Non Data Aided Estimation of the Modulation Index of Continuous Phase Modulations

Pascal Bianchi, Philippe Loubaton, François Sirven

Abstract

In this paper, a new non data aided estimator of the modulation index of continuous phase modulated (CPM) signals is proposed. It is based on the observation that the inverse of the index is the smallest positive real number a CPM signal should be raised to in order to generate a sinusoid of period $2T$, where T is the symbol period. The asymptotic behavior of the estimator is studied. If N is the sample size, the estimation error is shown to converge to a non Gaussian distribution at rate $1/N$. Simulation results sustain the conclusions of the theoretical asymptotic analysis.

Keywords

Continuous phase modulation, modulation index, non data aided estimation.

Edics

3-PERF, 3-SYNC

P. Bianchi and Ph. Loubaton are with the Université de Marne-la-Vallée and wtih the Laboratoire Traitement et Communication de l'Information, UMR 5141 CNRS/ENST, Université de Marne-la-Vallée, 5 boulevard Descartes, 77454 Marne-la-Vallée, France (e-mail: {bianchi, loubaton}@univ-mlv.fr).

F. Sirven is with Thalès Communications, 160, Boulevard de Valmy 92700 Colombes, France (e-mail: francois.sirven@fr.thalesgroup.com).

I. Introduction

Blind estimation of technical parameters characterizing the modulation used by a partially unknown transmitter is useful in certain civil or military applications. Spectrum monitoring is a typical civil application for which the values of the parameters of the active transmitters should be controlled. Passive listening is the most obvious military application which requires blind estimation procedures. Indeed, certain modulation parameters should be estimated in order to be able to extract the transmitted information and to recognize the type of transmitter.

Continuous phase modulation (CPM) is a widespread scheme thanks to its attractive spectral efficiency and its constant modulus property. In particular, it is used in the European second generation mobile system GSM, in the professional mobile communications system Tetrapol, as well as in a number of military systems. Non data aided estimation of CPM technical parameters is thus a problem of current interest.

A number of works have been devoted to the non data aided estimation of timing, phase and frequency offset of CPM signals (see e.g. [8], [9]). However, blind estimation of the modulation index of a CPM signal is comparatively less popular. This problem does not seem to have been investigated thoroughly in the literature except in [5] and [6] : in these works, estimators based on higher order statistics of the received signal are presented. However, these estimation methods work only in case of *full response* CPM signals, i.e. when the support of the shaping filter used at the transmitter side coincides with $[0, T]$, where T is the symbol period. We also mention that the use of such methods requires the prior knowledge of the symbol period, the time delay of arrival and the frequency offset.

In this paper, we propose a novel non data aided approach to estimate the modulation index and we analyze the performance of this estimator. As this contribution is motivated by applications to passive listening, the estimation procedure is expected to perform without any *a priori* information on the class of CPM signal used by the transmitter. Therefore, we assume that both the data sequence and the shaping filter are unknown at the receiver side. In this paper, the symbol period and the carrier frequency are supposed to be available for the sake of simplicity. However, the approach that is going to be introduced in the sequel can be adapted in order to estimate jointly the index, the symbol rate and if needed, the carrier frequency of the CPM signal. As the performance analysis of the joint estimate is far from being a direct generalization of the results presented in the present work, this joint estimation problem is not addressed in details here due to the lack of space. However, subsection IV-C briefly indicates how the joint estimation of the unknown parameters can be achieved in this case.

The present paper is organized as follows. Section II describes the signal model and the basic notations. In order to motivate the use of our estimate, we first review the existing approaches in section III. We discuss in subsection III-A on the maximum likelihood estimation of the modulation index. As the shaping filter of the modulation is unknown, the joint estimation of the modulation index and of a sequence of coefficients related to the shaping filter is required. The expectation maximization (EM) algorithm ([12], [13]) can be used to this end. However, each iteration of the algorithm leads to the maximization of a multi-dimensional non convex function. Therefore, the EM algorithm must be properly initialized. Moreover, as we shall see below, the computational cost of each iteration of the EM algorithm is proportional to N^2, where N is the number of symbols of the sample set. Of course, this makes this algorithm difficult to implement in practice. It is thus quite relevant to propose sub-optimum estimators. We therefore present briefly in subsection III-B the estimator proposed in [5], which makes sense only in the full response case.

Section IV introduces the proposed estimator. It is based on the observation that the inverse of the index is the smallest positive real number a CPM signal should be raised to in order to generate a sinusoid of period $2T$, where T is the symbol period. This property is a direct consequence of the exhaustive study [4] of the cyclic properties of CPM signals. We also briefly indicate how this property can be used in order to estimate jointly the modulation index, the carrier frequency and the symbol period.

In section V, we study the asymptotic performance of the proposed index estimate. The mean square error of the estimation is shown to be asymptotically proportional to $1/N^2$, where N is the number of symbols of the sample set. In particular, this claim implies that the rate of convergence of the estimator is higher than in many standard problems. Furthermore, the asymptotic behavior of the proposed estimate is rather unconventional: the estimation error converges in distribution toward a non Gaussian random variable constructed from certain Brownian motions. It is worth noting that the practical implementation of the estimate requires to raise the received signal to non integer powers. Thus, the phase of the received signal must be unwrapped by a procedure which can produce some errors in teh noisy case. These errors are unfortunately difficult to take into account in the asymptotic analysis of the performance. While the asymptotic results remains exact in the noiseless case, the study of the noisy case requires to neglect the influence of the phase unwrapping errors.

Section VI compares the theoretical asymptotic distributions with the empirical ones. Two different classes of CPM signals (LREC and LRC) are considered in the simulations. We study the values of the signal to noise ratio (SNR) for which our theoretical results allow to predict the behavior of the estimate. We finally compare the performance of the proposed estimator to the performance of the estimator of [5] based on higher-order statistics in case of full response CPM signals.

II. Signal model

The complex envelope $s_a(t)$ of a CPM signal can be expressed as follows:

$$s_a(t) = \exp i\psi_a(t) \, , \tag{1}$$

where the phase $\psi_a(t)$ is given by:

$$\begin{aligned} \psi_a(t) &= \pi h \int_{-\infty}^{t} \sum_{j \in \mathbb{Z}} a_j \, g_a(u - jT) du \\ &= \pi h \sum_{j \in \mathbb{Z}} a_j \, \phi_a(t - jT) \, . \end{aligned} \tag{2}$$

$(a_n)_{n \in \mathbb{Z}}$ represents the symbol sequence. It is assumed that for each n, a_n is equally likely ± 1 and that the sequence is independent identically distributed (i.i.d.). T is the symbol period, and function $g_a(t)$, classically called the shaping filter, is positive and non zero on the interval $[0, LT]$, and is zero outside $[0, LT]$, where L is a positive integer. $g_a(t)$ is normalized in such a way that $\int_0^{LT} g_a(t) dt = 1$. Therefore, function $\phi_a(t)$ defined by $\phi_a(t) = 0$ if $t < 0$ and $\phi_a(t) = \int_0^t g_a(s) ds$ satisfies $\phi_a(t) = 1$ for $t \geq LT$. The key parameter h, called the **modulation index**, is therefore characterized by the fact that the phase variation induced by a symbol is equal to $\pm\pi h$. It is quite useful to mention that if

$nT \leq t \leq (n+1)T$, $s_a(t)$ can be written as

$$s_a(t) = \exp\left(i\pi h\left[\sum_{k=-\infty}^{n-L} a_k + \sum_{k=0}^{L-1} \phi_a(t-kT)a_{n-k}\right]\right) . \tag{3}$$

Expression (3) is extensively used in the sequel.

As mentioned previously, our work is motivated by the passive listening context. Therefore, both the data sequence $(a_n)_{n\in\mathbb{N}}$ and the shaping filter are supposed to be unknown at the receiver side. However, for the sake of simplicity, we assume in the following that the symbol period T and the carrier frequency are known. This assumption is motivated by the observation that these parameters can be estimated prior to the modulation index h. In any case, we briefly show in section IV that our approach also allows to estimate jointly the index, the carrier frequency and the symbol period.

We recall that the carrier frequency is supposed to be known. Therefore, it can be assumed that it has been correctly compensated at the receiver side. Consequently, the complex envelope $y_a(t)$ of the received signal is a scaled and delayed version of the transmitted CPM signal corrupted by a complex additive Gaussian noise $b_a(t)$ of variance σ^2. The noise is assumed to be white in the bandwidth of the reception filter. We denote by N_0 its constant spectral density. The received signal can be written as follows:

$$y_a(t) = \alpha s_a(t-\tau) + b_a(t) . \tag{4}$$

Parameter τ represents the time delay between the transmitter and the receiver and is supposed to be unknown. Factor α is an unknown complex gain. Note that we implicitly assume that the reception filter does not produce any distortion on the modulated signal.

The received signal is observed during N signaling intervals (i.e. for $0 \leq t < NT$) and is sampled at rate T/M, where $M \in \mathbb{N}$. For convenience, we assume without restriction that $0 \leq T - \tau < \frac{T}{M}$. For each $k = 0, 1, 2, \ldots, NM-1$, we denote by $y(k)$, $s(k)$, $b(k)$ the discrete-time signals defined by $y(k) = y_a(T + kT/M)$, $s(k) = s_a(kT/M + T - \tau)$ and $b(k) = b_a(T + kT/M)$ respectively. The discrete-time version of (4) can be written as follows:

$$y(k) = \alpha s(k) + b(k) . \tag{5}$$

We also define sequence $(\psi(k))_{k\in\mathbb{Z}}$ by $\psi(k) = \psi_a(kT/M + T - \tau)$. As $nT \leq nT + mT/M + T - \tau \leq (n+1)T$ for $n \geq 0$ and $0 \leq m \leq M-1$, we obtain immediately from (3) that for each $n \geq 0$ and for each m such as $0 \leq m \leq M-1$,

$$\psi(nM+m) = \pi h \left(\sum_{j=-\infty}^{n-L} a_j + \sum_{j=0}^{L-1} a_{n-j}\phi_{j,m} \right) , \tag{6}$$

where $\phi_{j,m}$ is defined by $\phi_{j,m} = \phi_a(jT + mT/M + T - \tau)$. The assumption $0 \leq T - \tau < \frac{T}{M}$ is motivated by formula (6). If $\frac{k_0 T}{M} \leq T - \tau < \frac{(k_0+1)T}{M}$ for $1 \leq k_0 \leq (M-1)$, $\psi(nM+m)$ has to be replaced in (6) by $\psi(nM+m-k_0)$. This would complicate the notations of section 5 without modifying the results.

III. Review of the existing approaches.

A. The Maximum Likelihood Estimate.

The maximization of the likelihood function $p(y(0)\ldots y(NM-1)/h)$ w.r.t. h seems to be the most natural approach to estimate h. However, the use of the maximum likelihood criterion in our particular context gives rise to practical difficulties.

It is worth keeping in mind that the aim of the present section is not so much to investigate the maximum likelihood estimation of h in details than to point the problems that it poses. Consequently, we can make the following assumptions in order to simplify the notations. We assume in this section that the received signal is sampled at the symbol rate T, i.e. that $M = 1$, and that $b(k) = b_a(kT)$ is an i.i.d. sequence. Using (5) and (6), the received samples can be written as follows for each $n \geq 0$:

$$y(n) = |\alpha| \ \exp i \left(\pi h \sum_{k=-L+1}^{n-L} a_k + \pi h \sum_{k=0}^{L-1} \phi_k a_{n-k} + \Phi \right) + b(n), \tag{7}$$

where $\phi_k = \phi_a(kT + T - \tau)$ and where Φ represents the unknown phase $\Phi = \pi h \sum_{j=-\infty}^{-L} a_j +$ $\text{Arg}(\alpha)$. In order to simplify what follows, the noise variance σ^2 is supposed to be known. We assume without further restriction that factor $|\alpha|$ is equal to one. As parameter Φ and coefficients $(\phi_k)_{k=0,\ldots,L-1}$ are unknown, the maximum likelihood estimation of h is equivalent to the joint maximum likelihood estimation of h, Φ and coefficients $(\phi_k)_{k=0,\ldots,L-1}$. We thus consider $\boldsymbol{\theta} = (h, \Phi, \phi_0, \ldots, \phi_{L-1})^T$ and denote by $\mathbf{y}$ the set of received samples $\mathbf{y} = (y(0), \ldots, y(N-1))^T$ which is often called the *incomplete data*. Function $p(\mathbf{y}/\boldsymbol{\theta})$ represents the likelihood.

Since no information on the transmitted symbols is available, the computation of each value taken by the function $\boldsymbol{\theta} \to p(\mathbf{y}/\boldsymbol{\theta})$ requires about 2^N operations. The direct maximization of $\boldsymbol{\theta} \to p(\mathbf{y}/\boldsymbol{\theta})$ is thus impractical. The so-called Expectation Maximization (EM) algorithm ([13]) seems therefore to be the best way to maximize $p(\mathbf{y}/\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$. For a given $\boldsymbol{\theta}$, it can be seen on (7) that for each n, $y(n)$ depends on random vector $\boldsymbol{z}(n) = (z_1(n), a_{n-L+1}, \ldots, a_n)^T$, where $z_1(n) = \sum_{k=-L+1}^{n-L} a_k$. The *hidden data* refers to the sequence $\mathbf{Z} = (\boldsymbol{z}(0), \ldots, \boldsymbol{z}(N-1))^T$. The EM algorithm generates a sequence $(\boldsymbol{\theta}^{(k)})_{k \geq 1}$ which converges to a local maximum of the function $\boldsymbol{\theta} \to p(\mathbf{y}/\boldsymbol{\theta})$. At the step $(k+1)$, the estimate $\boldsymbol{\theta}^{(k+1)}$ is defined as $\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, where

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = E\left(\log p(\mathbf{y}, \mathbf{Z}/\boldsymbol{\theta}) \;/\mathbf{y}, \boldsymbol{\theta}^{(k)}\right). \tag{8}$$

At each iteration, it is thus necessary to evaluate function $\boldsymbol{\theta} \to Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ (so-called Expectation step) and to maximize this function w.r.t. $\boldsymbol{\theta}$ (Maximization step). In order to analyze the complexity of this algorithm, a more convenient expression of (8) can be provided by using the following remarks.

- For a given $\boldsymbol{\theta} = (h, \Phi, \phi_0, \ldots, \phi_{L-1})^T$, (7) can be written as $y(n) = F_{\boldsymbol{\theta}}(\boldsymbol{z}(n)) + b(n)$, where $F_{\boldsymbol{\theta}}(\boldsymbol{z}(n))$ is the mapping defined by

$$F_{\boldsymbol{\theta}}(\boldsymbol{z}(n)) = \exp\, i\left(\pi h z_1(n) + \pi h \sum_{k=0}^{L-1} \phi_k a_{n-k} + \Phi\right);$$

- For each $n = 0 \ldots N-1$, the first component $z_1(n)$ of $\boldsymbol{z}(n)$ is such that $-n \leq z_1(n) \leq n$. Moreover, $z_1(n)$ is even (resp. odd) if and only if n is even (resp. odd). As each of the L other components $(a_{n-L+1}, \ldots, a_n)$ of $\boldsymbol{z}(n)$ coincides with ± 1, vector $\boldsymbol{z}(n)$ belongs to a set $\mathcal{Z}_n$ which contains $2^L(n+1)$ values.

After some standard algebra, we obtain that $\boldsymbol{\theta}^{(k+1)}$ can as well be defined as the argument of the minimum of the following function $R(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$:

$$R(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \sum_{n=0}^{N-1} \sum_{\boldsymbol{z} \in \mathcal{Z}_n} |y(n) - F_{\boldsymbol{\theta}}(\boldsymbol{z})|^2\, p(\boldsymbol{z}(n) = \boldsymbol{z}/\mathbf{y}, \boldsymbol{\theta}^{(k)}) \tag{9}$$

where $\boldsymbol{z} \to p(\boldsymbol{z}(n) = \boldsymbol{z}/\mathbf{y}, \boldsymbol{\theta}^{(k)})$ denotes the conditional probability distribution of $\boldsymbol{z}(n)$ given $\mathbf{y}$ and the value of the parameter $\boldsymbol{\theta}^{(k)}$ obtained at step k. The Expectation step of the EM algorithm requires to evaluate the conditional distribution of $z(n)$ given $\mathbf{y}$ and

$\boldsymbol{\theta}^{(k)}$ for each $n = 0 \ldots N-1$. To that end, the so-called forward-backward algorithm can be employed ([13]). As $\mathcal{Z}_n$ contains $2^L(n+1)$ values for each n, it can be shown that its computation requires about $9\frac{N(N+1)}{2}2^L$ operations. As for the Maximization step, we mention that function $\boldsymbol{\theta} \rightarrow R(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ cannot be maximized in closed form. It is thus necessary to use an iterative maximization method, which costs $O(N^2)$ operations at each iteration. For example, the implementation of a gradient algorithm requires to evaluate the gradient of $\boldsymbol{\theta} \rightarrow R(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ at each iteration: each component of the gradient has a form similar to (9), and its evaluation in each point thus requires about $2^L N(N+1)/2$ operations. Furthermore, we stress that $\boldsymbol{\theta} \rightarrow R(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ is not a convex function of $\boldsymbol{\theta}$. Therefore, quite complicated algorithms have to be used in order to perform the maximization step if no relevant initial estimate of $\boldsymbol{\theta}$ is available.

The above remarks imply that the maximum likelihood estimate of h does not seem to be appropriate for a practical implementation even if the EM algorithm is used. It is thus necessary to propose sub-optimum estimates.

B. The estimator based on Higher-Order Statistics.

We now briefly present the estimator proposed in [5]. As mentioned in the introduction, this approach can be used only in case of **full response** CPM signals. Therefore, we assume that $L = 1$ in this subsection. Although the approach of [5] can be adapted to the case where $M > 1$ (see e.g. [6]), we moreover assume that $M = 1$, i.e. that the received signal is sampled at the symbol rate T. Keeping the notations of subsection III-A, the received signal $y(n)$ can be written as

$$y(n) = |\alpha| \, \exp i \left(\pi h \sum_{k=-1}^{n-1} a_k + \pi h \phi_0 a_n + \Phi \right) + b(n), \tag{10}$$

where $\phi_0 = \phi_a(T - \tau)$. It is worth noting that [5] assumes that the receiver is **perfectly synchronized**. In this case, the delay τ is equal $\tau = T$, and parameter ϕ_0 coincides with $\phi_a(T) = 1$. [5] is based on the simple observation that

$$\frac{\operatorname{cum}(y(n+1), y(n)^*, y(n), y(n)^*)}{\operatorname{cum}(y(n), y(n)^*, y(n), y(n)^*)} \tag{11}$$

coincides with $\cos \pi h$. Parameter h can therefore be estimated by $\frac{1}{\pi}\arccos(\hat{d})$ where $\hat{d}$ represents an estimate of (11). The asymptotic performance of this estimate was studied in [5]. It was shown that the estimate is asymptotically Gaussian, that it converges at rate $\frac{1}{\sqrt{N}}$, and that its variance converge to 0 if the noise variance converge to 0.

Although quite simple to implement, this estimate is irrelevant in the case $L > 1$. However, parameter L is greater than 1 in a number of communication systems in order to make the CPM modulation spectrally efficient. It is thus quite relevant to look for alternative blind estimates of h.

IV. The proposed Estimate

A. A cyclic Property of CPM Signals

The proposed estimate is based on a cyclic property of CPM signals which has been recently derived in [4].

Proposition 1: Let $r_a(t)$ be a CPM signal of modulation index f. Then,

- If f is non integer, then $E(r_a(t)) = 0$ for each t.
- If f is a non zero even integer, the function $t \to E(r_a(t))$ is periodic of period T.
- If f is an odd integer, the function $t \to E(r_a(t))$ is periodic of period $2T$.

We now consider a CPM signal $s_a(t)$ given by (1) and we denote by h its modulation index. In the following, we need to define the signal $s_a(t)^g$ for any positive real number g. This requires some care because when setting

$$s_a(t)^g = |s_a(t)|^g \exp(ig\text{Arg}(s_a(t)))$$

one has to specify which particular determination of $\text{Arg}(s_a(t))$ is chosen. Here, we choose the determination

$$\text{Arg}(s_a(t)) = \psi_a(t) ,$$

which can also be interpreted as the determination for which the function $t \to \text{Arg}(s_a(t))$ is continuous. Using the above definition, it is clear that $s_a(t)^g$ can be interpreted as a CPM signal of index gh.

The following result is an immediate corollary of Proposition 1.

Proposition 2: A positive real number g satisfies

$$\lim_{S \to \infty} \frac{1}{S} \int_0^S s_a(t)^g e^{-2i\pi \frac{t}{2T}} dt \neq 0 \tag{12}$$

only if g is an odd integer multiple of $g_0 = \frac{1}{h}$.

Sketch of the proof. We first observe that in order to show (12), it is sufficient to restrict to the case where $S = 2NT$ where $N \in \mathbb{N}$ is supposed to converge to $+\infty$. One can write

$$\frac{1}{2NT} \int_0^{2NT} \left(s_a(t)^g - E(s_a(t)^g)\right) e^{-2i\pi \frac{t}{2T}} dt = \frac{1}{N} \sum_{n=0}^{N-1} \xi_n, \tag{13}$$

where $(\xi_n)_{n\in\mathbb{N}}$ is the sequence defined by $\xi_n = \frac{1}{2T}\int_0^{2T}(s_a(t+2nT)^g - E(s_a(t+2nT)^g))e^{-2i\pi\frac{t}{2T}}dt$ for each n. It can be shown that $(\xi_n)_{n\in\mathbb{N}}$ is a zero mean stationary sequence which verifies the strong law of large numbers: in other words, the sum $\mathcal{S}_N$ defined by $\mathcal{S}_N = \frac{1}{N}\sum_{n=0}^{N-1} \xi_n$ converges to $E(\xi_0) = 0$. In order to briefly justify the latter claim, it can be shown that $E(|\mathcal{S}_N|^4) = O(\frac{1}{N^2})$. By the Borel-Cantelli lemma, this condition implies that $\mathcal{S}_N$ converges to zero almost surely. In other words, $\frac{1}{S}\int_0^S s_a(t)^g e^{-2i\pi\frac{t}{2T}}dt$ converges almost surely to $\lim_{S\to\infty} \frac{1}{S}\int_0^S E(s_a(t)^g)e^{-2i\pi\frac{t}{2T}}dt$ if S tends to infinity. Now we recall that for each g, $s_a(t)^g$ coincides with a CPM signal of index gh. Therefore, Proposition 1 implies that

$$\lim_{S\to\infty} \frac{1}{S} \int_0^S E(s_a(t)^g) e^{-2i\pi\frac{t}{2T}} dt \neq 0$$

only if g is an odd integer multiple of $g_0 = \frac{1}{h}$.

B. Presentation of the Sub-optimum Estimate

In this paragraph, we show how to use Proposition 2 in order to estimate $g_0 = \frac{1}{h}$ from the discrete-time received signal $(y(k))_{k=0,\ldots,MN-1}$ defined by (5). Using Propositions 1 and 2, we obtain immediately that g_0 is the smallest positive real number for which:

$$\lim_{N\to\infty} \frac{1}{NM} \sum_{k=0}^{MN-1} s(kT)^{g_0} e^{-i\pi\frac{k}{M}} \neq 0\,, \tag{14}$$

where we recall that discrete-time signal $s(k)$ is defined by $s(k) = s_a(kT/M + T - \tau)$. As y is a scaled and noisy version of $s(k)_{k\in\mathbb{Z}}$, this property suggests to estimate $g_0 = \frac{1}{h}$ as follows. For each $g > 0$, let $r_N(g)$ be defined by

$$r_N(g) = \frac{1}{NM} \sum_{k=0}^{NM-1} \left(\frac{y(k)}{|y(k)|}\right)^g e^{-i\pi\frac{k}{M}}\ . \tag{15}$$

In the noiseless case, $r_N(g)$ converges to 0 if g is not an odd integer multiple of g_0. Therefore, we propose to estimate g_0 by maximizing

$$J_N(g) = |r_N(g)|^2 \tag{16}$$

w.r.t. g over an interval $[g_{min}, g_{max}]$. The corresponding estimate of g_0 is denoted by $\hat{g}_N$, and the index h is of course estimated by $\hat{h}_N = \frac{1}{\hat{g}_N}$. In order to avoid ambiguities, search interval $[g_{min}, g_{max}]$ should be chosen so that $\frac{1}{h} \in [g_{min}, g_{max}]$ and $\frac{3}{h} > g_{max}$. This restriction is motivated by the observation that if N tends to infinity, the maximum of function $J_N(g)$ is reached when g coincides with an odd integer multiple of $\frac{1}{h}$: ambiguities may thus occur if the search interval contains an odd integer multiple of $\frac{1}{h}$. Of course, defining interval $[g_{min}, g_{max}]$ so that the previous property holds does not seem so easy at the receiver side because the modulation index h is unknown. However, it is reasonable to assume that an *a priori* information on the possible values of h is available and that it allows to determine an acceptable search interval. Assume for example that h is known to belong to an interval $[h_{min}, h_{max}]$, where $h_{max} < 3\,h_{min}$. Then every element h of $[h_{min}, h_{max}]$ is such that $\frac{1}{h} \in [\frac{1}{h_{max}}, \frac{1}{h_{min}}]$ and $\frac{3}{h} > \frac{1}{h_{min}}$. Therefore, interval $[\frac{1}{h_{max}}, \frac{1}{h_{min}}]$ does not contain any odd integer multiple of the modulation index. The interval $[h_{min}, h_{max}] = [0.3, 0.9]$ which contains the most often encountered values of the modulation indices of CPM modulations satisfies the above condition.

Remark 1: **Motivation for the use of sequence** $\left(\frac{y(k)}{|y(k)|}\right)$ **in** (15).
One could also replace $\frac{y(k)}{|y(k)|}$ by $y(k)$ in (15). In the noiseless case, ratio $\frac{y(k)}{|y(k)|}$ is of course equal to sample $y(k)$ up to a constant term. But in the presence of noise, the use of (15) cancels in some sense the contribution of the noise to the modulus of the received signal, and provides better experimental results. This observation could be confirmed theoretically by comparing the asymptotic behaviors of the estimators derived from both possible definitions of function $r_N(g)$. Due to the lack of space, this point is not developed in the sequel.

Remark 2: **Practical implementation**.
Function $g \to J_N(g)$ is not convex. Therefore, we propose to compute the values taken by J_N on a discrete grid of the search interval (coarse search), and then to initialize a Newton

maximization algorithm with the argument of the maximum of $J_N(g)$ on the grid in order to refine the estimate (fine search).

Remark 3: **On the computation of sequence** $(y(k)^g)$.

The main problem of the present estimate comes from the computation of sequence $(y(k)^g)_{k=0,\dots,MN-1}$ for each g of the search interval. This crucial step requires indeed to unwrap the phase of y adequately. In other words, for each $k = 0, \dots, MN-1$, a relevant determination of $\mathrm{Arg}(y(k))$ should be chosen to compute $y(k)^g = |y(k)| \exp(ig\mathrm{Arg}(y(k)))$. We first address the noiseless case. Then, $y(k)$ coincides with $\alpha s(k)$. In order to obtain a consistent estimate of g_0, one has to select for each k the determination defined by $\mathrm{Arg}(y(k)) = \mathrm{Arg}(\alpha) + \psi(k)$. In the absence of noise, this particular determination of $\mathrm{Arg}(y(k))$ can always be selected among every possible ones. Indeed, using equation (6) and the fact that coefficients $(\phi_{j,m})_{j,m}$ belong to the interval $[0,1]$, we obtain that the phase variation $\psi(k)-\psi(k-1)$ between two consecutive samples is so that $\psi(k)-\psi(k-1) < \pi h$ for each k. If as usual h satisfies $h < 1$, then $\psi(k) - \psi(k-1)$ is also strictly less than π. Thanks to this remark, one can easily check that the following recursive procedure allows to extract the correct determination $\mathrm{Arg}(y(k)) = \mathrm{Arg}(\alpha) + \psi(k)$ for each k: assuming that the correct determination $\mathrm{Arg}(y(k-1))$ has been identified at time $k-1$, $\mathrm{Arg}(y(k))$ is defined as the determination for which $|\mathrm{Arg}(y(k)) - \mathrm{Arg}(y(k-1))|$ is minimum. In the less usual case $h > 1$, the inequality $\psi(k)-\psi(k-1) < \pi$ holds as long as the oversampling factor M is large enough. Therefore, the correct determination of $\mathrm{Arg}(y(k))$ can still be selected thanks to the same procedure. In the noisy case, this unwrapping procedure is still used. However, the influence of the additive noise $b(k)$ on the associated determination $\mathrm{Arg}(y(k))$ may produce some phase unwrapping errors. Numerical results of section VI illustrate the effect of such errors on the performance of the estimator.

Remark 4: **Computational cost**.

The evaluation of $J_N(g) = |r_N(g)|^2$ requires about N operations for each g. The unwrapping step requires about N operations. But since this procedure only depends on sequence $(y(k))_{k\in\mathbb{Z}}$, it does not have to be applied for each value of g. The computational cost of the coarse search is thus equal to $N_g N + N$ where N_g represents the number of points of the grid over which J_N is evaluated. In our simulations, we have chosen $N_g = N/10$: about $\frac{N^2}{10}$

operations are therefore necessary to the coarse search. The complexity of the fine search depends on the number of iterations of the Newton algorithm. In our experiments, the number of iterations is set to 10. The corresponding number of operations is about $10N$, and can thus be neglected w.r.t. the cost of the coarse search. Although our procedure is of course computationally more expensive than the estimate of [5], it is less demanding than the Expectation/ Maximization algorithm. Indeed, we recall that the Expectation step requires about $4.5N^2 2^L$ operations at each iteration, while the complexity of the maximization step depends on the algorithm used to implement the minimization of function R defined by (9) ; in any case, this step also requires a multiple of N^2 operations. It is therefore clear that the complexity of our new sub-optimum estimate is significantly lower than the complexity of the maximum likelihood estimate.

C. The case where the symbol period and the carrier frequency are unknown.

We now briefly indicate how it is possible to estimate h when T and a possible residual carrier frequency f_0 are unknown. If $f_0 \neq 0$, the received signal $y_a(t)$ can be written as:

$$y_a(t) = \alpha s_a(t - \tau) e^{2i\pi f_0 t} + b_a(t) \; .$$

Assume for the sake of simplicity that the additive noise $b_a(t)$ is zero or negligible. Using Propositions 1 and 2, it is easy to check that

$$\lim_{S \to \infty} \frac{1}{S} \int_0^S y_a(t)^g e^{-2i\pi\alpha t} dt \neq 0$$

if and only if

- g is an odd integer multiple of $\frac{1}{h}$ and $\alpha = n\frac{1}{2T} + g f_0$ for some integer $n \in \mathbb{Z}$.

or

- g is an even integer multiple of $\frac{1}{h}$ and $\alpha = n\frac{1}{T} + g f_0$ for some integer $n \in \mathbb{Z}$.

This property potentially allows to estimate h by a joint search w.r.t. (g, α). However, some *a priori* knowledge about the range of values taken by the 3 parameters (h, f_0, T) is required in order to avoid ambiguities. More precisely, this *a priori* information should provide search intervals I_g for g and I_α for α such that

$$\lim_{S \to \infty} \frac{1}{S} \int_0^S y_a(t)^g e^{-2i\pi\alpha t} dt \neq 0$$

for $g \in I_g$ and $\alpha \in I_\alpha$ if and only if

$$g = g_0 = \frac{1}{h} \text{ and } \alpha = \alpha_0 = \frac{1}{2T} + \frac{f_0}{h} \quad .$$

In this case, one can jointly estimate g_0, α_0, and thus parameter h itself. In practice, such intervals I_g and I_α can be defined from initial estimates of f_0 and T. Such initial estimates can be obtained as follows. T can be estimated by noticing that $\frac{1}{T}$ is the smallest cyclic frequency of the instantaneous frequency of $y_a(t)$[1]. Similarly, a coarse estimate of f_0 can be obtained by observing that f_0 coincides with the average of the instantaneous frequency of $y_a(t)$.

Because f_0 and T may in practice be unknown in the context of passive listening, we should study in detail the performance of this joint estimation procedure of (g_0, α_0). Due to the lack of space, we rather focus on the case where f_0 and T are known, and will be address the general case in a forthcoming paper. Note that the results on the asymptotic behavior of the joint estimate of (g_0, α_0) cannot be directly deduced from the content of the present paper.

V. Asymptotic Analysis of the proposed Estimate

In this section, we assume without restriction that parameter α is equal to 1 in order to simplify the notations. We moreover suppose that $\exp(i\pi h \sum_{-\infty}^{-1} a_j) = 1$, so that sequence $\psi(nM+m)$ defined by (6) is equal to $\pi h \left(\sum_{j=0}^{n-L} a_j + \sum_{j=0}^{L-1} a_{n-j}\phi_{j,m} \right)$ modulo 2π if $n \geq (L-1)$. It is worth noting that this assumption does not imply any restriction to the asymptotic analysis which follows. Indeed, $\pi h \sum_{j=-\infty}^{-1} a_j$ can be considered as a random initial phase which does not have any influence on cost function $J_N(g)$ (and *a fortiori* on the asymptotic behavior of the estimate). This claim can be shown by noticing that for each g, $J_N(g)$ can also be written as $J_N(g) = \left| \exp\left(-i\pi h g \sum_{j=-\infty}^{-1} a_j \right) r_N(g) \right|^2$: plugging (15) and (6) into the previous equation, we conclude that $J_N(g)$ does actually not depend on the initial phase.

[1] $\frac{1}{T}$ is also the smallest strictly positive cyclic frequency of signal $y_a(t)$ itself, but due to bandwidth constraints, the second order cyclic statistics of $y_a(t)$ at $\frac{1}{T}$ are close to zero, and thus provide estimates of T which are more sensitive to the presence of noise.

A. The Noiseless Case

We first study the asymptotic distribution of the estimate $\hat{g}_N$ of g_0 in the noiseless case, and deduce from this the distribution of $\hat{h}_N$.

Before presenting the main results of this section, we first study the behavior of $r_N(g_0)$ in order to get some insights on the parameters which may influence the performance of the estimate. Using (15) and (33), and noticing that:

$$e^{-2i\pi \frac{nM+m}{2M}} = (-1)^n e^{-i\pi \frac{m}{M}},$$

$r_N(g)$ has the following form:

$$r_N(g) = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \exp\left[ig\psi(nM+m)\right](-1)^n e^{-i\pi \frac{m}{M}}. \tag{17}$$

We now use the above expression to evaluate (17) at the point g_0. We note that $\exp\left[i\pi \sum_{j=0}^{n-L} a_j\right]$ coincides with $\pm(-1)^n$, and assume that it is equal to $(-1)^n$. As $g_0 h = 1$, $\exp[ig_0\psi(nM+m)]$ can be written as

$$\begin{aligned} \exp[ig_0\psi(nM+m)] &= (-1)^n \prod_{j=0}^{L-1} \exp\left[i\pi a_{n-j}\phi_{j,m}\right], \\ &= (-1)^n \prod_{j=0}^{L-1} \left(\cos(\pi\phi_{j,m}) + i a_{n-j} \sin(\pi\phi_{j,m})\right). \end{aligned} \tag{18}$$

Expanding the right-hand side of (18), it is clear that $(-1)^n \exp ig_0\psi(nM+m)$ coincides with the sum $(\prod_{j=0}^{L-1} \cos(\pi\phi_{j,m})) + \epsilon(nM+m)$, where $\epsilon(nM+m)$ represents a zero mean random variable which only depends on data symbols a_n, a_{n-1}, $\ldots$, a_{n-L+1} and on coefficients $(\phi_{j,m})_{j=0,\ldots,L-1}$. Now, plugging the above expression of $(-1)^n \exp ig_0\psi(nM+m)$ into (17) leads to rewrite $r_N(g_0)$ as

$$r_N(g_0) = \lambda + \frac{1}{N} \sum_{n=0}^{N-1} \tilde{\epsilon}(n) \tag{19}$$

where $\tilde{\epsilon}(n)$ represents the random process defined by $\tilde{\epsilon}(n) = \frac{1}{M} \sum_{m=0}^{M-1} \epsilon(nM+m) e^{-i\pi \frac{m}{M}}$ and where λ is the deterministic constant defined by

$$\lambda = \frac{1}{M} \sum_{m=0}^{M-1} \prod_{j=0}^{L-1} \cos(\pi\phi_{j,m}) e^{-i\pi \frac{m}{M}}. \tag{20}$$

$(\tilde{\epsilon}(n))_{n\in\mathbb{Z}}$ is a zero mean stationary sequence for which the strong law of large numbers holds, i.e. $\frac{1}{N}\sum_{n=0}^{N-1}\tilde{\epsilon}(n)$ converges almost surely toward zero as N tends to infinity. Thus, using (19), $r_N(g_0)$ converges toward λ. In other words, $|\lambda|^2$ can be interpreted as the asymptotic magnitude of the cost function J_N at g_0. The above result gives us already the insight that the modulus $|\lambda|$ of λ has a crucial influence on the performance of the proposed estimate.

In the sequel, we define $\rho = |\lambda|$ and $\theta = \mathrm{Arg}(\lambda)$. We also need to introduce the random process $\overline{\epsilon}(n) = \frac{1}{\rho}i(\lambda\tilde{\epsilon}(n)^* - \lambda^*\tilde{\epsilon}(n))$.

In order to study the asymptotic behavior $\hat{g}_N$, we first note that $J_N'(\hat{g}_N)$ is equal to zero. Using the Taylor expansion of $J_N'(\hat{g}_N)$ around g_0, we conjecture as usual that $N(\hat{g}_N - g_0)$ and $-\frac{J_N'(g_0)}{J_N''(g_0)/N}$ have the same asymptotic behavior. Thus, the asymptotic study of the estimate requires the separate study of $J_N'(g_0)$ and $J_N''(g_0)/N$.

The asymptotic behaviors of $J_N'(g_0)$ and $J_N''(g_0)$ are given by the following results which are proved in the appendix :

Lemma 1: The first derivative $J_N'(g_0)$ of J_N at point g_0 can be written as

$$J_N'(g_0) = \pi h\rho\left(\mu + (\sum_{n=0}^{N-1}\frac{(\sum_{j=0}^{n-L}a_j)}{N^{3/2}})(\frac{\sum_{n=0}^{N-1}\overline{\epsilon}(n)}{N^{1/2}}) - \frac{1}{N}\sum_{n=0}^{N-1}(\sum_{j=0}^{n-L}a_j)\overline{\epsilon}(n)\right) + \mathrm{O}_P(\frac{1}{N^{1/2}}) \tag{21}$$

where notation O_P stands for *bounded in probability* and where μ is the deterministic term defined by

$$\mu = -\frac{2}{M}\sum_{m=0}^{M-1}\left(\sum_{j=0}^{L-1}\phi_{j,m}\,\sin(\pi\phi_{j,m})\prod_{\substack{k=0\\k\neq j}}^{L-1}\cos(\pi\phi_{k,m})\right)\cos(\frac{\pi m}{M}+\theta). \tag{22}$$

Lemma 2: $-\frac{J_N''(g_0)}{N}$ can be written as

$$-\frac{J_N''(g_0)}{N} = 2\rho^2(\pi h)^2\left(\frac{1}{N^2}\sum_{n=0}^{N-1}(\sum_{j=0}^{n-L}a_j)^2 - \frac{1}{N^3}(\sum_{n=0}^{N-1}\sum_{j=0}^{n-L}a_j)^2\right) + O_P(\frac{1}{N^{1/2}}). \tag{23}$$

The asymptotic behavior of $N(\hat{g}_N - g_0)$ can be characterized by using the so-called functional central limit theorem ([1]). In order to introduce the main result of this paragraph, we define on $[0,1]$ the two following stochastic processes

$$W_1^{(N)}(t) = \frac{1}{N^{1/2}} \sum_{n=0}^{[Nt]} a_{n-L} , \tag{24}$$

$$W_2^{(N)}(t) = \frac{1}{N^{1/2}} \sum_{n=0}^{[Nt]} \overline{\epsilon}(n) , \tag{25}$$

where $[x]$ denotes the greatest integer less than or equal to x. We also define the 2-dimensional stochastic process $\mathbf{W}^{(N)}(t) = (W_1^{(N)}(t), W_2^{(N)}(t))^T$. The standard central limit theorem asserts that the sequence of random vectors $(\mathbf{W}^{(N)}(1))_{N\in\mathbb{Z}}$ converges in distribution towards a 2-dimensional Gaussian distribution whose covariance matrix $\mathbf{\Gamma}$ is given by

$$\mathbf{\Gamma} = \sum_{k\in\mathbb{Z}} E\left(\begin{bmatrix} a_{n+k-L} \\ \overline{\epsilon}(n+k) \end{bmatrix} [a_{n-L}\ \overline{\epsilon}(n)] \right) . \tag{26}$$

The functional central limit theorem is a much more powerful result stating that the probability distribution of the stochastic process $(\mathbf{W}^{(N)}(t))_{t\in[0,1]}$ converges toward the probability distribution of a 2-dimensional Brownian motion $(\mathbf{W}(t))_{t\in[0,1]}$ of covariance matrix $\mathbf{\Gamma}$ as N tend to infinity. We recall that a Brownian motion $(\mathbf{W}(t))_{t\in[0,1]}$ of covariance matrix $\mathbf{\Gamma}$ is a Gaussian stochastic process such that for each t and each s in the interval $[0,1]$,

- $E(\mathbf{W}(t)) = 0$,
- $E(\mathbf{W}(t)\mathbf{W}(s)^T) = \mathbf{\Gamma}\ \min(t,s)$.

In particular, this implies that if f is a functional defined on a certain space of functions defined on $[0,1]$ (see [1] for details), then the random variable $f(\mathbf{W}^{(N)})$ converges in distribution toward the random variable $f(\mathbf{W})$. In order to illustrate this, we mention the following corollaries of the functional central limit theorem:

$$\int_0^1 \mathbf{W}^{(N)}(t)dt = \frac{1}{N^{3/2}} \begin{pmatrix} \sum_{n=0}^{N-1} \sum_{j=0}^{n} a_{j-L} \\ \sum_{n=0}^{N-1} \sum_{j=0}^{n} \overline{\epsilon}(j) \end{pmatrix}$$

converges in distribution toward the Gaussian random vector $\int_0^1 \mathbf{W}(t)dt$. Moreover,

$$\int_0^1 (W_1^{(N)}(t))^2 dt = \frac{1}{N^2} \sum_{n=0}^{N-1} (\sum_{j=0}^{n} a_{j-L})^2$$

converges in distribution toward the random variable $\int_0^1 (W_1(t))^2 dt$. Finally, using more involved arguments,

$$\int_0^1 W_1^{(N)}(t) dW_2^{(N)}(t) = \frac{1}{N} \sum_{n=0}^{N-1} (\sum_{j=0}^{n} a_{j-L}) \overline{\epsilon}(n)$$

converges in distribution toward the stochastic integral (in the Ito's sense) $\int_0^1 W_1(t)dW_2(t)$. Using in an appropriate way these ideas, it is straightforward to show the following theorem.

Theorem 1: Denote by ζ the random variable defined by

$$\zeta = \frac{1}{2\pi} \left[\frac{\mu + (\int_0^1 W_1(t)dt)W_2(1) - \int_0^1 W_1(t)dW_2(t)}{(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2} \right] . \tag{27}$$

Then, $N(\hat{g}_N - g_0)$ converges in distribution toward the random variable $\frac{1}{h\rho}\zeta$ ■

As $\hat{h}_N = \frac{1}{\hat{g}_N}$, the transfer theorem ([2]) implies that $N(\hat{h}_N - h)$ and $-\frac{1}{g_0^2} N(\hat{g}_N - g_0) = -h^2 N(\hat{g}_N - g_0)$ have the same asymptotic distribution. We immediately deduce from this the following theorem:

Theorem 2: $N(\hat{h}_N - h)$ converges in distribution toward the random variable $-\frac{h}{\rho}\zeta$ ■

♦ **General comments.**

- We first observe that the convergence rate of $\hat{h}_N$ to h is equal to $\frac{1}{N}$ in the sense that $\hat{h}_N = h + O_P(\frac{1}{N})$. This is in contrast with more standard estimation problems in which the convergence rate is equal to $\frac{1}{N^{1/2}}$. We mention in particular that the estimation procedures presented in [5] and [6] have a convergence rate equal to $\frac{1}{N^{1/2}}$.
- The asymptotic mean square error of $\hat{h}_N$ increases with h^2. Therefore, the smaller the modulation index, the better the performance of the estimate.
- The asymptotic mean square error of $\hat{h}_N$ increases with $\frac{1}{\rho^2}$. As expected, this shows that the performance of the estimate depends crucially on the value of ρ.

• By the Jensen inequality ([7]), $\int_0^1 W_1(t)^2dt - (\int_0^1 W_1(t)dt)^2$ is almost surely positive. It is thus reasonable to conjecture that $E(\zeta)$ is in general non zero, and that its sign is equal to the sign of

$$E\left(\mu + (\int_0^1 W_1(t)dt)W_2(1) - \int_0^1 W_1(t)dW_2(t)\right). \tag{28}$$

We now give a more informative expression of (28). We first note that $E(W_1(t)W_2(1)) = \Gamma_{1,2}|1-t|$, where $\Gamma_{1,2}$ denotes the non-diagonal coefficient of matrix $\mathbf{\Gamma}$ (i.e. the correlation coefficient between processes W_1 and W_2). Thus, we get that $E(\int_0^1 W_1(t)dt)W_2(1)) = \frac{\Gamma_{1,2}}{2}$. Now it can be shown that $\int_0^1 W_1(t)dW_2(t)$ is a random term of zero mean. Putting all the pieces together, we conclude that the expectation given by (28) is equal to $\mu + \frac{\Gamma_{1,2}}{2}$. This remark shows that $\hat{h}_N$ is biased, and that it is positively (resp. negatively) biased if and only if $\mu + \frac{\Gamma_{1,2}}{2} < 0$ (resp. $\mu + \frac{\Gamma_{1,2}}{2} > 0$).

♦ **The case** $L = 1$.

In order to get more insights on the significance of Theorem 2, we now study further the case $L = 1$ for which more informative expressions of the various parameters can be obtained. We also assume that the shaping filter satisfies the following condition:

$$g_a(t) = g_a(T-t)$$

for $t \in [0, T]$. This condition is equivalent to

$$\phi_a(T-t) = 1 - \phi_a(t) \tag{29}$$

for $0 \leq t \leq T$. Substituting 1 for L in (18), we obtain immediately that for each $n = 0 \ldots N-1$ and for each $m = 0 \ldots M-1$, $\epsilon(nM+m) = i\sin(\pi\phi_m)a_n$. Therefore, sequence $\overline{\epsilon}(n)$ is given by

$$\overline{\epsilon}(n) = [\frac{2}{M}\sum_{m=0}^{M-1} \sin(\pi\phi_m)\cos(\theta + \frac{\pi m}{M})]\, a_n. \tag{30}$$

The above relation implies that the second component $W_2(t)$ of $\mathbf{W}(t)$ coincides with $[\frac{2}{M}\sum_{m=0}^{M-1} \sin(\pi\phi_m)\cos(\theta + \frac{\pi m}{M})]\, W_1(t)$. Thus, the limit distribution of $\hat{h}_N$ depends only on the single unit-variance Brownian motion $W_1(t)$.

In order to discuss on the influence of the oversampling factor M, we first consider the case $M = 1$.

Proposition 3: Assume that $M = 1$. If ϕ_0 denotes $\phi_0 = \phi_a(T - \tau)$, and if we assume that $\tau > \frac{T}{2}$, then, $N(\hat{h}_N - h)$ converges in distribution toward the random variable

$$\frac{h \sin \pi\phi_0}{\pi \cos \pi\phi_0} \left[\frac{\phi_0 + \int_0^1 W_1(t) dW_1(t) - (\int_0^1 W_1(t)dt) W_1(1)}{(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2} \right] . \tag{31}$$

Proof. It is easy to check that parameter λ coincides with $\cos \pi\phi_0$, and is thus real. Recalling that function $\phi_a(t)$ is increasing and that $\phi_a(T/2) = \frac{1}{2}$ by (29), it is clear that $0 \leq \phi_a(t) \leq \frac{1}{2}$ on $[0, \frac{T}{2}]$. Hence, $0 \leq \phi_0 \leq \frac{\pi}{2}$ and $\lambda \geq 0$. The argument θ of λ is equal to 0, and μ reduces to $\mu = -2\phi_0 \sin \pi\phi_0$. This yields immediately (31).

Remark 5: As $\phi_0 = \phi_a(T - \tau)$, the above expression shows that the performance of the estimator depends in a crucial way of the (unknown) time delay τ. If τ is close to $\frac{T}{2}$, then $\cos \pi\phi_0$ is close to zero, and the asymptotic variance of the proposed estimate increases.

That is one of the reasons why it is beneficial to oversample the received signal: this allows to make the parameters characterizing the asymptotic distribution $\hat{h}_N - h$ nearly independent of τ.

In order to show the latter claim, we now assume that M tends to infinity.

Proposition 4: If M tends to infinity, random variable ζ is such that

$$\zeta \to \frac{T}{\pi} \frac{\left[\int_0^{T/2} (1 - 2\phi_a(t)) \sin \pi\phi_a(t) \cos \pi t/T \, dt \right]}{(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2} . \tag{32}$$

Proof. It is easy to check that

$$\lambda \to \frac{1}{T} \int_0^T \cos \pi\phi_a(t) e^{-i\pi t/T} \, dt .$$

Using property (29), it can be shown that $\frac{2}{T} \int_0^T \sin \pi\phi_a(t) \cos(\pi t/T) dt = 0$. Therefore, λ converges a strictly positive real number given by

$$\lambda \to \frac{1}{T} \int_0^T \cos \pi\phi_a(t) \cos \pi t/T \, dt .$$

Thus, θ tends to zero. Moreover, it is easy to check that

$$\mu \to \frac{2}{T} \int_0^{T/2} (1 - 2\phi_a(t)) \sin \pi\phi_a(t) \cos \pi t/T \, dt .$$

Finally, $\overline{\epsilon}(n) \to \left[\frac{2}{T}\int_0^T \sin \pi\phi_a(t)\cos(\pi t/T)dt\right] a_n = 0$. This implies that the second component $W_2(t)$ of Brownian process $\mathbf{W}(t)$ is actually equal to zero when the oversampling factor M tends to infinity. This remark leads immediately to (32).

Remarks

- As expected, these calculations show that the parameters characterizing the asymptotic distribution of the estimate do not depend on the timing. It is therefore quite useful in practice to oversample the received signal.
- As $\phi_a(t) \leq \frac{1}{2}$ on $[0, \frac{T}{2}]$, μ converges toward a strictly positive limit. As $(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2$ is almost surely positive, the estimation error $\hat{h}_N - h$ is almost surely negative as M tends to infinity.
- We finally remark that the critical parameter $\rho = \lambda$ is maximum if and only if $\phi_a(t)$ coincides with $\frac{t}{T}$ on $[0, T]$ (this can be shown by using the Schwartz inequality). This shows that if $L = 1$, then the performance of the estimator is optimum if the shaping filter coincides with the rectangular window on $[0, T]$.

B. The Noisy Case

Here we assume that the received samples are corrupted by an additive noise: $y(k) = \alpha s(k) + b(k)$, and assume without restriction that parameter α is equal to $\alpha = 1$ in order to simplify the notations. We first distinguish between the contribution of the noise to the phase and its contribution to the modulus of the received signal: for each $k = 0 \ldots NM - 1$, $y(k) = \exp\left[i\psi(k)\right]\left(1 + \tilde{b}(k)\right)$, where $\tilde{b}(k) = e^{-i\psi(k)}b(k)$. As $b(k)$ is assumed to be a complex Gaussian random variable which is independent of $e^{i\psi(k)}$, $\tilde{b}(k)$ is still a Gaussian random variable which is independent of $e^{i\psi(k)}$. We set $1 + \tilde{b}(k) = r(k)e^{i\delta(k)}$ where $\delta(k) \in]-\pi, \pi]$. $(\delta(k))_{k\in\mathbb{Z}}$ is a stationary zero mean sequence. The received samples can finally be written as follows:

$$y(k) = r(k) \exp i \left[\psi(k) + \delta(k)\right] . \tag{33}$$

As mentioned in subsection IV-B, the calculation of $y(k)^g$ requires to unwrap the phase of $y(k)$. The additive noise can of course produce some phase unwrapping errors which are unfortunately very difficult to take into account. A rigorous extension of the above asymptotic analysis seems therefore difficult. However, the results of subsection V-A may be extended if phase unwrapping errors are neglected, i.e. if it is assumed that the unwrapped phase of $y(k)$ coincides with $\psi(k) + \delta(k)$ for each k. Under this assumption, we establish in this subsection that $N(\hat{h}_N - h)$ converges in distribution towards a certain non Gaussian random variable. In section VI, we compare our theoretical predictions with empirical estimates of the probability density of the $\hat{h}_N$, and conclude on the domain of validity of our hypotheses.

Using some algebra and the functional central limit theorem, it is possible to show the following result. In the sequel, notation *no* stands for the noisy case.

Theorem 3: Denote by ζ_{no} the random variable defined by

$$\zeta_{no} = \frac{1}{2\pi} \left[\frac{\mu_{no} + (\int_0^1 W_1(t)dt) W_{2,no}(1) - \int_0^1 W_1(t) dW_{2,no}(t)}{(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2} \right] , \tag{34}$$

where μ_{no} denotes the deterministic term defined by:

$$\mu_{no} = \mu - 2 \, \frac{E(\delta \sin(g_0 \delta))}{E(\cos(g_0 \delta))} \, . \tag{35}$$

and where $W_1(t)$ is the same Brownian motion as in the noiseless case. $W_{2,no}(t)$ is such that $\mathbf{W}_{no}(t) = (W_1(t), W_{2,no}(t))^T$ is a 2-dimensional Brownian motion whose covariance matrix $\mathbf{\Gamma}_{no}$ verifies $\Gamma_{1,2,no} = \Gamma_{1,2}$, where $\Gamma_{1,2}$ denotes the non-diagonal coefficient of matrix $\mathbf{\Gamma}$ in the noiseless case. Finally,

- $N(\hat{g}_N - g_0)$ converges in distribution toward the random variable $\frac{1}{h\rho}\zeta_{no}$.
- $N(\hat{h}_N - h)$ converges in distribution toward the random variable $-\frac{h}{\rho}\zeta_{no}$ ■

We first note that the form of expression (34) is quite similar to the form of the limit distribution in the noiseless case, which is given by Theorem 1. However, the additive noise has of course an influence on the bias and on the variance of the estimate $\hat{h}_N$. Indeed, it can be shown that the variance of the second Brownian motion $W_{2,no}(t)$ increases in the presence of noise. It is therefore reasonable to conjecture that the variance of variable ζ_{no} is also modified. Moreover, it can be shown as in the previous subsection that the estimate $\hat{h}_N$ is biased, and that it is positively (resp. negatively) biased if and only if $\mu_{no} + \frac{\Gamma_{1,2}}{2} < 0$

(resp. $\mu_{no} + \frac{\Gamma_{1,2}}{2} > 0$). As $E(\delta \sin(g_0\delta))$ is a positive number, we note that the term

$$\frac{1}{2\pi} \left[\frac{-2\, E(\delta \sin(g_0\delta))/E(\cos(g_0\delta))}{(\int_0^1 W_1(t)^2 dt) - (\int_0^1 W_1(t)dt)^2} \right] \tag{36}$$

corresponds to an almost surely negative additional term which is specifically due to the additive noise. If ζ represents the random variable defined by (27) in the noiseless case, this implies that $E(\zeta_{no}) - E(\zeta) < 0$. Recalling that $N(\hat{h}_N - h)$ converges in distribution toward the random variable $-\frac{h}{\rho}\zeta_{no}$, we deduce from the previous claim that the additive noise produces an additional positive bias.

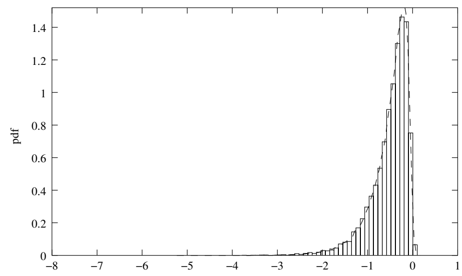
We finally stress that the previous result rests on the assumption that no error occurs during the unwrapping step. Consequently, theorem 3 allows to predict the asymptotic behavior of the practical estimator only when the signal to noise ratio (SNR) is large enough so that the unwrapping procedure performs well. The question of which level of SNR is required so that our theoretical results fit to the experimental ones is addressed in the next section.

VI. Performances and simulations

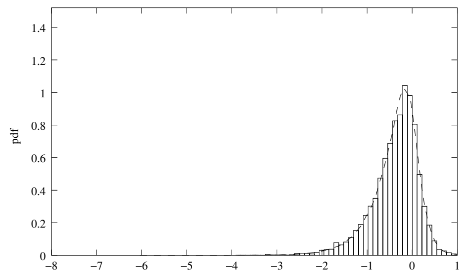
A. *Comparison to empirical results*

Here, we compare our theoretical predictions with empirical results. We first give the parameters used for the simulations. The number N of signaling intervals is set to $N = 1000$. The time delay τ is equal to $\tau = 0.21\,T$. The additive noise is assumed to be white in the frequency interval $[-\frac{1}{T}, \frac{1}{T}]$, so that its variance σ^2 is given by $\sigma^2 = \frac{2}{T}N_0$. Results presented in the sequel are obtained by using either 1REC modulated signals (i.e. the shaping filter $g_a(t)$ is given by $g_a(t) = \frac{1}{T}$ on $[0, T[$ and $g_a(t) = 0$ elsewhere) of modulation index $h = 0.5$ or 3RC modulated signals (i.e. the shaping filter is the raised-cosine of order $L = 3$, given by $g_a(t) = \frac{1}{LT}(1 - \cos\frac{2\pi t}{LT})\mathbf{I}_{0 \le t \le LT}$) of modulation index $h = 0.7$. Finally, the oversampling factor M is equal to $M = 4$ unless otherwise stated.

As mentioned above, function J_N is first evaluated on a discrete grid, and the argument of the maximum is used to initialize a Newton algorithm. The search interval is equal to $[0.3, 0.9]$ and the number of points of the grid is set to 120. 10 iterations of the Newton algorithm are used. Previous theorems assert that the mean square error of $\hat{h}_N$ is proportional to h^2/N^2 as N tends to infinity. Therefore, we present in the sequel the distribution of the normalized random variable $N(\hat{h}_N - h)/h$, so that it is possible to compare the results obtained in both simulation contexts (i.e. 1REC modulation of index $h = 0.5$ and 3RC modulation of index $h = 0.7$). In each case, the empirical distribution is represented by an histogram based on 10^4 realizations of the random variable $N(\hat{h}_N - h)/h$. The corresponding theoretical probability density function (pdf) is approximated thanks to an histogram based on 10^5 realizations of the limit random variable $-\frac{h}{\rho}\zeta_{no}/h$. Figures 1 and 2 represent the empirical and theoretical distributions respectively for 1REC and 3RC signals for $\frac{E_b}{N_0} = 25$dB and $\frac{E_b}{N_0} = 15$dB, were E_b represents the signal energy per bit.

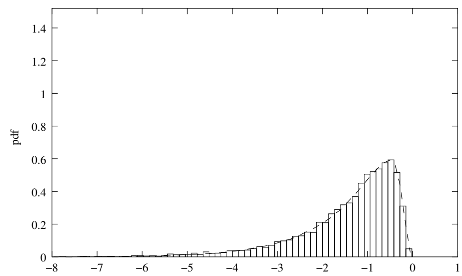


(a) $E_b/N_0 = 25$dB

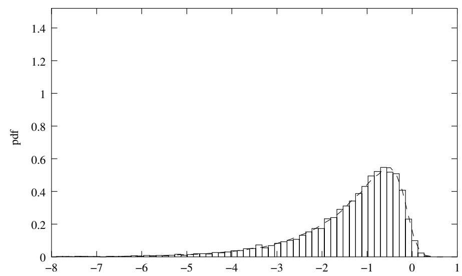


(b) $E_b/N_0 = 15$dB

Fig. 1. Normalized histogram of $N(\hat{h}_N - h)/h$ and limit pdf in case of 1REC signals - $\tau = 0.21T$



(a) $E_b/N_0 = 25$dB

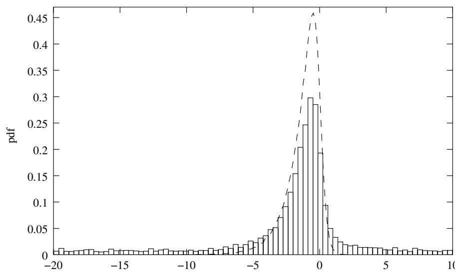


(b) $E_b/N_0 = 15$dB

Fig. 2. Normalized histogram of $N(\hat{h}_N - h)/h$ and limit pdf in case of 3RC signals - $\tau = 0.21T$

We observe that the histogram of the limit distribution $-\frac{h}{\rho}\zeta_{no}/h$ fits to the empirical histogram of $N(\hat{h}_N - h)/h$. The performance is better when a 1REC signal is transmitted. However, in case of a 3RC modulation, the performance obtained either at $\frac{E_b}{N_0} = 25$dB or at $\frac{E_b}{N_0} = 15$dB is almost the same ; in case of 1REC signals, it seems to be slightly more sensitive to the noise level. In the previous section, we have shown that in case of a 1REC modulation and in the absence of noise, the estimation error $\hat{h}_N - h$ is almost surely negative when the oversampling factor M tends to infinity. Figure 1a sustains this claim: when the ratio $\frac{E_b}{N_0}$ is large enough, almost every realizations of $\hat{h}_N - h$ are negative. The above theoretical claim seems therefore to hold whereas the oversampling factor M is just equal to $M = 4$. Furthermore, Figure 2a allows to conjecture that this remark still holds in case of 3RC signals. On the opposite, when $\frac{E_b}{N_0}$ is equal to 15dB, one can see on Figures 1b and 2b that a significant number of realizations are such that $\hat{h}_N - h$ is positive. This illustrates the fact that the additive noise produces an additional positive bias.

We now consider less favorable signal to noise ratios. In Figure 3, $\frac{E_b}{N_0}$ is equal to 10dB and the transmitted signal corresponds to a 3RC modulation. In this case, the approximations that were required in order to derive the limit distribution are no longer valid because of a significant number of phase unwrapping errors. Therefore, theoretical predictions do not fit to the empirical results.



$E_b/N_0 = 10$dB

Fig. 3. Normalized histogram of $N(\hat{h}_N - h)/h$ and limit pdf in case of 3RC signals - $\tau = 0.21T$

B. Asymptotic performances

Here, we study the behavior of the theoretical limit distribution of $\hat{h}_N$ as the number N of signaling intervals tends to infinity. Realizations of random variable $-\frac{h}{\rho}\zeta_{no}/h$ are computed using theorem 3. In order to evaluate the asymptotic performances, we compute $\Delta_\infty = \lim_{N\to\infty} NE(\hat{h}_N/h - 1)$ and $\sigma_\infty^2 = \lim_{N\to\infty} N^2 E(((\hat{h}_N - E(\hat{h}_N))/h)^2)$. Note that Δ_∞ and σ_∞ do not depend on the value of the modulation index. Since the limit distribution of the random variable $N(\hat{h}_N - h)/h$ is not Gaussian, Δ_∞ and σ_∞ do not provide an exhaustive information on the asymptotic behavior of the estimate. However, these values provide a satisfying illustration of the limit distribution. 10^5 realizations of random variable $-\frac{h}{\rho}\zeta_{no}/h$ have been computed in order to estimate Δ_∞ and σ_∞.

1REC ($h = 0.5$) and 3RC ($h = 0.7$) signals are considered in the following simulations. Figures 4 and 5 illustrate the influence of time delay τ on Δ_∞ and σ_∞ in the noiseless case.

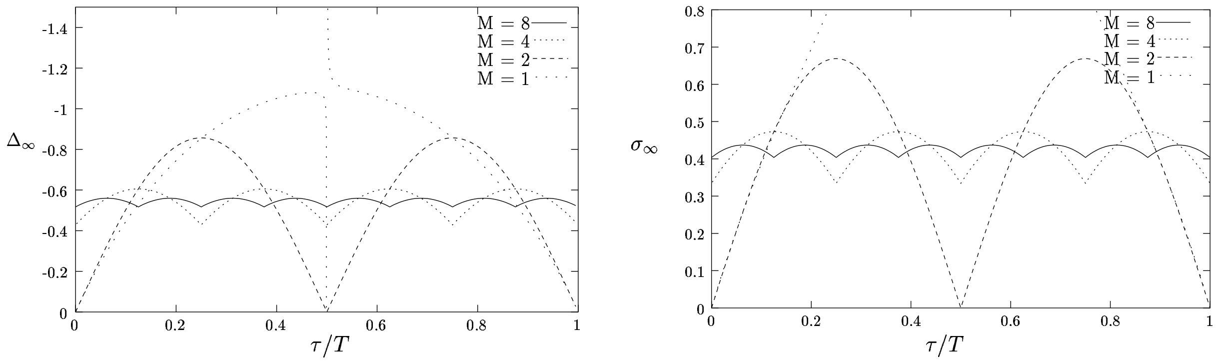


Fig. 4. Δ_∞ and σ_∞ as functions of τ - 1REC signals - noiseless case

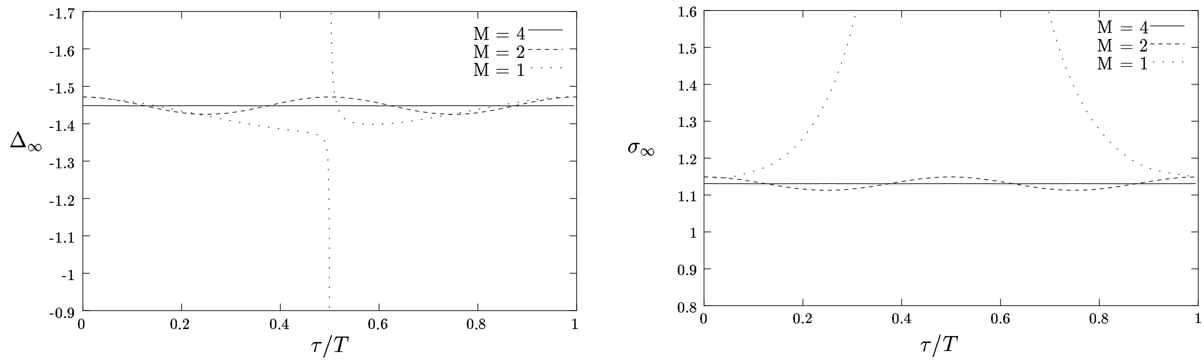
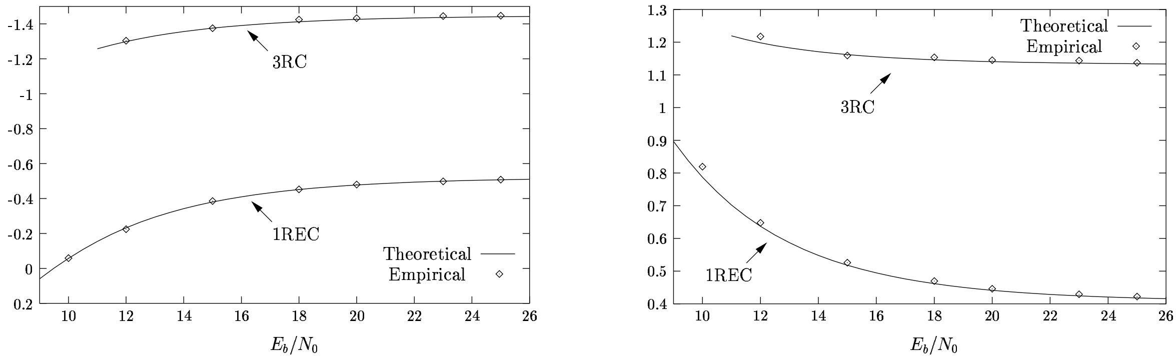


Fig. 5. Δ_∞ and σ_∞ as functions of τ - 3RC signals - noiseless case

As above, we first remark that 1REC signals provide better performances than 3RC signals as long as the oversampling factor is large enough. When the oversampling factor M is equal to $M = 1$, the bias and the variance crucially depend on time delay τ. In particular, Figures 4 and 5 confirm that if $M = 1$ and if τ is close to $\frac{T}{2}$, the asymptotic normalized standard deviation σ_∞ tends to infinity. As mentioned in the previous section, the largest the oversampling factor M, the less the influence of τ on the performances. Oversampling at rate $M = 4$ seems to be sufficient to make the bias and the variance of the estimator nearly independent of the time delay τ.

We now study the influence of the additive noise on the asymptotic performances. The oversampling factor is set to $M = 4$, so that the next results are nearly independent of time delay. As in the previous subsection, the time delay τ is set to $\tau = 0.21T$ and the noise $b_a(t)$ is white in the frequency interval $[-\frac{1}{T}, \frac{1}{T}]$. Figures 6a and 6b represent the behavior of the normalized bias Δ_∞ and the normalized standard deviation σ_∞ respectively, as the signal to noise ratio is varying. We also compare the theoretical and empirical biases and standard deviations.



(a) Normalized bias

(b) Normalized standard deviation

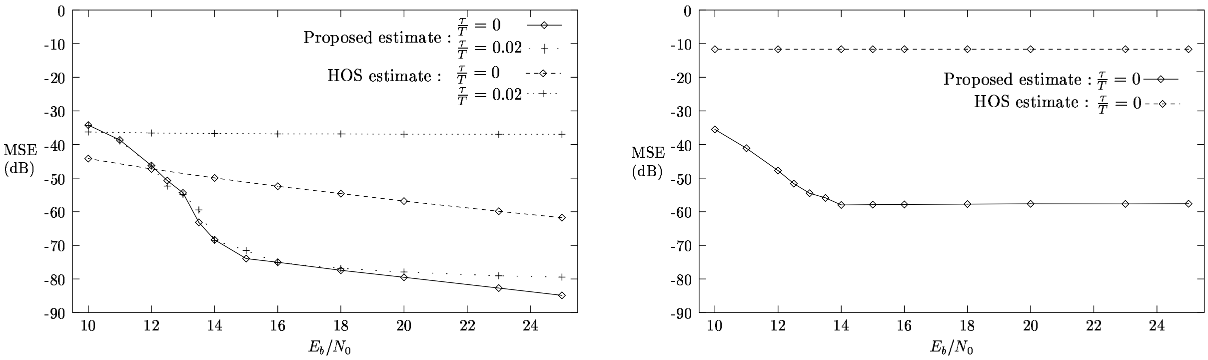
Fig. 6. Theoretical and empirical performance as a function of E_b/N_0

We first note the good fit between the theoretical and empirical results as long as $\frac{E_b}{N_0} \geq$ 10dB in case of a 1REC modulation, $\frac{E_b}{N_0} \geq$ 12dB in case of a 3RC modulation. We also remark that noise level variations seem to have only a slight influence on the performance, specially in case of 3RC signals.

C. Comparison to the estimator based on higher-order statistics (HOS)

We finally compare the performance of the proposed estimator to the performance of the estimator based on higher-order statistics which has been briefly presented in subsection III-B. We recall that the latter estimation method can be applied only when the transmitted signal is a full response CPM signal (i.e. $L = 1$) and when the time delay τ has been correctly compensated (i.e. $\tau = 0$). We denote by $\hat{h}_N^F$ the corresponding (HOS) estimate and we recall that $\hat{h}_N^F$ has been shown to converge toward h at rate $\frac{1}{N^{1/2}}$. In the sequel, we still consider 1REC and 3RC signals. We assume as in [5] that the oversampling factor M is equal to 1. The modulation index is equal to $h = 0.7$. The number of received samples is set to $N = 1000$ unless otherwise stated.

In order to evaluate the performance, we first compare the Mean Square Errors (MSE) of both estimators for different values of E_b/N_0. Mean Square Errors are estimated thanks to 5000 realizations of $\hat{h}_N$ and $\hat{h}_N^F$.



(a) 1REC signals

(b) 3RC signals

Fig. 7. Mean Square Errors as a function of E_b/N_0 - $N = 1000$, $h = 0.7$, $M = 1$.

Figures 7a and 7b illustrate *i)* the effect of the additive noise on the performance, *ii)* the effect of a possible slightly defective synchronization on the performance. When the time delay is equal to zero and a 1REC signal is transmitted, Figure 7a shows that the MSE of both estimators tend to zero as the SNR tends to infinity. This remark confirms previous claims. In the present simulation context, the proposed method outperforms the method based on higher-order statistics as long as the SNR is greater than about 12dB. Nevertheless, methods based on higher-order statistics do not suffer from phase unwrapping errors: if the time delay has been perfectly compensated (i.e. $\tau = 0$), the HOS estimate outperforms the proposed method at low SNR. On the other hand, it can be seen on Figure 7a that the proposed estimator is not significantly influenced by a possible slight residual time delay, whereas the MSE of the estimator based on fourth-order cumulants increases. The practical performance of the latter method is therefore crucially influenced by the accuracy of the prior time synchronization.

Figure 7b represents the MSE when a 3RC signal is transmitted. It can be seen that the HOS estimator yields a very poor performance. This remark confirms that the HOS estimator is not appropriate in case of partial response CPM signals i.e. when $L > 1$.

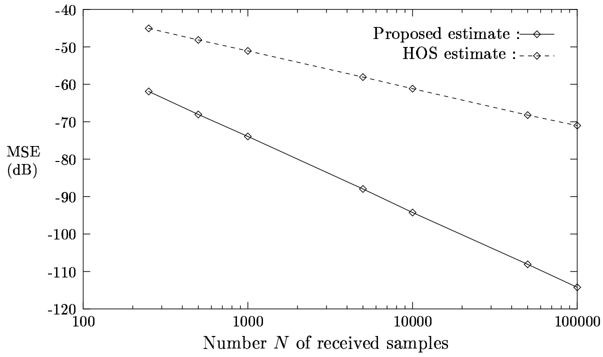


Fig. 8. Mean Square Error as a function of N - $E_b/N_0 = 15$dB, $h = 0.7$, $M = 1$.

Figure 8 illustrates the influence of the number N of received samples on the Mean Square Error (MSE) when E_b/N_0 is set to 15dB. As expected, the gap between the performances of both approaches increases with the number of received samples at least for a sufficiently large SNR.

VII. Conclusion

In this paper, the non data aided estimation of the modulation index of a CPM signal has been addressed. A new estimate has been proposed, and its statistical performance has been studied thoroughly using the functional central limit theorem. In contrast with existing estimators whose convergence rate corresponds to $\frac{1}{\sqrt{N}}$, the rate of convergence of the proposed estimate is equal to $\frac{1}{N}$. It should however be mentioned that the practical implementation of the proposed estimate requires to unwrap the phase of the received signal, a difficult task when the signal to noise ratio is low. This study has been undertaken if the carrier frequency and the symbol period of the CPM are known. Although it has been briefly indicated how to adapt the present approach in this context, the study of the corresponding estimators will be presented in a forthcoming paper.

Appendices

I. Proof of Lemma 1

As $J_n(g_0) = |r_N(g_0)|^2$, $J_N'(g_0)$ can be written as follows:

$$J_N'(g_0) = r_N'(g_0) r_N(g_0)^* + r_N'(g_0)^* r_N(g_0). \tag{37}$$

In order to obtain a more convenient expression of $r_N'(g_0)$, we need to rewrite $\exp[ig_0\psi(nM + m)]$, for each $n = 0 \dots N-1$ and $m = 0 \dots M-1$. Expanding the right-hand side of (18), we remark that $(-1)^n \exp[ig_0\psi(nM+m)]$ can be written as

$$(-1)^n \exp[ig_0\psi(nM+m)] = \lambda(m) + \epsilon(nM+m) \tag{38}$$

where $\lambda(m)$ is the deterministic constant defined by $\lambda(m) = \prod_{j=0}^{L-1} \cos(\pi\phi_{j,m})$ and where $\epsilon(nM+m)$ represents a random variable of zero mean which is defined as follows:

$$\epsilon(nM+m) = \sum_{p=1}^{L} i^p \sum_{\nu_p \in \mathcal{N}_p} \beta(\nu_p, m) \prod_{j \in \nu_p} a_{n-j}, \tag{39}$$

where for each $p \geq 1$, $\mathcal{N}_p$ denotes the set

$$\mathcal{N}_p = \left\{ (j_1, \dots, j_p) \in \{0, \dots, L-1\}^p \middle/ j_1 < j_2 < \dots < j_p \right\},$$

and where coefficients $\beta(\nu_p, m)$ are defined for each $\nu_p \in \mathcal{N}_p$ and for each $m = 0 \dots M-1$ by

$$\beta(\nu_p, m) = \prod_{j \in \nu_p} \sin(\pi\phi_{j,m}) \prod_{j \in \{0,\dots,L-1\} \setminus \nu_p} \cos(\pi\phi_{j,m}). \tag{40}$$

As $(a_n)_{n\in\mathbb{Z}}$ is a zero mean i.i.d. sequence, ϵ is a zero mean periodically correlated sequence of period M. In order to evaluate r_N' at the point g_0, we note that (17) leads to

$$r_N'(g) = \frac{i}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \psi(nM+m) \exp\left[ig\psi(nM+m)\right] (-1)^n e^{-i\pi\frac{m}{M}}. \tag{41}$$

Noticing that $e^{-2i\pi\frac{nM+m}{2M}}$ coincides with $(-1)^n e^{-i\pi\frac{m}{M}}$ and using (6) and (38), one can rewrite (41) as

$$\begin{aligned} r'_N(g_0) &= \frac{i\pi h}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{j=0}^{n-L} a_j\right)(\lambda(m)+\epsilon(nM+m))\,e^{-i\pi\frac{m}{M}} \\ &+ \frac{i\pi h}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{j=0}^{L-1} a_{n-j}\phi_{j,m}\right)\lambda(m)e^{-i\pi\frac{m}{M}} \\ &+ \frac{i\pi h}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{j=0}^{L-1} a_{n-j}\phi_{j,m}\right)\epsilon(nM+m)e^{-i\pi\frac{m}{M}}. \end{aligned} \tag{42}$$

We first study the asymptotic behavior of the third term of the right-hand side of (42). For each $j = 0 \dots L-1$ and for each $m = 0 \dots M-1$, (39) yields

$$\frac{1}{N}\sum_{n=0}^{N-1} a_{n-j}\epsilon(nM+m) = i\beta(\{j\},m) + \sum_{p=1}^{L} i^p \sum_{\nu_p\in\mathcal{N}_p\setminus\{j\}} \beta(\nu_p,m)\frac{1}{N}\sum_{n=0}^{N-1} a_{n-j}\prod_{k\in\nu_p} a_{n-k}. \tag{43}$$

Given a set $\nu_p \in \mathcal{N}_p$, we obtain easily that $E(a_{n-j}\prod_{k\in\nu_p} a_{n-k}) = 0$ as long as ν_p is different from the singleton set $\{j\}$. In this case, the classical central limit theorem implies that the random variable $\frac{1}{N^{1/2}}\sum_{n=0}^{N-1} a_{n-j}\prod_{k\in\nu_p} a_{n-k}$ converges in distribution toward a zero mean Gaussian variable. Thus, (43) can also be written as:

$$\frac{1}{N}\sum_{n=0}^{N-1} a_{n-j}\epsilon(nM+m) = i\beta(\{j\},m) + O_P(\frac{1}{N^{1/2}}). \tag{44}$$

We now consider the second term of the right-hand side of (42). The central limit theorem can be used again to assert that for each $j = 0 \dots L-1$ and for each $m = 0 \dots M-1$:

$$\frac{1}{N}\sum_{n=0}^{N-1} a_{n-j}\lambda(m)e^{-i\pi\frac{m}{M}} = O_P(\frac{1}{N^{1/2}}). \tag{45}$$

Thus the second term of (42) corresponds also to $O_P(\frac{1}{N^{1/2}})$. A more convenient expression of $r'_N(g_0)$ can be obtained by considering sequence $\tilde{\epsilon}(n) = \frac{1}{M}\sum \epsilon(nM+m)e^{-i\pi\frac{m}{M}}$ and by using constant λ which is defined by (20) as the mean w.r.t. m of coefficients $\lambda(m)e^{-i\pi\frac{m}{M}}$. Using the previous remarks, we obtain finally:

$$r'_N(g_0) = \frac{i\pi h\lambda}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{n-L} a_j + \frac{i\pi h}{N}\sum_{n=0}^{N-1}\left(\sum_{j=0}^{n-L} a_j\right)\tilde{\epsilon}(n) - \gamma + O_P(\frac{1}{N^{1/2}}), \tag{46}$$

where

$$\gamma = \frac{\pi h}{M} \sum_{m=0}^{M-1} \sum_{j=0}^{L-1} \beta(\{j\}, m) \phi_{j,m} e^{-i\pi \frac{m}{M}}.$$

It can be shown that the behavior of the second term of the right-hand side of (46) is such as:

$$\frac{i\pi h}{N} \sum_{n=0}^{N-1} \left(\sum_{j=0}^{n-L} a_j \right) \tilde{\epsilon}(n) = O_P(1).$$

Now we plug the final expressions of $r_N(g_0)$ and $r'_N(g_0)$ (given by (19) and (46)) into (37). Noticing that

$$\frac{1}{N} \sum_{n=0}^{N-1} \tilde{\epsilon}(n) = O_P(\frac{1}{N^{1/2}}),$$

we obtain the following result:

$$\begin{aligned} J'_N(g_0) &= 2\Re e \left[\lambda \left(\frac{-i\pi h}{N} \sum_{n=0}^{N-1} \left(\sum_{j=0}^{n-L} a_j \right) \tilde{\epsilon}(n)^* \right) \right. \\ &- \left. \lambda \gamma^* + \left(\frac{1}{N} \sum_{n=0}^{N-1} \tilde{\epsilon}(n) \right) \left(\frac{-i\pi h \lambda^*}{N} \sum_{n=0}^{N-1} \sum_{j=0}^{n-L} a_j \right) \right] + O_P(\frac{1}{N^{1/2}}). \end{aligned} \tag{47}$$

Considering sequence $\overline{\epsilon}(n) = \frac{1}{\rho} i(\lambda \tilde{\epsilon}(n)^* - \lambda^* \tilde{\epsilon}(n))$ and writing λ in the polar form $\lambda = \rho e^{i\theta}$, $J'_N(g_0)$ can finally be written as

$$\begin{aligned} J'_N(g_0) &= -\frac{\pi h \rho}{N} \sum_{n=0}^{N-1} \left(\sum_{j=0}^{n-L} a_j \right) \overline{\epsilon}(n) + \pi h \rho \mu \\ &+ \left(\frac{\pi h \rho}{N} \sum_{n=0}^{N-1} \overline{\epsilon}(n) \right) \left(\frac{1}{N} \sum_{n=0}^{N-1} \sum_{j=0}^{n-L} a_j \right) + O_P(\frac{1}{N^{1/2}}), \end{aligned}$$

where μ is defined by (22).

II. Proof of Lemma 2

We first note that:

$$J''_N(g_0) = r''_N(g_0) r_N(g_0)^* + r''_N(g_0)^* r_N(g_0) + 2 r'_N(g_0) r'_N(g_0)^*. \tag{48}$$

Using (41), $r''_N(g)$ can be written as follows:

$$r''_N(g) = \frac{-1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \psi(nM+m)^2 \exp\left[ig\psi(nM+m)\right] (-1)^n e^{-i\pi \frac{m}{M}}. \tag{49}$$

We now evaluate $r''_N(g_0)/N$ by plugging (6) into (49):

$$\frac{r''_N(g_0)}{N} = \frac{r''_{N,1}(g_0)}{N} + \frac{r''_{N,2}(g_0)}{N} + \frac{r''_{N,3}(g_0)}{N}, \tag{50}$$

where

$$\frac{r''_{N,1}(g_0)}{N} = \frac{-\pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 e^{ig_0 \psi(nM+m)} (-1)^n e^{-i\pi \frac{m}{M}}, \tag{51}$$

$$\frac{r''_{N,2}(g_0)}{N} = \frac{-2\pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{n-L} a_j \right) \left(\sum_{j=0}^{L-1} a_{n-j} \phi_{j,m} \right) e^{ig_0 \psi(nM+m)} (-1)^n e^{-i\pi \frac{m}{M}} \tag{52}$$

$$\frac{r''_{N,3}(g_0)}{N} = \frac{-\pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{L-1} a_{n-j} \phi_{j,m} \right)^2 e^{ig_0 \psi(nM+m)} (-1)^n e^{-i\pi \frac{m}{M}}. \tag{53}$$

$r''_N(g_0)/N$ has actually the same asymptotic behavior than $r''_{N,1}(g_0)/N$. In order to prove this claim, we study the asymptotic behavior of $r''_{N,k}(g_0)/N$ for each $k = 1, 2, 3$. We first consider the third term $r''_{N,3}(g_0)/N$. As $|r''_{N,3}(g_0)/N| \leq \frac{\pi^2 h^2 L^2}{N}$, $r''_{N,3}(g_0)/N$ converges toward zero as N tends to infinity. In order to study the asymptotic behavior of the second term $r''_{N,2}(g_0)/N$, we need to consider the following stochastic process: we define $W_1^{(N)}(t) = \frac{1}{N^{1/2}} \sum_{j=0}^{[Nt]} a_{j-L}$ for each $t \in [0, 1]$. The functional central limit theorem states that $W_1^{(N)}(t)$ converges in distribution toward a Brownian motion, say $W_1(t)$. Therefore, the random variable

$$\frac{1}{N^{3/2}} \sum_{n=0}^{N-1} \left| \sum_{j=0}^{n-L} a_j \right| = \int_0^1 |W_1^{(N)}(t)| dt$$

converges in distribution toward $\int_0^1 |W_1(t)| dt$. Since

$$\left| \frac{r''_{N,2}(g_0)}{N} \right| \leq \frac{2\pi^2 h^2 L^2}{N^2} \sum_{n=0}^{N-1} \left| \sum_{j=0}^{n-L} a_j \right|,$$

we obtain that $r''_{N,2}(g_0)/N = o_P(1)$. We finally study the asymptotic behavior of $r''_{N,1}(g_0)/N$. Plugging (38) into (51) leads to write $r''_{N,1}(g_0)/N$ as a sum of two terms, say $r''_{N,11}(g_0)/N$ and $r''_{N,12}(g_0)/N$, which are given by:

$$\frac{r''_{N,11}(g_0)}{N} = \frac{-\pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 \epsilon(nM+m) e^{-i\pi \frac{m}{M}}$$

$$\frac{r''_{N,12}(g_0)}{N} = \frac{-\pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 \lambda(m) e^{-i\pi \frac{m}{M}}$$

We now show that $r''_{N,11}(g_0)/N$ converges toward zero in probability as N tends to infinity. To that end, we use the functional central limit theorem in order to prove that $r''_{N,11}(g_0)$ converges in distribution toward a certain random variable: this implies indeed that $r''_{N,11}(g_0) = O_P(1)$ and thus, that $r''_{N,11}(g_0)/N$ converges toward zero in probability. First, we note that:

$$\frac{r''_{N,11}(g_0)}{N} = \frac{-\pi^2 h^2}{M} \sum_{m=0}^{M-1} \sum_{p=1}^{L} i^p \sum_{\nu_p \in \mathcal{N}_p} \beta(\nu_p, m) e^{-i\pi \frac{m}{M}} \left(\frac{1}{N^2} \sum_{n=0}^{N-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 \prod_{k \in \nu_p} a_{n-k} \right),$$

and we define for each $p = 1 \dots L$ and for each set $\nu_p \in \mathcal{N}_p$,

$$W_{\nu_p}^{(N)}(t) = \frac{1}{N^{1/2}} \sum_{n=0}^{[Nt]} \prod_{k \in \nu_p} a_{n-k-1}.$$

The functional central limit theorem asserts that the 2-dimensional stochastic process $(W^{(N)}(t), W_{\nu_p}^{(N)}(t))^T$ converges toward the 2-dimensional Brownian motion $[W(t), W_{\nu_p}(t)]^T$. Thus, the following random variable

$$\frac{1}{N^{3/2}} \sum_{n=0}^{N-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 \prod_{k \in \nu_p} a_{n-k} = \int_0^1 \left(W^{(N)}(t) \right)^2 dW_{\nu_p}^{(N)}(t)$$

converges in distribution toward $\int_0^1 W(t)^2 dW_{\nu_p}(t)$. Therefore, $r''_{N,11}(g_0)/N$ converges toward zero in probability as N tends to infinity. Therefore, $r''_N(g_0)$ reduces to

$$\frac{r''_N(g_0)}{N} = \frac{-\lambda \pi^2 h^2}{N^2 M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(\sum_{j=0}^{n-L} a_j \right)^2 + o_P(1). \tag{54}$$

Note that $r_N(g_0)$ can be written in the following form:

$$r_N(g_0) = \lambda + o_P(1), \tag{55}$$

and that (46) can be rewritten as:

$$\frac{r'_N(g_0)}{N^{1/2}} = \frac{i\pi h \lambda}{N^{3/2}} \sum_{n=0}^{N-1} \sum_{j=0}^{n-L} a_j + o_P(1). \tag{56}$$

Plugging (54), (55) and (56) into (48) we obtain the final result:

$$\frac{J''_N(g_0)}{N} = 2\rho^2 (\pi h)^2 \left(\frac{1}{N^2} \sum_{n=0}^{N-1} (\sum_{j=0}^{n-L} a_j)^2 - \frac{1}{N^3} (\sum_{n=0}^{N-1} \sum_{j=0}^{n-L} a_j)^2 \right) + o_P(1).$$

References

[1] P. Billingsley, "Convergence of Probability Measures", John Wiley, 1968.

[2] P.J. Brockwell, R.A. Davies, "Time Series: Theory and Methods", Springer Series in Statistics, 1991.

[3] P.A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulse", *IEEE Trans. on Communications*, vol. COM-34, no. 2, pp. 150-160, February 1986.

[4] A. Napolitano, C. Spooner, "Cyclic spectral analysis of continuous-phase modulated signals", *IEEE Trans. on Signal Processing*, vol. 49, no. 1, pp. 30-44, January 2001.

[5] J.R. Fonollosa, J.A.R. Fonollosa, "Estimation of the modulation index of CPM signals using higher-order statistics", in Proc. ICASSP-93, pp. 268-271, vol. 4, Minneapolis, 1993.

[6] J.R. Fonollosa, C.L. Nikias, "Analysis of CPM signals using higher-order statistics", MILCOM '93, pp. 663-667, vol. 2, Boston, 1993.

[7] W. Rudin, "Real and Complex Analysis", Mc Graw-Hill, New-York, 1970.

[8] U. Mengali, A. d'Andrea, "Synchronization Techniques for Digital Receivers", Plenum Press, 1997.

[9] J. Riba, G. Vasquez, "Parameter estimation of binary CPM signals", in Proc. ICASSP-01, pp. 2561-2564, Salt-Lake City, 2001.

[10] G.K. Kaleh, R.Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels", *IEEE Trans. on Communication*, vol. 42, pp. 2406-2413, July 1994.

[11] H. Cirpan, M.K. Tsatsanis, "Blind receiver for non linearly modulated signals in multipath", in Proc. SPAWC, pp. 149-152, Paris, 1997.

[12] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Royal Statistical Soc., Ser. B*, vol. 39, no.1, pp. 1-38, 1977.

[13] T.K. Moon, "The Expectation Maximization algorithm", *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47-60, November 1996.