative [1]–[8]. We cannot deny that these methods provide good performance in terms of error variance, but unfortunately they require, in counterpart, a sufficiently close initial guess to converge to the global maximum of the likelihood function. Otherwise, the result may be a local maximum, which does not correspond to the true time delay value. This is why a suboptimal algorithm needs to be applied firstly and then its output is considered as an initial value for any iterative technique.

To avoid this challenging drawback of iterative methods, we propose in this paper an entirely different technique which does not claim any initial guess of the time delay parameter. We apply the global maximization method earlier proposed by Pincus [13] which provides a powerful tool for accomplishing nonlinear optimization and guarantees finding the global maximum without any initialization concerns. In fact, the theorem of Pincus states that the maximum of $L_c(\boldsymbol{y};\tau)$ is given by

$$\widehat{\tau^*} = \lim_{\rho\to\infty}\int_J \tau L'_{c,\rho}(\tau)d\tau \quad (16)$$

where

$$L'_{c,\rho}(\tau) = \frac{\exp\{\rho L_c(\boldsymbol{y};\tau)\}}{\int_J \exp\{\rho L_c(\boldsymbol{y};\tau)\}d\tau} \quad (17)$$

can be viewed as the normalized function of $\exp\{\rho L_c(\boldsymbol{y};\tau)\}$. Note that in (16) and (17), J is the integration interval in which τ is supposed to be confined. In a certain way, $L'_{c,\rho}(\tau)$ can be viewed as a pdf (since it verifies all the properties of a pdf), but since τ is actually deterministic, $L'_{c,\rho}(\tau)$ is more conveniently called a pseudo-pdf [9]. It is also worth noting that, as $\rho \to \infty$, $L'_{c,\rho}(\tau)$ becomes a Dirac delta function centered at the location of its original maximum. We leave broad details on this point in Appendix A.

The ML estimator for the time delay parameter, obtained from the location of the global maximum of $L_c(\boldsymbol{y};\tau)$ is given, for a large value of ρ_0, by

$$\widehat{\tau^*} = \int_J \tau L'_{c,\rho_0}(\tau)d\tau. \quad (18)$$

Now, we need to evaluate the integral given in (18), although a direct integration remains always difficult if not impossible. However, this integral is in a way the mean value of a random variable distributed according to $L'_{c,\rho_0}(.)$. It was shown in [14] that this type of integral can be efficiently evaluated using Monte Carlo simulations as follows:

$$\widehat{\tau^*} = \frac{1}{R}\sum_{k=1}^R \tau_k \quad (19)$$

where $\{\tau_k\}_{k=1}^R$ are realizations of τ distributed according to the pseudo-pdf, $L'_{c,\rho_0}(\tau)$, and hence the global maximization problem reduces simply to a generation of random variables. Yet, since it is a nonlinear function of τ, the direct generation of realizations according to $L'_{c,\rho_0}(\tau)$ is computationally hard. Thus, instead of pursuing a fruitless path, we use the importance

sampling technique, as done in [9]–[11] for the estimation of the signal directions of arrival, the carrier frequency and the Doppler frequencies, instead of directly using (17).

V. THE IMPORTANCE SAMPLING TECHNIQUE

It has been shown that the importance sampling technique is a powerful tool to compute multiple integrals; in particular the one given in (18). In fact, it can be easily seen that for any function[1] $f(.)$:

$$\int_J f(\tau) L'_{c,\rho_0}(\tau) d\tau = \int_J f(\tau) \frac{L'_{c,\rho_0}(\tau)}{g'(\tau)} g'(\tau) d\tau \qquad (20)$$

where $g'(.)$, called the normalized importance function, is another pseudo-pdf which must be chosen as a simple function of τ so that realizations distributed according to $g'(.)$ can be easily generated. Then, the Monte Carlo method is used to empirically compute the integral in (20) simply via the following summation:

$$\int_J f(\tau) L'_{c,\rho_0}(\tau) d\tau = \frac{1}{R} \sum_{k=1}^R f(\tau_k) \frac{L'_{c,\rho_0}(\tau_k)}{g'(\tau_k)} \qquad (21)$$

where τ_k is the kth realization of τ according to the normalized importance function $g'(.)$ and R is the number of realizations. Typically, $g'(.)$ and $L'_{c,\rho_0}(.)$ should be very similar to reduce the variance of the estimates. However, $L'_{c,\rho_0}(.)$ remains a complex function and in counterpart $g'(.)$ needs to be as simple as possible. Therefore, some tradeoffs must be found in the construction of the importance function. In fact, the inverse matrix $\left(A_\tau^T A_\tau \right)^{-1}$ in the actual compressed likelihood function, $L_c(y; \tau)$ (or equivalently $L'_{c,\rho_0}(.)$), is very nonlinear with respect to τ. Intuitively, one can replace this inverse matrix by the diagonal matrix $\frac{T_s}{E_h} I_K$. Hence, a reasonable approximation of the compressed likelihood function is

$$L_c(y; \tau) \approx \frac{T_s}{E_h} y^H A_\tau A_\tau^T y. \qquad (22)$$

The approximation of $\left(A_\tau^T A_\tau \right)^{-1}$ with $\frac{T_s}{E_h} I_K$ is very reasonable for most of the conventional pulse shaping functions. For instance, it can be verified that for the widely used square root-raised cosine pulse, the diagonal elements of $\left(A_\tau^T A_\tau \right)^{-1}$ are dominant compared to its off-diagonal ones. In fact, as defined in (5), the columns of A_τ are built upon shifted versions of the shaping pulse $h(.)$, therefore every element of $A_\tau^T A_\tau$ can be seen as the convolution of two shifted versions of $h(.)$ (the shift being an integer multiple of T), which value is maximum when the shift is the same, i.e., in the diagonal elements. Whereas, when the shift is not the same, the value of the convolution is very low. See Appendix B for more details about this observation. In the particular case where the pulse shape does not generate inter-symbol interference, the approximation becomes

strict equality and (22) yields the exact compressed likelihood function. Then, a reasonable importance function is given by

$$g_{\rho_1}(\tau) = \exp\left\{ \rho_1 \sum_{k=0}^{K-1} \frac{T_s}{E_h} \left| \sum_{i=0}^{M-1} y^*(iT_s) h(iT_s - kT - \tau) \right|^2 \right\} \qquad (23)$$

where ρ_1 is another constant different[2] from ρ_0. Note that the normalization of $g_{\rho_1}(.)$ by $\int_J g_{\rho_1}(x) dx$ yields the normalized importance function $g'_{\rho_1}(.)$ (i.e., $g'_{\rho_1}(\tau) = \frac{g_{\rho_1}(\tau)}{\int_J g_{\rho_1}(x) dx}$). But since the periodogram of the data evaluated at the time delay τ, $I_k(\tau)$, is given by

$$I_k(\tau) = \left| \sum_{i=0}^{M-1} y^*(iT_s) h(iT_s - kT - \tau) \right|^2 ,$$
$$k = 0, 1, \ldots, K-1 \qquad (24)$$

then, we rewrite the importance function as follows:

$$g_{\rho'_1}(\tau) = \prod_{k=0}^{K-1} \exp\left\{ \rho'_1 I_k(\tau) \right\} \qquad (25)$$

with

$$\rho'_1 = \frac{T_s}{E_h} \rho_1. \qquad (26)$$

The normalization of (25) leads to the pseudo-pdf $g'(.)$ which will be used, hereafter, to generate the realizations involved in (21):

$$g'_{\rho'_1}(\tau) = \frac{\prod_{k=0}^{K-1} \exp\left\{ \rho'_1 I_k(\tau) \right\}}{\int_J \prod_{k=0}^{K-1} \exp\left\{ \rho'_1 I_k(v) \right\} dv}. \qquad (27)$$

It is also worth noting that the performance of the new maximum-likelihood estimator depends on the choice of ρ'_1. In fact, our ultimate goal is to find the global maximum of the function $L_c(y; \tau) = y^H A_\tau (A_\tau^T A_\tau)^{-1} A_\tau^T y$. However, this function exhibits many local maxima even in the total absence of noise, and it is often difficult to distinguish between the global and a local maximum. For this purpose, ρ'_1 is chosen to render the objective function in (27) more peaked around its global maximum which will have a relatively higher peak compared to the local maxima. This behavior is illustrated in Fig. 1, which plots the function $g'_{\rho'_1}(.)$ for $\rho'_1 = 10$ and $\rho'_1 = 20$, in the total absence of the additive noise. Moreover, we show in Appendix C how this parameter renders $g'_{\rho'_1}(.)$ more peaked around its global maximum. Based on this fact, it can be stated, a priori, that it is better to arbitrarily increase ρ'_1 in order to achieve better performance. However, this is much easier said than done since, in practice, this leads to numerical overflows. Actually, the best value of ρ'_1 is the highest possible without resulting in any overflow in the computation of $g'_{\rho'_1}(.)$ as it will be seen in Section VII. Same argument is valid for

---

[1]In our case, we have $f(\tau) = \tau$.

[2]In our case, ρ_1 can be equal to ρ_0, unlike for the multiple parameters estimation where ρ_1 should be different from ρ_0.
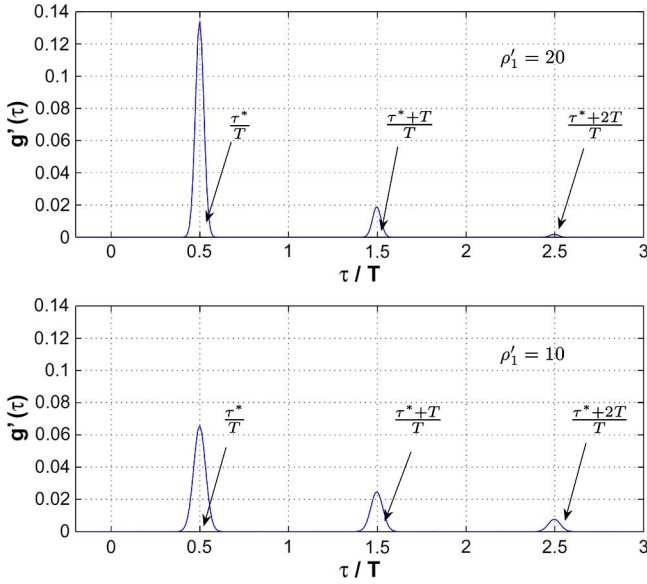
Fig. 1. Plot of $g'_{\rho'_1}(.)$ for $\rho'_1 = 20$ and $\rho'_1 = 10$ using a root-raised cosine pulse and for $K = 100$.

ρ_0. In fact, the approximation of $(\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau)^{-1}$ by $\frac{T_s}{E_h} \boldsymbol{I}_K$ means that the quantity $L_c(\boldsymbol{y}; \tau) = \boldsymbol{y}^H \boldsymbol{A}_\tau (\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau)^{-1} \boldsymbol{A}_\tau^T \boldsymbol{y}$ is almost equal to $\frac{T_s}{E_h} L_c^{(a)}(\tau)$ where $L_c^{(a)}(\tau) = \boldsymbol{y}^H \boldsymbol{A}_\tau \boldsymbol{A}_\tau^T \boldsymbol{y}$ (note that the superscript (a) in $L_c^{(a)}(\tau)$ refers to the word "approximate" where we approximate $L_c(\boldsymbol{y}; \tau)$ with $L_c^{(a)}(\tau)$ by replacing $(\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau)^{-1}$ by $\frac{T_s}{E_h} \boldsymbol{I}_K$). This means that ρ_0 will results in an overflow in $\exp\{\rho_0 L_c(\boldsymbol{y}; \tau)\}$ as far as ρ_1 results in an overflow in $\exp\left\{\rho_1 \frac{T_s}{E_h} L_c^{(a)}(\tau)\right\}$. Therefore, the optimal value, ρ_0^{opt}, of ρ_0 verifies $\rho_0^{opt} = \rho_1^{opt}$ where ρ_1^{opt} is again the optimal value of $\rho_1 = \frac{E_h}{T_s} \rho'_1$.

We also note that the matrix $\boldsymbol{A}_\tau$ in (22) exhibits an interesting structure since its columns are simply shifted versions of the pulse shape. Hence, the matrix-by-vector operation $\boldsymbol{A}_\tau^T \boldsymbol{y}$ can be viewed as a filtering operation of the received samples $\boldsymbol{y}$ with a filter $h(.)$ whose coefficients are the central row of $\boldsymbol{A}_\tau^T$. Therefore, the computation of (27) is quite simple and realizations distributed according to $g'_{\rho'_1}(.)$ in (27) can be easier generated than according to $L'_{c,\rho_0}(.)$ in (17).

VI. ESTIMATION OF THE TIME DELAY

In the sequel, two scenarios will be proposed depending on the range of the unknown time delay parameter. In the first scenario, we assume that the time delay parameter takes values within $[0, PT]$, where P is a strictly positive integer (i.e., $P \geq 1$). In the second scenario, we assume that the time delay parameter takes values within $[0, T]$.

A. First Scenario: $\tau^* \in [0, PT]$

As already mentioned in the introduction, in many applications such as radar or sonar transmissions, the actual time delay introduced by the channel may exceed the symbol's duration. In this subsection, we assume however that the time delay does not exceed PT, where P is a given strictly positive integer,

i.e., $\tau^* \in [0, PT]$. This upper limitation of the interval is justified since, in each communication system, we always have an *a priori* idea about the maximum range[3] of τ. As we have seen in the previous section, the maxima of $g'_{\rho'_1}(.)$ are periodic, with period T. Therefore, many secondary peaks may appear which ultimately affects the estimate $\widehat{\tau^*}$ of τ^*. In fact, to obtain unbiased estimates of τ^*, the expected value of the estimation error $\tau_e = \tau^* - \widehat{\tau^*}$ should be equal to zero, i.e.,

$$\mathrm{E}\{\tau_e\} = \mathrm{E}\{\tau^* - \widehat{\tau^*}\} = 0. \tag{28}$$

However, it may occur that the difference between τ^* and $\widehat{\tau^*}$ is very important. In fact, to simplify, assume that $g'_{\rho'_1}(.)$ has only two peaks and neglect the others. Then the generated values will take values around τ^* and $T + \tau^*$, with higher probability around τ^* where the highest peak is located. If we denote by C_1 the set of realizations taking values near τ^* and C_2 the set of realizations taking values near $\tau^* + T$, then from (21) the estimated, $\widehat{\tau^*}$, can be approximated by

$$\widehat{\tau^*} = \frac{\mathrm{card}(C_1)}{R} \tau^* + \frac{\mathrm{card}(C_2)}{R} (\tau^* + T) \tag{29}$$

with $\mathrm{card}(C)$ denoting the cardinal of C, i.e., the number of elements of C. Therefore, $\widehat{\tau^*} \neq \tau^*$ since $R = \mathrm{card}(C_1) + \mathrm{card}(C_2)$ and $\mathrm{card}(C_2)$ is always not equal to zero. Moreover, the bias is larger at low SNRs and/or short data records. This property was also previously observed in the case of frequency estimation in [15].

To circumvent this problem, the pseudo-pdf, $g'_{\rho'_1}(.)$, must be centered around τ^*. To that end, two intuitive methods may be envisioned. We may either eliminate the secondary peaks to keep only the principal one, or we can generate other peaks in a way that the number of secondary lobes on either side of the principal lobe is the same. The first idea seems to be the most efficient, but it is unfortunately unrealizable and we opt for the second alternative. Indeed, as we have seen, the estimation bias stems from the peaks taking place after the principal lobe. Thus, we have to modify $g'_{\rho'_1}(.)$ so that it becomes quasi-symmetric around τ^*. To that end, the simplest way is to suppose, virtually, that τ takes negative values although τ is always positive. We extend the interval of definition of $g'_{\rho'_1}(.)$ from $[0, PT]$ to $[-QT, PT]$, where Q is a positive integer smaller than P. In this way, virtual secondary lobes appear before as well as after τ^*. Moreover, as it can be seen from Fig. 2, the probability of generating realizations around $\tau^* + iT$ is almost the same as the one of generating realizations around $\tau^* - iT$. Hence, the estimator becomes unbiased and the estimate $\widehat{\tau^*}$ is more accurate. So far, we have established an unbiased estimator based on a linear average of the generated realizations. But through simulations, we noted a performance change according to the constellation type. In fact, for a constant-envelope constellation such as phase-shift keying (PSK), the estimator works perfectly. However, its performance degrades dramatically for nonconstant-modulus constellations such as pulse-amplitude modulation (PAM), quadrature amplitude modulation (QAM), etc. In fact, as we have previously seen, the main problem that faces

---

[3]Note that P can be always chosen as large as desired to ensure that $\tau^* \in [0, PT]$.
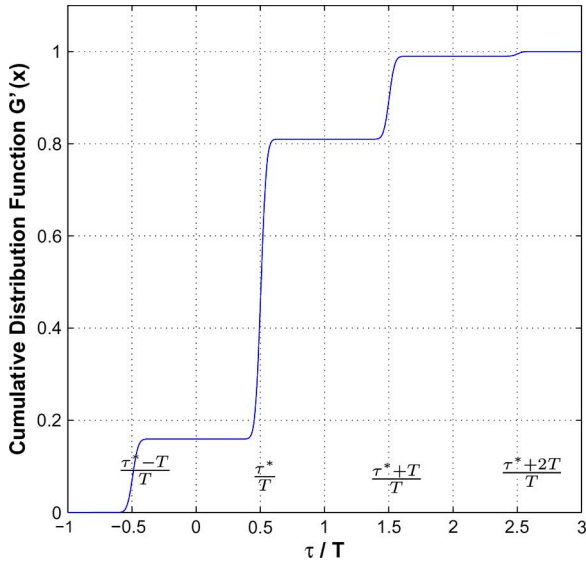
Fig. 2. Plot of the cumulative distribution function (CDF), $G'(\tau)$ whose pdf is $g'(\tau)$, $\mathrm{SNR} = 5$ dB.

the new ML estimator is the presence of secondary peaks. Although we have explained, in Section V, how to reduce the adverse effects of these peaks, they generate irreversible errors in the case of nonconstant-envelope constellations. To simplify the problem, without loss of generality, we suppose again that $g'_{\rho'_1}(.)$ exhibits only two peaks, one located at τ^* and the other located at $\tau^* + T$. From (25), $g_{\rho'_1}(\tau^*)$ and $g_{\rho'_1}(\tau^* + T)$ can be written as follows:

$$g_{\rho'_1}(\tau^*) = \exp\left\{\rho'_1 \sum_{k=1}^{K-1} \left|\sum_{i=0}^{M-1} y^*(iT_s)h(iT_s - kT - \tau^*)\right|^2\right\}$$
$$\times \exp\left\{\rho'_1 \left|\sum_{i=0}^{M-1} y^*(iT_s)h(iT_s - \tau^*)\right|^2\right\} \quad (30)$$

and

$$g_{\rho'_1}(\tau^* + T) = \exp\left\{\rho'_1 \sum_{k=1}^{K-1} \left|\sum_{i=0}^{M-1} y^*(iT_s)h(iT_s - kT - \tau^*)\right|^2\right\}$$
$$\times \exp\left\{\rho'_1 \left|\sum_{i=0}^{M-1} y^*(iT_s)h(iT_s - KT - \tau^*)\right|^2\right\}. \quad (31)$$

Noting that the samples of the received signal are limited in time to $[0, KT]$ and the magnitudes of $h(t - KT - \tau)$ for $t \in [0, KT]$ are very small compared to those of $h(t - KT - \tau)$ for $t \in [KT, (K+1)T]$, then, the term $\sum_{i=0}^{M-1} y^*(iT_s)h(iT_s - KT - \tau)$ can be neglected. Considering this result, we express $g_{\rho'_1}(\tau^*)$ as a function of $g_{\rho'_1}(\tau^* + T)$:

$$g_{\rho'_1}(\tau^*) \approx g_{\rho'_1}(\tau^* + T) \exp\left\{\rho'_1 I_0(\tau^*)\right\} \quad (32)$$

where $I_0(\tau^*)$ depends mainly on the amplitude of the first symbol. In practice, it may occur that the amplitude of the first transmitted symbol is the smallest one. In this case, the contribution of $\exp\{\rho'_1 I_0(\tau^*)\}$ in $g_{\rho'_1}(\tau^*)$ is far less important

than the other terms, i.e., $\exp\{\rho'_1 I_0(\tau^*)\} \ll \exp\{\rho'_1 I_i(\tau^*)\}$ for $i = 1, 2, \ldots, K - 1$. As a result, $g_{\rho'_1}(\tau^*)$ will be closer to $g_{\rho'_1}(\tau^* + T)$, which is a local maximum making the estimate $\widehat{\tau^*}$ shift toward $\tau^* + T$. The same problem occurs when the last transmitted symbol has the smallest amplitude with the only difference that the shift will be toward $\tau^* - T$.

To avoid these problems, $I_0(\tau^*)$ must be as large as possible. To that end, we slightly modify the algorithm in the case of non-constant-modulus constellations by sending two *a priori* known symbols: one at the beginning and one at the end of the frame. Moreover, these two known symbols must be of highest energy among the constellation points. Then, $I_0(\tau^*)$ is no longer negligible compared to $I_i(\tau^*)$, for $i \neq 0$, and the difference between the magnitude of $g_{\rho'_1}(\tau^*)$ and $g_{\rho'_1}(\tau^* + T)$ is large enough to avoid an important detection error. The same thing holds for $g_{\rho'_1}(\tau^*)$ and $g_{\rho'_1}(\tau^* - T)$.

B. Second Scenario: $\tau^* \in [0, T]$

In many cases, the time delay does not exceed the symbol duration T. Therefore, we must look for the global maximum only in $[0, T]$. As previously explained, the maxima of the importance function are periodically located, with a period equal to T. Moreover, since we know *a priori* that τ does not exceed T, then we can more conveniently use the circular[4] (instead of the linear) mean to evaluate the mean in (18). It will be seen in Section VII that the use of the circular mean provides considerable performance enhancements in the low-SNR region. As it will be explained later, the use of the circular mean considerably reduces the computational cost.

To introduce the concept of a circular mean, consider a circular random variable which takes values in a finite interval that can be mapped into the unit circle. For instance, let α be a random variable defined in [0,1] with pdf $P(\alpha)$. Then, the circular mean of α is defined as:

$$E_c\{\alpha\} = \frac{1}{2\pi} \angle \int \exp\{j2\pi\alpha\} P(\alpha) d\alpha \quad (33)$$

where $\angle$ denotes the angle in radians. Having R realizations of α, its circular mean is [16]

$$\widehat{\alpha} = \frac{1}{2\pi} \angle \frac{1}{R} \sum_{r=1}^{R} \exp\{j2\pi\alpha_r\}. \quad (34)$$

In our case, if the time delay is not confined within the interval [0,1], it can be easily transposed into this interval by normalizing τ^* by T. Then, the resulting transposed estimate is inversed to obtain an estimate in the original interval. Hence, the IS estimate of τ^* using (34) and (21) is

$$\widehat{\tau^*} = \frac{T}{2\pi} \angle \frac{1}{R} \sum_{k=1}^{R} \frac{L'_c(\tau_i)}{g'(\tau_i)} \exp\left\{j\frac{2\pi\tau_k}{T}\right\} \quad (35)$$

or, finally,

$$\widehat{\tau^*} = \frac{T}{2\pi} \angle \frac{1}{R} \sum_{i=1}^{R} F(\tau_i) \exp\left\{j\frac{2\pi\tau_i}{T}\right\} \quad eqno(36)$$

---

[4]Note that the circular mean cannot be used in the first scenario when τ may exceed T since it always returns an estimate in $[0, T]$ by virtually bringing, into this interval, all the secondary lobes of the normalized importance function.

where

$$F(x) = \frac{L'_{c,\rho_0}(x)}{g'_{\rho'_1}(x)}. \tag{37}$$

Note that we need to find the angle of a complex number, and thus, we can remove any positive real factor taking place in (36) without affecting the final result. This means that the two strictly positive normalization constants $\int_J L'(v)dv$ and $\int_J g'(v)dv$ can be simply dropped. Moreover, an overflow may occur since both the numerator and the denominator are exponentials. To circumvent this problem, we replace[5] $F(\tau_i)$ by $F'(\tau_i)$:

$$F'(\tau_i) = \exp\left\{\rho_0 L_c(\tau_i) - \rho'_1 \sum_{k=0}^{K-1} I_k(\tau_i)\right.$$
$$\left. - \max_{1 \leq l \leq R}\left(\rho_0 L_c(\tau_l) - \rho'_1 \sum_{k=0}^{K-1} I_k(\tau_l)\right)\right\} \tag{38}$$

where we multiply $F(\tau_i)$ by a real scalar factor.

C. Summary of Steps

In the following, we summarize the steps of the new algorithm for the two considered scenarios.

1) Based on the sampled data $y(iT)$, $i = 0, 1, \ldots, M - 1$, evaluate the periodogram $I_k(\tau)$ according to (24).
2) Compute the normalized importance function in (27). Note that, in practice, we use a discreet model by substituting the integration in the denominator of (27) by a summation as follows:

$$g'_{\rho'_1}(\tau) = \frac{\prod_{k=0}^{K-1} \exp\{\rho'_1 I_k(\tau)\}}{\sum_{n=1}^{N} \prod_{k=0}^{K-1} \exp\{\rho'_1 I_k(\tau_n)\}} \tag{39}$$

where N is the total number of points in the time delay interval.

3) Generate R realizations of the parameter $\{\tau_i\}_{i=1}^R$ using the inverse probability integration as detailed in Appendix D.
4) Evaluate the weight coefficient $F(\tau_i)$ defined in (37) (or $F'(\tau_i)$ defined in (38) if we consider that τ is in $[0, T]$) for each generated value τ_i.
5) Compute the mean of the generated variables multiplied by the weight coefficients to find the ML estimate of the time delay.

VII. SIMULATION RESULTS

In this section, we will present numerical results to substantiate the performance of the new ML estimator as a function of the SNR. We will also refer to our new IS-based ML estimator as "IS algorithm." The normalized (by T^2) mean-square error (NMSE), defined in (40), will be used as our performance measure:

$$\text{NMSE}(\tau) = \frac{E\left\{(\widehat{\tau^*} - \tau^*)^2\right\}}{T^2} \tag{40}$$

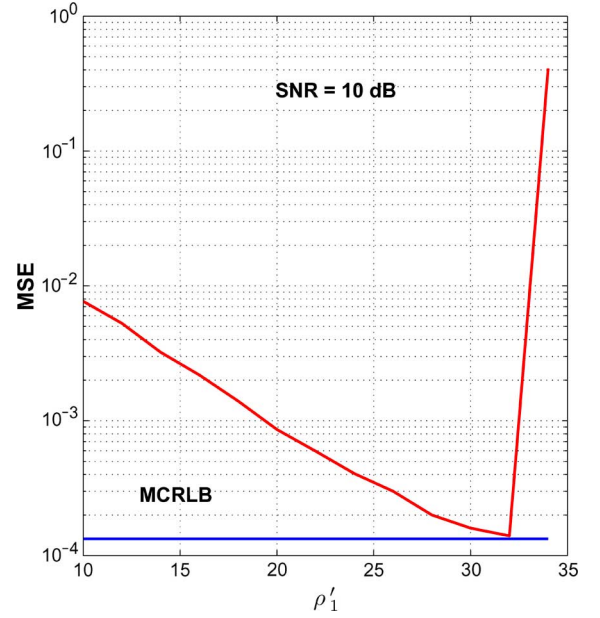[5]Note that the same simplifications have been used in [9] to estimate the signal DOA.

Fig. 3. Performance versus ρ'_1 for $\text{SNR} = 10$ B.

and computed over 1000 Monte Carlo runs. The modified Cramér–Rao lower bound (MCRLB) is also normalized by T^2 and the total number of transmitted symbols, K, in the observation window is set to $K = 100$ and ρ'_1 is taken equal to 28. Unless specified otherwise, a root-raised cosine shaping pulse of roll-off factor of 0.5 is used. First, the effect of ρ_1 (or equivalently ρ'_1) on the performance of our IS-based ML estimator is shown in Fig. 3 at an SNR of 10 dB. As it could be predicted, the mean-square error decreases as ρ_1 increases toward its optimal value and, for too large values, the performance deteriorates due to numerical overflows. In the implementation, ρ'_1 can be set as a function of the power of the received samples $\boldsymbol{y}$. Moreover, using computer simulations, we verify that for a root-raised cosine filter the ratio $\frac{L'_{c,\rho_0}(\tau)}{g'(\tau)}$ is almost equal to 1. Then to reduce the computational complexity, we can set this ratio to 1 in (21) and (36). In fact, Fig. 4 shows the NMSE of the IS-based time delay ML estimator when this ratio is preserved in the importance function and when it is set to 1. As it can be seen, this simplification does not degrade the performance of the estimator while reducing the computational complexity considerably. This simplification is also valid for any linear modulation scheme. Therefore, in the following simulations, we consider that

$$\frac{L'_{c,\rho_0}(\tau)}{g'_{\rho'_1}(\tau)} = 1. \tag{41}$$

Note that this simplification remains valid when the intersymbol interference is not important (for high values of the roll-off factor). As the roll-off factor tends to 0, it appears necessary to consider the ratio in order to achieve better performance of the estimator. Moreover, we implement the iterative CML estimator, called CML-TED, proposed in [8] and compare its performance to the performance of our IS-based CML estimator. As far as we know, among all the existing synchronization techniques, the CML-TED algorithm achieves the best
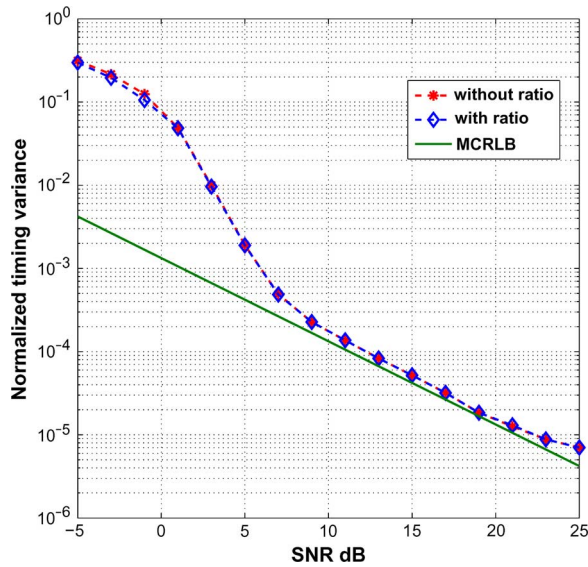
Fig. 4. Estimation performance considering $\frac{L'_{c,\rho_0}(\tau)}{g'(\tau)}$ and setting $\frac{L'_{c,\rho_0}(\tau)}{g'(\tau)}$ to 1 with QPSK modulation.
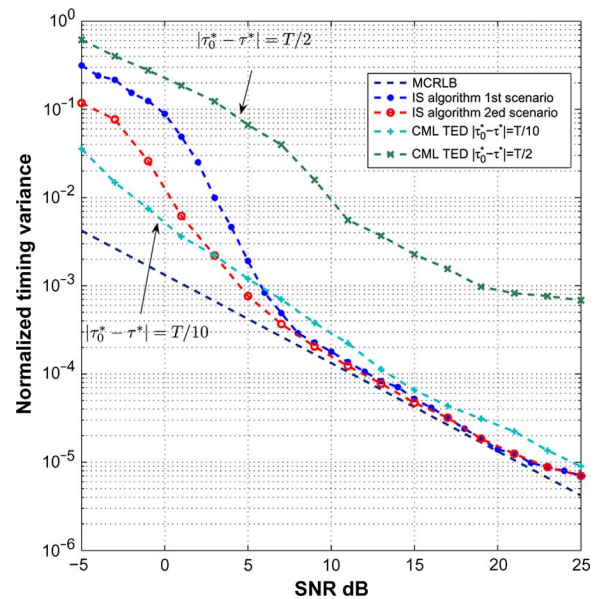

Fig. 5. Comparison between the estimation performance of the IS algorithm using the two scenarios and the tracking performance of the CML-TED using QPSK modulation.

performance, but, as an iterative procedure, its performance depends strictly on the initial guess. To corroborate our claims, we consider in Fig. 5 two initial values of τ^* for the CML-TED, which should be seen as the result of another estimator. In Fig. 5, the small crosses represent the normalized variance where the initial value is very close to the true time delay value, i.e., verifies $|\tau_0^* - \tau^*| = \frac{T}{10}$, with τ^* being the true time delay value to be estimated and τ_0^* is the initial guess. As it can be seen from this figure, even with a close-enough initial guess, our IS-based estimator outperforms the CML-TED estimator in the high SNR region although the CML-TED achieves better performance at low SNR values. We also note that, at high SNR values, the performance of our IS-based estimator is close to the MCRLB. This means that, in this region, our new time delay estimator exhibits performances equivalent to those that could be achieved if the transmitted data were perfectly known to the receiver. However, if we consider $|\tau_0^* - \tau^*| = \frac{T}{2}$, the performance of the CML-TED deteriorates considerably over the entire SNR region. This illustrates the fact that the CML-TED algorithm fails to estimate the time delay if the initial value is not appropriately chosen, while no initialization concerns are raised with our new IS-based CML estimator. Moreover, the second variant of the IS algorithm, namely considering the time delay as a circular variable, is also represented in Fig. 5. We see in this case that the variance error is reduced in the low SNR region. In addition, in both cases, starting from an SNR value of about 5 dB, our IS-based algorithm surpasses the iterative algorithm, even when assuming a sufficiently accurate initial guess.

Furthermore, in Fig. 6, the CML-TED algorithm exhibits a variance penalty for a roll-off equal to 0.2. This penalty is higher for smaller excess bandwidth. It has been shown in [8] that the CML-TED reaches the asymptotic compressed CRLB (CRLB_c), and the difference between the MCRLB and the CRLB_c becomes more important as the roll-off factor decreases. Then the performance of the CML-TED cannot approach asymptotically the MCRLB for small roll-off factors.
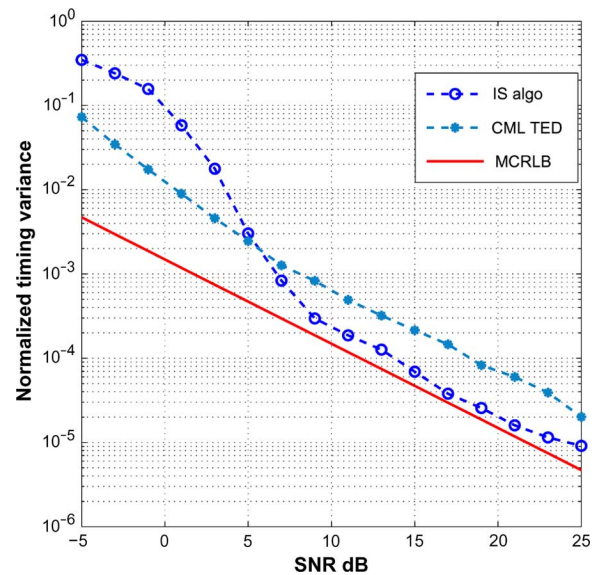

Fig. 6. Comparison between the estimation performance of IS algorithm and the tracking performance of CML TED using QPSK modulation and for a roll-off factor of 0.2.

In contrast, the new IS-based algorithm always reaches the MCRLB in the high SNR range, irrespectively of the roll-off factor value.

In Fig. 7, performance curves are drawn for 16-QAM and 64-QAM, as examples for non-constant-modulus constellations. As explained in Section VI, we force the first and the last transmitted symbols to be of maximum energy. To illustrate the performance degradation in the case of higher-order modulations, we also plot the NMSE for QPSK. As we can see, the IS algorithm achieves close performance for the three modulations orders, with, however, a small improvement for the QPSK modulation. To illustrate the performance enhancement achieved by forcing the first and the last symbols to
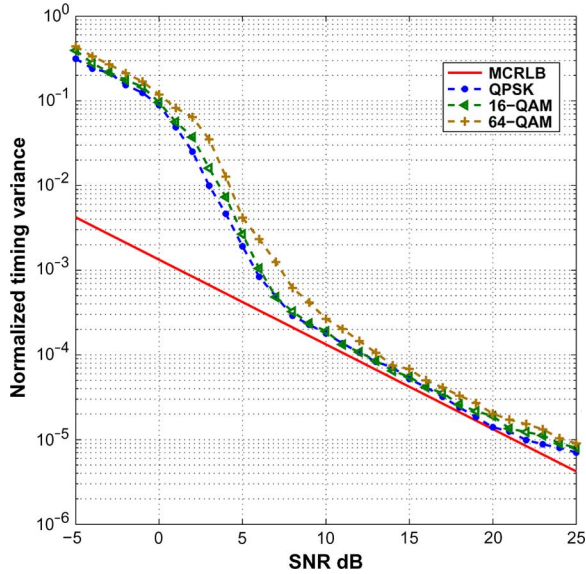
Fig. 7. Normalized MSE of the time delay estimate for different QAM modulation order, using a root-raised cosine filter with a roll-off factor 0.5.
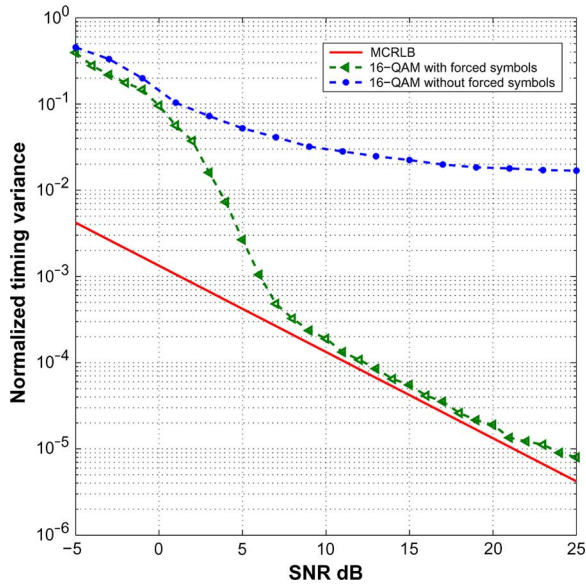

Fig. 8. Comparison of the estimation performance with and without forced symbols using 16-QAM modulation and for a roll-off factor of 0.5.

have maximum constellation magnitude, we plot, in Fig. 8 the performance of the new IS-based ML estimator without this constraint. As anticipated, the performance is strongly affected by the two edge symbols since the curve corresponding to the nonforced symbols does not approach the MCRLB. Therefore, for non-constant-envelope modulated constellations, it is essential to force the first and the last transmitted symbols to have maximum energy, as explained in Section VI.

VIII. CONCLUSION

A computationally efficient technique has been developed to implement the CML estimator of the time delay parameter. Based on a discrete-time model, the transmitted symbols are supposed to be unknown and no restriction on their distribution was assumed. To avoid iterative techniques and their drawbacks,

the importance sampling method was used to find the ML solution. Its main advantage over the iterative procedures is that it does not require any initial guess of the time delay parameter and that it is far less computationally expensive while retaining good performances. Moreover, its convergence to the global maximum is guaranteed. Relative to other proposed methods such as the CML-TED, the IS-based estimator exhibits better performance at high SNR. In practice, the choice of the algorithm parameters ρ_0 and ρ_1' is critical for the estimation performance and for a good choice of these parameters, a small number of generated realizations can be sufficient to achieve satisfactory performance and reduce the computation burden.

APPENDIX A
PROOF OF $\lim_{\rho_0 \to +\infty} L'_{c,\rho_0}(\tau) = \delta_\tau$

In the following, we prove that $L'_{c,\rho_0}(\tau)$ defined in (17) tends to a Dirac delta function centered at the location of its global maximum as $\rho_0 \to +\infty$. To do so, consider the general case where $f(x)$ is an integrable function having one global maximum, denoted a:

$$a = \arg \max_{x \in \mathbb{R}} f(x) \qquad (42)$$

and denoting by $F(x)$ the following normalized function:

$$F(x) = \frac{\exp \{\rho_0 f(x)\}}{\int_I \exp \{\rho_0 f(u)\} \, du} \qquad (43)$$

where I is the definition domain of $F(.)$. Then, for a given real number $b \neq a$, we have

$$F(b) = \frac{\exp \{\rho_0 f(b)\}}{\int_I \exp \{\rho_0 f(x)\} \, dx} < \frac{\exp \{\rho_0 f(a)\}}{\int_I \exp \{\rho_0 f(x)\} \, dx}. \qquad (44)$$

However, since $f(a)$ is the maximum value of the function $f(x)$, then $f(b) - f(a)$ is a negative number and, therefore, $\exp\{\rho_0(f(b) - f(a))\}$ tends to 0, as well as $F(b)$, when ρ_0 tends to ∞. As a result

$$\lim_{\rho_0 \to \infty} F(x) = 0, \qquad (45)$$

for any real $x \neq a$. Moreover, if we consider that $\lim_{\rho_0 \to \infty} F(a) = 0$, then whatever $x \in \mathbb{R}$, we have $\lim_{\rho_0 \to \infty} F(x) = 0$ and $\lim_{\rho_0 \to \infty} \int_{-\infty}^{+\infty} F(u) du = 0$, which is in conflict with the assumption that $\int_{-\infty}^{+\infty} F(u) du = 1$.

Finally, we conclude that $\lim_{\rho_0 \to \infty} F(a) \neq 0$ and since $\lim_{\rho_0 \to \infty} \int_{-\infty}^{+\infty} F(u) du = 1$, $F(x)$ becomes a Dirac delta function centered at a when ρ_0 tends to $+\infty$.

APPENDIX B
JUSTIFICATION OF THE APPROXIMATION $\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau \approx \frac{E_h}{T_s} \boldsymbol{I}_K$

The diagonal elements of $\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau$ are the convolution of the same shifted version of $h(.)$ ($[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{i;i} = \boldsymbol{a}_i^T(\tau)\boldsymbol{a}_i(\tau)$ for $i = 0, 1, \ldots K - 1$ where $\boldsymbol{a}_i(\tau)$ is defined in (6)). Whereas, when the shift is not the same (i.e., $[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{i;j} = \boldsymbol{a}_i^T(\tau)\boldsymbol{a}_j(\tau)$, $i \neq j$), the value of the convolution, $\boldsymbol{a}_i^T(\tau)\boldsymbol{a}_j(\tau)$, is, according to the Nyquist criteria, equal to zero. However, since we take some samples of $h(.)$, $[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{i;j} = \boldsymbol{a}_i^T(\tau)\boldsymbol{a}_j(\tau)$, $i \neq j$ is not really equal to zero but still very small and negligible in front of
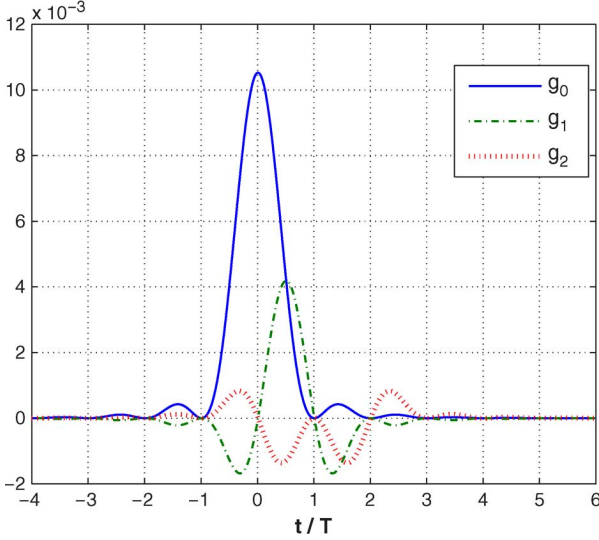
Fig. 9. Plot of $g_0(.), g_1(.)$ and $g_2(.)$.

$[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{i;i}$. To clarify, we plotted in Fig. 9 the following three functions: $g_0(x) = h(x) \times h(x) = h(x)^2, g_1(x) = h(x) \times h(x-T)$ and $g_2(x) = h(x) \times h(x-2T)$ and we take for example $[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{1;2} = \boldsymbol{a}_1^T(\tau)\boldsymbol{a}_2(\tau) = \sum_{m=0}^{M-1} g_1(mT_s)$. Then, notice from this figure that some samples of $g_1(.)$ are negative, which will compensate for positive samples in $\sum_{m=0}^{M-1} g_1(mT_s)$ from the same function thereby resulting in very low convolution. However, the samples of $g_0(x)$ are always positive and there is no compensation effect in $[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{1;1} = \sum_{m=0}^{M-1} g_0(mT_s)$ added to the fact that $g_0(x) \gg g_1(x)$, this results in the following conclusion: $[\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{1;1} \gg [\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau]_{1;2}$. Using the same arguments for the other off-diagonal elements of $\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau$ leads to the following approximation:

$$\boldsymbol{A}_\tau^T \boldsymbol{A}_\tau \approx \frac{E_h}{T_s} \boldsymbol{I}_K. \tag{46}$$

APPENDIX C
PROOF OF THE EFFECT OF ρ_1' ON $g_{\rho_1'}(.)$

In the following, we briefly show how ρ_1' (or equivalently ρ_1) can render $g_{\rho_1'}(.)$ more peaked around its global maximum. Starting from $g_{\rho_1'}(\tau^*)$, define the function $H_{\tau^*}(\rho_1') = g_{\rho_1'}(\tau^*)$:

$$H(\rho_1') = \frac{\exp\left\{\rho_1' L_c^{(a)}(\tau^*)\right\}}{\int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv} \tag{47}$$

where τ^* is the true time delay value to be estimated and $L_c^{(a)}(\tau)$ is the approximation of $L_c(\boldsymbol{y}; \tau)$ defined in the right-hand side of (22), i.e., $L_c^{(a)}(\tau) = \boldsymbol{y}^H \boldsymbol{A}_\tau \boldsymbol{A}_\tau^T \boldsymbol{y}$. The first derivative of $H(\rho_1')$ with respect to ρ_1' is given by

$$H'(\rho_1') = \frac{1}{\left(\int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv\right)^2} \times \left[L_c^{(a)}(\tau^*) \exp\left\{\rho_1' L_c^{(a)}(\tau^*)\right\}\right.$$

$$\times \int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv - \exp\left\{\rho_1' L_c^{(a)}(\tau^*)\right\}$$

$$\left.\times \int_J L_c^{(a)}(v) \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv\right]$$

$$= \frac{1}{\left(\int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv\right)^2}$$

$$\times \left[\exp\left\{\rho_1' L_c^{(a)}(\tau^*)\right\}\right.$$

$$\times \left(L_c^{(a)}(\tau^*) \int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv\right.$$

$$\left.\left.- \int_J L_c^{(a)}(v) \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv\right)\right]. \tag{48}$$

And noting that $\tau^* = \arg\max_\tau L_c^{(a)}(\tau)$, it follows that

$$\int_J L_c^{(a)}(v) \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv$$

$$< L_c^{(a)}(\tau^*) \int_J \exp\left\{\rho_1' L_c^{(a)}(v)\right\} dv. \tag{49}$$

Therefore, $H'_{\tau^*}(\rho_1') > 0 \forall \rho_1'$. Hence, $H_{\tau^*}(\rho_1')$ is an increasing function with respect to ρ_1', i.e., for every $\rho_1'^{(2)} > \rho_1'^{(1)}$ we have $H_{\tau^*}(\rho_1'^{(2)}) > H_{\tau^*}(\rho_1'^{(1)})$, which means $g'_{\rho_1'^{(2)}}(\tau^*) > g'_{\rho_1'^{(1)}}(\tau^*)$. We conclude that ρ_1' renders the objective function more and more peaked around its global maximum.

APPENDIX D
METHOD TO GENERATE $[\tau_i]_{i=1}^R$

In this appendix, we detail how to generate realizations according to $g'(\tau)$. First, we generate a vector $\boldsymbol{u} = [u_1, u_2, \ldots, u_R]$ of R realizations uniformly distributed in [0,1]. Then, we search $\tau_i = G'^{-1}(u_i)$, where $G'^{-1}(.)$ is the reciprocal function of the cumulative distribution function (CDF) $G'(x)$ of x defined as

$$G'(x) = \int_0^x g'_{\rho_1'}(v) dv. \tag{50}$$

Unfortunately, a closed-form expression of $G'^{-1}(x)$ is not analytically tractable. Moreover, since $G'(x)$ is a steep-slope function, a fine search to find τ_i as $\arg\min_\tau |u_i - G'(\tau)|$ is required and makes the process computationally intensive. However, since $G'(\tau)$ is an increasing function of τ, the function $S(\tau) = |u_i - G'(\tau)|$ is unimodal. This observation allows us to adopt the golden search [17] to find the location of the minimum of $S(\tau)$. The golden search is appropriate for this problem because it converges after a small number of iterations and requires only one function evaluation per iteration.

REFERENCES

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

[2] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed.   New York: Mc-Graw-Hill, 2001.

[3] G. C. Carter, *Coherence and Time Delay Estimation: An Applied Tutorial for Research, Development, Test and Evaluation Engineers*.   New York: IEEE Press, 1993.

[4] J. W. M. Bergmans, *Digital Baseband Transmission and Recording*. Boston, MA: Kluwer, 1996.

[5] M. Olsson, H. Johansson, and P. Lowenborg, "Time-delay estimation using farrow-based fractional-delay FIR filters: Filter approximation versus estimation errors," in *Proc. XIV Eur. Signal Process. Conf. (EU-SIPCO) 2006*, Florence, Italy, Sep. 2006.

[6] N. Noels *et al.*, "Turbo symchronization: An EM algorithm interpretation," in *IEEE Conf. Commum.*, Anchorage, AK, May 2003, pp. 2933–2937.

[7] S. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*.   Englewood Cliffs, NJ: Prentice-Hall, 1993.

[8] J. Riba, J. Sala, and G. Vazquez, "Conditional maximum likelihood timing recovery: Estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835–850, Apr. 2001.

[9] S. Saha and S. Kay, "An exact maximum likelihood narrowband direction of arrival estimator," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 1082–1092, Oct. 2008.

[10] S. Saha and S. Kay, "Mean likelihood frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1937–1946, Jul. 2000.

[11] H. Wang and S. Kay, "A maximum likelihood angle-doppler estimator using importance sampling," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 2, pp. 610–622, Apr. 2010.

[12] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 1783–1795, Oct. 1990.

[13] M. Pincus, "A closed form solution for certain programming problems," *Oper. Res.*, pp. 690–694, 1962.

[14] L. Stewart, "Bayesian analysis using Monte Carlo integration: A powerful methodology for handling some difficult problems," *Amer. Stat.*, pp. 195–200, 1983.

[15] S. Kay, "Comments on frequency estimation by linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 198–199, Apr. 1979.

[16] K. V. Mardia, *Statistics of Directional Data*.   New York: Academic, 1972.

[17] D. J. Wilde, *Globally Optimum Design*.   Englewood Cliffs, NJ: Prentice-Hall, 1964.

Mr. Bellili was selected by the INRS as its candidate for the 2009–2010 competition of the very prestigious Vanier Canada Graduate Scholarships program. He also received the Academic Gold Medal of the Governor General of Canada for the year 2009–2010 and the Excellence Grant of the Director General of INRS for the year 2009–2010. He also received the award of the best M.Sc. thesis of INRS-EMT for the year 2009–2010 and twice—for both the M.Sc. and Ph.D. programs—the National Grant of Excellence from the Tunisian Government. He was also rewarded in 2011 the Merit Scholarship for Foreign Students from the Ministère de l'Education, du Loisir et du Sport (MELS), Québec, Canada. He serves regularly as a reviewer for many international scientific journals and conferences.
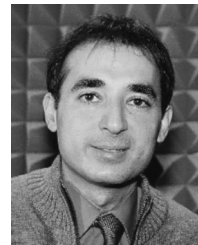
Sofiène Affes (S'94–M'95–SM'04) received the Diplôme d'Ingénieur degree in electrical engineering in 1992, and the Ph.D. degree (with hons.) in signal processing in 1995, both from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France. He has since been with INRS-EMT, University of Quebec, Montreal, QC, Canada, as a Research Associate from 1995 until 1997, then as an Assistant Professor until 2000. Currently, he is an Associate Professor in the Wireless Communications Group. His research interests are in wireless communications, statistical signal and array processing, adaptive space-time processing, and MIMO. From 1998 to 2002, he has been leading the radio design and signal processing activities of the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications at INRS-EMT, Montreal, QC, Canada. Since 2004, he has been actively involved in major projects in wireless communication of PROMPT (Partnerships for Research on Microelectronics, Photonics and Telecommunications). Prof. Affes was the corecipient of the 2002 Prize for Research Excellence of INRS. He currently holds a Canada Research Chair in Wireless Communications and a Discovery Accelerator Supplement Award from NSERC (Natural Sciences & Engineering Research Council of Canada). In 2006, he served as a General Co-Chair of the IEEE Vehicular Technology Conference (VTC) 2006-Fall, Montreal, QC, Canada. In 2008, he received from the IEEE Vehicular Technology Society the IEEE VTC Chair Recognition Award for exemplary contributions to the success of IEEE VTC. He currently serves as a member of the Editorial Board of the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, and the Wiley Journal on Wireless Communications and Mobile Computing.

Ahmed Masmoudi was born in Ariana, Tunisia, on February 10, 1987. He received the Diplôme d'Ingénieur degree in telecommunication from the Ecole Supérieure des Communications de Tunis-Sup'Com (Higher School of Communication of Tunis), Tunisia, in 2010. Since September 2010, he has been working toward the M.Sc. degree in the Institut National de la Recherche Scientifique (INRS), Montréal, QC, Canada.

His research activities include signal processing and parameters estimation for wireless communication.

Mr. Masmoudi is the recipient of the National Grant of Excellence from the Tunisian Government.

Faouzi Bellili was born in Sbeitla, Kasserine, Tunisia, on June 16, 1983. He received the Diplôme d'Ingénieur degree (with Hons.) in signals and systems from the Tunisia Polytechnic School in 2007 and the M.Sc. degree (with exceptional grade) at the Institut National de la Recherche Scientifique-Energie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, Montréal, QC, Canada, in 2009. He is currently working towards the Ph.D. degree at the INRS-EMT.

His research focuses on statistical signal processing and array processing with an emphasis on parameters estimation for wireless communications. During his M.Sc. studies, he has authored/coauthored six international journal papers and more than ten international conference papers.

Alex Stéphenne (S'94–M'95–SM'04) was born in Quebec, Canada, on May 8, 1969. He received the B. Eng. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1992 and the M.Sc. degree and Ph.D. degree in telecommunications from INRS-Télécommunications, Université du Québec, Montreal, QC, Canada, in 1994 and 2000, respectively.

In 1999, he joined SITA, Inc., Montreal, QC, Canada, where he worked on the design of remote management strategies for the computer systems of airline companies. In 2000, he became a DSP Design Specialist for Dataradio, Inc., Montreal, QC, Canada, a company specializing in the design and manufacturing of advanced wireless data products and systems for mission critical applications. In January 2001, he joined Ericsson and worked for over two years in Sweden, where he was responsible for the design of baseband algorithms for WCDMA commercial base station receivers. From June 2003 to December 2008, he was still working for Ericsson, but was based in Montreal, where he was a researcher focusing on issues related to the physical layer of wireless communication systems. Since 2004, he has also been an Adjunct Professor at INRS, where he has been continuously supervising the research activities of multiple students. His current research interests include Coordinated Multi-Point (CoMP) transmission and reception, Inter-Cell Interference Coordination (ICIC) and mitigation techniques in Heterogeneous Networks (HetNets), wireless channel modeling/characterization/estimation, statistical signal processing, array processing, and adaptive filtering for wireless telecom.

Dr. Stéphenne is a member of the organizing committee and a Co-Chair of the Technical Program Committee (TPC) for the 2012-Fall IEEE Vehicular Technology Conference (VTC'12-Fall) in Quebec City, Canada. He has served as a Co-Chair for the Multiple Antenna Systems and Space-Time Processing track for VTC'08-Fall in Calgary, Canada, and as a Co-Chair of the TPC for VTC'06-Fall in Montreal, Canada.