

Bayesian and Hybrid Cramér–Rao Bounds for the Carrier Recovery Under Dynamic Phase Uncertain Channels

Jianxiao Yang, Benoit Geller, Stéphanie Bay

▶ **To cite this version:**

Jianxiao Yang, Benoit Geller, Stéphanie Bay. Bayesian and Hybrid Cramér–Rao Bounds for the Carrier Recovery Under Dynamic Phase Uncertain Channels. IEEE Transactions on Signal Processing, 2011, pp.667-680. 10.1109/TSP.2010.2081981 . hal-01180818

HAL Id: hal-01180818
https://hal.science/hal-01180818

Submitted on 4 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Bayesian and Hybrid Cramér-Rao Bounds for the Carrier Recovery under Dynamic Phase Uncertain Channels

Jianxiao Yang, Benoît Geller, and Stephanie Bay[1]

***Abstract*—In this paper, we study Bayesian and hybrid Cramér-Rao bounds (BCRB and HCRB) for the code-aided (CA), the data-aided (DA) and the non-data-aided (NDA) dynamical phase estimation of QAM modulated signals. We address the bounds derivation for both the off-line scenario, for which the whole observation frame is used, and the on-line which only takes into account the current and the previous observations. For the CA scenario we show that the computation of the Bayesian information matrix (BIM) and of the hybrid information matrix (HIM) is NP hard. We then resort to the belief-propagation (BP) algorithm or to the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm to obtain some approximate values. Moreover, in order to avoid the calculus of the inverse of the BIM and of the HIM, we present some closed form expressions for the various CRBs, which greatly reduces the computation complexity. Finally, some simulations allow us to compare the possible improvements enabled by the off-line and the CA scenarios.**

***Index Terms*—Bayesian Cramér-Rao Bound (BCRB), Code-Aided (CA) Bound, Data-Aided (DA) Bound, Dynamical Phase Estimation, Hybrid Cramér-Rao Bound (HCRB), Non-Data-Aided (NDA), On-line, Off-line**

I. Introduction

As well-known, optimal estimators cannot always be built in practical implementations. Assessing the achievable estimation performance may be difficult, and one often has to resort to simulations and then compare the performance to some lower bounds corresponding to the optimum performance. Lower bounds give an indication of the performance limitations, and consequently, they can also be used to determine whether some imposed performance requirements are realistic or not. Although there exists many lower bounds, the Cramér-Rao bounds (CRB) are the most commonly used [1]-[3] because they achieve a good

[1] This work was partially funded by the ANR LURGA program and was presented in part at ICC 2009.
Jianxiao Yang and Benoît Geller are with UEI ENSTA, ParisTech, 32 Boulevard Victor, 75015 Paris, France (e-mail: yang@ensta.fr and geller@ensta.fr);
Stephanie Bay is with SATIE, ENS Cachan, CNRS, UniverSud, 61 Avenue du Président Wilson, 94230 Cachan, France (e-mail: bay@satie.ens-cachan.fr).

accuracy at the price of a reasonable computation difficulty.

There are three ways to use data in a telecommunication system [4]: data aided (DA), code aided (CA) and non data aided (NDA) estimations. Earlier attempts of signal synchronization in the low-SNR regime focused either on the so-called data-aided (DA) or non-data-aided (NDA) synchronization mode [5], [6]. On the one hand, DA parameter estimation techniques rely on the presence of pilot symbols in the data frame and may lead to unacceptable losses in terms of power and spectral efficiency. On the other hand, NDA synchronization algorithms drop some statistical information about the transmitted data and may lead to very poor results at low SNR. As a consequence, it was recognized [7],[8] that the only way to achieve a good performance compromise both in terms of acquisition time and steady-state accuracy (the so-called jitter variance) is to take advantage of the coding gain not only for data detection, but also for synchronization as well. This led to the notion of code aided (CA) synchronization, *i.e.*, explicitly using the channel code structure and properties to perform a satisfying synchronization without any known pilot.

Many works concern the Cramér-Rao bounds for the carrier phase and frequency estimation. Most of them are related to constant (i.e., non-dynamic) carrier phase and frequency. For instance the CRB for phase and /or frequency estimation with known data has been derived in [9]-[11]. Some phase and frequency CRBs for DA and NDA PSK or QAM signals have been derived in [12]-[18]. In particular some analytical expressions of those CRBs at low SNR have been derived in [19],[20]. For PSK signals, the CRBs for DA and NDA estimators were computed in [21],[22] for both the case of phase estimation and the case of joint phase and frequency estimation. Still for the static carrier phase, the CA standard CRB (SCRB) has been derived in [23]-[25] using the first derivative of the log-likelihood function expressed in terms of the marginal a posteriori probabilities (APPs) of the coded symbols. One of the most important contributions of these papers is the application of the BCJR algorithm [37] to the APP computation which was first applied to calculate the APPs of the convolutional code (CC) aided SCRB [23] and was then extended to the evaluation of the APPs of turbo code (TC) aided SCRB in [24]. Moreover, the CA CRBs for joint static parameters estimation problem (carrier phase, carrier frequency and timing estimation) was achieved with the use of the BCJR algorithm in [25].

However, in modern high rate communication systems, one cannot rely on a constant phase model and must take into account time-varying phase noise due to the oscillator instabilities [26]-[33]. To quantify the resulting performance degradation, [26] considered the data aided CRB for the frequency offset estimation with a phase noise variance. [31] has derived a Bayesian CRB (BCRB) for the NDA BPSK signal with dynamical phase offset. When a deterministic parameter (such as a linear drift) is jointly to be considered, the hybrid CRB (HCRB) is relevant; the HCRB was applied to the case of the dynamical phase estimation as briefly sketched in [32] for NDA BPSK signal and in [33] for coded QAM signals. The goal and the contribution of this paper is to give both the BCRB and the HCRB for the dynamical time-varying phase estimation in the case of QAM modulated signals and for the different scenarios (NDA, CA and DA). We present some closed form expressions for the various CRBs based on the second derivatives of the log-likelihood function; by detailing the derivations, we show that the exact calculation of the HCRB is

theoretically NP hard and cannot be easily calculated (the SCRB is just a special case of the HCRB). The obtained closed form expressions give a clearer insight of the corresponding information matrix elements and this enables us to justify some simplifications. Moreover, we display how the different bounds for the deterministic and stochastic parameters of interest behave in the different scenarios (off-line / on-line, DA / CA / NDA) and we give asymptotic results. Note that similarly to [31],[32] which provide a satisfying benchmark for the non coded BPSK phase estimation algorithm [34], these derived CRBs provide very good benchmarks for the more difficult QAM phase estimation algorithms both in NDA [35] and CA [36] scenarios.

This paper is thus organized as follows. In section II, we recall the various kinds of Cramér-Rao bounds. After describing the system model in section III, we derive the Bayesian and the hybrid Cramér-Rao bounds for off-line estimations for the DA, CA and NDA scenarios in section IV. The on-line bound is derived in section V and the different results are illustrated and interpreted in the final section. [2]

II. CRAMÉR-RAO BOUNDS REVIEW

Different kinds of CRBs can be considered. In the following, we briefly describe the links between on one side the HCRB, and on the other side the standard CRB and the BCRB.

We consider the most general case including both deterministic and random parameters for hybrid estimation. Denote this parameter vector as $\mathbf{u}=\left(\mathbf{u}_r^T,\mathbf{u}_d^T\right)^T$, where $\mathbf{u}_d$ is assumed to be a $(n-m)\times 1$ deterministic vector and $\mathbf{u}_r$ is assumed to be a $m\times 1$ random vector with an a priori probability density function (pdf) $p(\mathbf{u}_r)$. The true value of $\mathbf{u}_d$ is denoted $\tilde{\mathbf{u}}_d$. We consider $\hat{\mathbf{u}}(\mathbf{y})$ as an estimator of $\mathbf{u}$ where $\mathbf{y}$ is the observation vector. The HCRB satisfies the following inequality on the MSE

$$E_{\mathbf{y},\mathbf{u}_r|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[\left(\hat{\mathbf{u}}(\mathbf{y})-\mathbf{u}\right)\left(\hat{\mathbf{u}}(\mathbf{y})-\mathbf{u}\right)^T_{|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\right]\geq \mathbf{H}^{-1}(\tilde{\mathbf{u}}_d), \tag{1}$$

where $\mathbf{H}(\tilde{\mathbf{u}}_d)$ is the so-called hybrid information matrix (HIM) and is defined as

$$\mathbf{H}(\tilde{\mathbf{u}}_d)=E_{\mathbf{y},\mathbf{u}_r|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[-\Delta_{\mathbf{u}}^{\mathbf{u}}\log p(\mathbf{y},\mathbf{u}_r\,|\,\mathbf{u}_d)_{|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\right]. \tag{2}$$

It is shown in [32] that inequality (1) is still respected when the deterministic and the random parts of the parameter vector are dependent. By expanding the log-likelihood as $\log p(\mathbf{y},\mathbf{u}_r\,|\,\mathbf{u}_d)=\ln p(\mathbf{y}\,|\,\mathbf{u}_r,\mathbf{u}_d)+\ln p(\mathbf{u}_r\,|\,\mathbf{u}_d)$, the HIM can be rewritten as

[2] The notational convention adopted is as follows: italic indicates a scalar quantity, as in a ; boldface indicates a vector quantity, as in $\mathbf{a}$ and capital boldface indicates a matrix quantity as in $\mathbf{A}$. The $(m,n)^{th}$ entry of matrix $\mathbf{A}$ is denoted as $[A]_{m,n}$. The transpose matrix of $\mathbf{A}$ is indicated by a superscript $\mathbf{A}^T$, and $|\mathbf{A}|$ is the determinant of $\mathbf{A}$. $\mathbf{a}_m^n$ represents the vector $[a_m,\cdots,a_n]^T$, where m and n are positive integers ($m<n$). $\mathrm{Re}\{a\}$ and $\mathrm{Im}\{a\}$ are respectively the real and imaginary parts of a. $E_{xy}[\,]$ denotes the expectation over x and y. $\nabla_{\mathbf{u}}$ and $\Delta_{\mathbf{u}}^{\mathbf{v}}$ represent the first and second order derivative operators. Finally, $\tilde{a}$ is designated as the value taken by variable a.

$$\mathbf{H}(\tilde{\mathbf{u}}_d) = E_{\mathbf{u}_r|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[\mathbf{F}(\tilde{\mathbf{u}}_d,\mathbf{u}_r)\right] + E_{\mathbf{u}_r|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[-\Delta_{\mathbf{u}}^{\mathbf{u}} \ln p(\mathbf{u}_r \mid \mathbf{u}_d)_{\mathbf{u}_d=\tilde{\mathbf{u}}_d}\right], \tag{3}$$

where $\mathbf{F}(\tilde{\mathbf{u}}_d,\mathbf{u}_r)$ is the Fisher information matrix (FIM) [2] defined by

$$\mathbf{F}(\tilde{\mathbf{u}}_d,\mathbf{u}_r) = E_{\mathbf{y}|\mathbf{u}_r,\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[-\Delta_{\mathbf{u}}^{\mathbf{u}} \ln p(\mathbf{y} \mid \mathbf{u}_d,\mathbf{u}_r)_{|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\right]. \tag{4}$$

In (3), the Fisher information matrix $\mathbf{F}(\tilde{\mathbf{u}}_d,\mathbf{u}_r)$ gives the contribution of the observations whereas the second term on the right side corresponds to the contribution of the a priori distribution. It is then straightforward to re-obtain the standard and the Bayesian CRBs as particular cases. If $\mathbf{u}=\mathbf{u}_d$, then $\mathbf{H}$ reduces to

$$\mathbf{H}(\tilde{\mathbf{u}}_d) = \mathbf{F}(\tilde{\mathbf{u}}_d) = E_{\mathbf{y}|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\left[-\Delta_{\mathbf{u}_d}^{\mathbf{u}_d} \ln p(\mathbf{y} \mid \mathbf{u}_d)_{|\mathbf{u}_d=\tilde{\mathbf{u}}_d}\right], \tag{5}$$

and the inverse of (5) is just the standard CRB [1].

If $\mathbf{u}=\mathbf{u}_r$, we have that

$$\mathbf{H} = E_{\mathbf{u}_r}\left[\mathbf{F}(\mathbf{u}_r)\right] + E_{\mathbf{u}_r}\left[-\Delta_{\mathbf{u}_r}^{\mathbf{u}_r} \ln p(\mathbf{u}_r)\right], \tag{6}$$

where

$$\mathbf{F}(\mathbf{u}_r) = E_{\mathbf{y}|\mathbf{u}_r}\left[-\Delta_{\mathbf{u}_r}^{\mathbf{u}_r} \ln p(\mathbf{y} \mid \mathbf{u}_r)\right] \tag{7}$$

and the inverse of $\mathbf{H}$ in (6) is the Bayesian CRB [2].

Because (5)-(6) can be considered as special cases of the general definition given in (3), in section IV, we first give the detailed derivation of the HCRB for the QAM dynamical phase estimation and it is then easy to obtain the BCRB for our problem.

III. System Model

We consider the transmission over an additive white Gaussian noise (AWGN) channel of a modulated sequence $\mathbf{s}=[s_1,\cdots,s_L]^T$, where the symbols s_l belong to a constellation set $\mathbf{S}_M=\{\tilde{s}_1,\cdots,\tilde{s}_M\}$ (M-QAM, M-PSK or M-APSK), affected by some carrier phase offsets stacked in a vector $\boldsymbol{\theta}=[\theta_1,\cdots,\theta_L]^T$. Assuming that the timing recovery is perfect and that there is no inter-symbol interference (ISI), the sampled baseband signal $\mathbf{y}=[y_1,\cdots,y_L]^T$ is written as

$$y_l = s_l e^{j\theta_l} + n_l = (a_l + jb_l)e^{j\theta_l} + n_l, \tag{8}$$

where s_l, θ_l and n_l are respectively the l^{th} transmitted complex symbol ($s_l = a_l + jb_l$ with $E\{s_l\}=0$ and $E\left\{|s_l|^2\right\} = \frac{1}{M}\sum_{s_l\in\mathbf{S}_M}|s_l|^2 = \sigma_s^2$ is the transmitted signal power), the residual phase distortion that must be eliminated and the zero mean circular Gaussian noise with known variance σ_n^2.

For the data-aided (DA) system, the transmitted symbols are known. As the observations and the transmitted symbols are independent and the additive noise is Gaussian distributed, the conditional probability based on the known phase $\boldsymbol{\theta}$ is

$$p(\mathbf{y}\,|\,\boldsymbol{\theta}) = \prod_{l=1}^{L} p(y_l\,|\,\theta_l) = \left(\frac{1}{\pi\sigma_n^2}\right)^L \prod_{l=1}^{L} \exp\left\{-\frac{|s_l|^2+|y_l|^2}{\sigma_n^2}\right\} \exp\left\{2\frac{\mathrm{Re}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}\right\}. \tag{9}$$

For the non-data-aided (NDA) system, the transmitted symbols are independent and identically distributed (*i.i.d.*). Hence the conditional probability based on the known phase $\boldsymbol{\theta}$ is

$$p(\mathbf{y}\,|\,\boldsymbol{\theta}) = \prod_{l=1}^{L} p(y_l\,|\,\theta_l) = \left(\frac{1}{\pi\sigma_n^2}\right)^L \prod_{l=1}^{L} \sum_{s_l\in\mathbf{S}_M} \frac{1}{M} \exp\left\{-\frac{|s_l|^2+|y_l|^2}{\sigma_n^2}\right\} \exp\left\{\frac{2\,\mathrm{Re}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}\right\}. \tag{10}$$

For the code-aided (CA) system, the independent and identically distributed (*i.i.d.*) condition between the transmitted symbols does not hold any more, instead the *i.i.d.* condition holds between the codewords. A message of K bits is encoded into a N bit codeword $\mathbf{c} = (c_1, c_2, \cdots, c_N) \in \left\{\tilde{\mathbf{c}}_v \,|\, v = 1, \cdots, 2^K\right\}$ which is then in practice further mapped as a symbol vector $\tilde{\mathbf{s}}(\tilde{\mathbf{c}}_v) = \left(\tilde{s}_1(\tilde{\mathbf{c}}_v), \cdots, \tilde{s}_L(\tilde{\mathbf{c}}_v)\right)$. Hence, the conditional probability based on the known phase vector $\boldsymbol{\theta}$ is

$$\begin{aligned} p(\mathbf{y}\,|\,\boldsymbol{\theta}) &= \sum_{v=1}^{2^K} p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v, \boldsymbol{\theta})\, p(\mathbf{c}=\tilde{\mathbf{c}}_v) = \sum_{v=1}^{2^K} p\left(\mathbf{y}\,|\,\mathbf{s}=\tilde{\mathbf{s}}(\tilde{\mathbf{c}}_v), \boldsymbol{\theta}\right) p\left(\mathbf{s}=\tilde{\mathbf{s}}(\tilde{\mathbf{c}}_v)\right) \\ &= \left(\frac{1}{\pi\sigma_n^2}\right)^L \sum_{v=1}^{2^K} \frac{1}{2^K} \prod_{l=1}^{L} \exp\left\{-\frac{\left|\tilde{s}_l(\tilde{\mathbf{c}}_v)\right|^2+|y_l|^2}{\sigma_n^2}\right\} \exp\left\{2\frac{\mathrm{Re}\left\{y_l \tilde{s}_l^*(\tilde{\mathbf{c}}_v) e^{-j\theta_l}\right\}}{\sigma_n^2}\right\}. \end{aligned} \tag{11}$$

We now consider the phase model. In practice, there is a constant frequency shift between the transmitter's clock and the receiver's clock. The corresponding phase distortion is linear. Furthermore, clocks are never perfect and oscillators suffer from jitters. This results in a Brownian phase model with a linear drift

$$\theta_l = \theta_{l-1} + \xi + w_l, \tag{12}$$

where θ_l is the unknown phase offset at time l, ξ is the unknown constant frequency offset (linear drift), w_l is a white Gaussian noise with zero mean and variance σ_w^2. This model is commonly used [26]-[34] in order to describe the behavior of practical oscillators for which the frequency is randomly perturbed. The corresponding conditional probability can be expressed as

$$p(\theta_l\,|\,\theta_{l-1}, \xi) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{(\theta_l - \theta_{l-1} - \xi)^2}{2\sigma_w^2}\right\}. \tag{13}$$

Note that from (12), the joint pdf $p(\boldsymbol{\theta}\,|\,\xi)$ can be written as

$$p(\boldsymbol{\theta}\,|\,\xi) = p(\theta_1) \prod_{l=2}^{L} p(\theta_l\,|\,\theta_{l-1}, \xi) \quad . \tag{14}$$

Since the observation y_l in (8) does not depend on the deterministic parameter ξ, we also have $p(y_l\,|\,\theta_l, \xi) = p(y_l\,|\,\theta_l)$.

IV. CRBs FOR THE DYNAMICAL PHASE ESTIMATION

In practical receivers, phase estimation can actually be considered following two main scenarios: off-line and on-line. With off-line synchronization, the carrier phase offset $\boldsymbol{\theta}$ is not estimated until the whole observation frame $\mathbf{y}$ has been received. In the rest of this section, we derive some analytical expressions for bounds corresponding to the off-line carrier phase offsets estimation. In a subsequent paragraph, these expressions will further allow us to find some bounds for the on-line scenario.

The parameters of the phase model (12) include some random parameters $\boldsymbol{\theta}=[\theta_1,\cdots,\theta_L]^T$ (*i.e.* the dynamical phase) and a deterministic parameter ξ (*i.e.* the scalar linear drift), so that the parameter vector can be written as

$$\mathbf{u}=\begin{bmatrix}\mathbf{u}_r\\ \mathbf{u}_d\end{bmatrix}=\begin{bmatrix}\boldsymbol{\theta}\\ \xi\end{bmatrix}. \tag{15}$$

Equation (3) thus becomes

$$\begin{aligned}\mathbf{H}\left(\tilde{\xi}\right)&=E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left(E_{\mathbf{y}|\boldsymbol{\theta},\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\xi\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\end{pmatrix}\right)+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\xi}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\end{pmatrix}\\ &=E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)\right]+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\xi}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\end{pmatrix}\end{aligned}. \tag{16}$$

where $\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)=E_{\mathbf{y}|\boldsymbol{\theta},\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\xi\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\end{pmatrix}$.

We then decompose the hybrid information matrix (HIM) $\mathbf{H}$ into a block matrix that will be useful in the sequel

$$\mathbf{H}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{12}^T & \mathbf{H}_{22}\end{bmatrix}, \tag{17}$$

where

$$\begin{cases}\mathbf{H}_{11}=E_{\mathbf{y},\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right]+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right]\\ \mathbf{H}_{12}=E_{\mathbf{y},\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right]+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right]=\mathbf{H}_{21}^T\\ \mathbf{H}_{22}=E_{\mathbf{y},\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right]+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}\right].\end{cases} \tag{18}$$

From (8)-(11), one can see that the log likelihood function $\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\xi\right)$ does not depend on ξ; consequently, the partial derivatives $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}}\log p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}$ and $\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}}\log p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}}$ are both equal to zero. The hybrid block matrix can thus be written as:

$$\begin{aligned}\mathbf{H}\left(\tilde{\xi}\right)&=E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\begin{pmatrix}E_{\mathbf{y}|\boldsymbol{\theta},\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\mathbf{y}\,|\,\boldsymbol{\theta},\xi\right)_{|\xi=\tilde{\xi}}\right] & \mathbf{0}_{L\times 1}\\ \mathbf{0}_{1\times L} & 0\end{pmatrix}+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\xi}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\end{pmatrix}\\ &=E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[\mathbf{F}\left(\boldsymbol{\theta}\right)\right]+E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\begin{pmatrix}-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\ln p\left(\boldsymbol{\theta}\,|\,\boldsymbol{\xi}\right)_{|\xi=\tilde{\xi}} & -\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\\ \left(-\Delta_{\boldsymbol{\theta}}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\right)^T & -\Delta_{\xi}^{\xi}\ln p\left(\boldsymbol{\theta}\,|\,\xi\right)_{|\xi=\tilde{\xi}}\end{pmatrix}\end{aligned}, \tag{19}$$

where $\mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} E_{\mathbf{y}|\boldsymbol{\theta},\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)_{|\xi=\tilde{\xi}}\right] & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & 0 \end{pmatrix}$ and

$$\begin{cases} \mathbf{H}_{11} = E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[E_{\mathbf{y}|\boldsymbol{\theta},\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)_{|\xi=\tilde{\xi}}\right]\right] + E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|\boldsymbol{\xi})_{|\xi=\tilde{\xi}}\right] \\ \mathbf{H}_{12} = E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\xi}} \ln p(\boldsymbol{\theta}|\boldsymbol{\xi})_{|\xi=\tilde{\xi}}\right] = \mathbf{H}_{21}^T \\ \mathbf{H}_{22} = E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\Delta_{\boldsymbol{\xi}}^{\boldsymbol{\xi}} \ln p(\boldsymbol{\theta}|\boldsymbol{\xi})_{|\xi=\tilde{\xi}}\right]. \end{cases} . \tag{20}$$

First, the DA and NDA scenarios are briefly reviewed in sections IV.A and IV.B, as a simplified version has already been presented in [33]. Then the more difficult to tackle CA scenario is detailed in section IV.C.

A. Computation of $E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)\right]$ for the DA scenario

The evaluation of the term $E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)\right]$ requires the computation of the FIM $\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)$, which in turn requires the evaluation of the Hessian of the log-likelihood function $\ln p(\mathbf{y}|\boldsymbol{\theta},\xi)$. For both the DA and NDA scenarios, according to the observation model defined previously, after a marginalization on the independent symbols, and then using both the independence of the transmitted symbols and the whiteness of the noise, one finds that:

$$\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta},\xi) = \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\left(\sum_{l=1}^{L}\left(\ln \sum_{s_l} p(y_l|s_l,\theta_l,\xi)\, p(s_l)\right)\right) = \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}\left(\sum_{l=1}^{L} \ln p(y_l|\theta_l)\right). \tag{21}$$

One can realize that each term of the sum (21) is a matrix with only one non-zero element at most, namely,

$$\left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)\right]_{l,l} = \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(y_l|\theta_l). \tag{22}$$

As a direct consequence, the Hessian $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)$ is a diagonal matrix with the l^{th} diagonal element given by (22). Moreover, because of the circularity of the observation noise, the expectation of (22) with respect to $p(y_l|\theta_l)$ does not depend on θ_l. Discarding the last zero line and column, one then obtains

$$E_{\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[\mathbf{F}(\boldsymbol{\theta})\right] = J_D \mathbf{I}_L, \tag{23}$$

where $\mathbf{I}_L$ is the $L\times L$ identity matrix and J_D is defined as follows

$$J_D = E_{\mathbf{y},\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)}{\partial \theta_l^2}\right]. \tag{24}$$

Starting from (9) with the DA scenario where $p(y_l|\theta_l,\xi) = p(y_l|\theta_l,\xi,s_l)$ and taking the first and the second derivatives, one easily obtains that

$$\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta},\xi)}{\partial \theta_l} = \frac{\partial \ln p(y_l|\theta_l,\xi,s_l)}{\partial \theta_l} = \frac{2\,\mathrm{Im}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2} \tag{25}$$

$$\text{and } \frac{\partial^2 \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l^2} = \frac{\partial^2 \ln p(y_l\,|\,\theta_l,\xi,s_l)}{\partial \theta_l^2} = -\frac{2\mathrm{Re}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}. \tag{26}$$

So in the DA case, (24) becomes

$$J_D^{(\mathrm{DA})} = E_{\mathbf{y},\boldsymbol{\theta}}\left[-\frac{\partial^2 \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l^2}\right] = \frac{2\sigma_s^2}{\sigma_n^2} = 2\cdot \mathrm{SNR}\,, \tag{27}$$

where the signal-to-noise ratio is defined as $\mathrm{SNR} = \sigma_s^2/\sigma_n^2$.

B. Computation of $E_{\boldsymbol{\theta}|\tilde{\xi}}\left[\mathbf{F}\left(\tilde{\xi},\boldsymbol{\theta}\right)\right]$ *for the NDA scenario*

We now turn to the NDA scenario. With an appropriate marginalization on the transmitted symbols, we obtain by deriving (10)

$$\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta})}{\partial \theta_l} = \frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l} = \frac{\partial \ln p(y_l\,|\,\theta_l,\xi)}{\partial \theta_l} = \frac{\partial \ln \sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}. \tag{28}$$

As the derivation is a linear operator, we find that

$$\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l} = \frac{\dfrac{\partial \sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}}{\sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)} = \frac{\sum_{s_l\in\mathbf{S}_M} \dfrac{\partial p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}}{\sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)} = \sum_{s_l\in\mathbf{S}_M} \frac{\dfrac{\partial p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}}{p(y_l\,|\,\theta_l,\xi)} \tag{29}$$

Moreover as $\dfrac{\partial \ln p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l} = \dfrac{1}{p(y_l\,|\,s_l,\theta_l,\xi)}\dfrac{\partial p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}$, using (25) we have $\dfrac{\partial p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l} = p(y_l\,|\,s_l,\theta_l,\xi)\dfrac{2\mathrm{Im}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}$.

Consequently we obtain

$$\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l} = \sum_{s_l\in\mathbf{S}_M} \Pr(s_l\,|\,y_l,\theta_l,\xi)\frac{2\mathrm{Im}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}, \tag{30}$$

where $\Pr(s_l\,|\,y_l,\theta_l,\xi) = \dfrac{p(y_l\,|\,s_l,\theta_l,\xi)\,p(s_l)}{p(y_l\,|\,\theta_l,\xi)}$ is the marginalized a posteriori probability (APP) of s_l based on the observation y_l with known phase.

Taking the second derivative of $\ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)$, using (25), (26), (30) and following a similar calculus, we further find that

$$\begin{aligned}
\frac{\partial^2 \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_l^2} &= \frac{\partial^2 \ln p(y_l\,|\,\theta_l,\xi)}{\partial \theta_l^2} \\
&= \frac{\sum_{s_l\in\mathbf{S}_M} \dfrac{\partial^2 p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l^2}}{\sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)} - \left(\frac{\sum_{s_l\in\mathbf{S}_M} \dfrac{\partial p(y_l\,|\,s_l,\theta_l,\xi)}{\partial \theta_l}}{\sum_{s_l\in\mathbf{S}_M} p(y_l\,|\,s_l,\theta_l,\xi)}\right)^2 \\
&= \sum_{s_l\in\mathbf{S}_M} \Pr(s_l\,|\,y_l,\theta_l,\xi)\left(\left(\frac{2\mathrm{Im}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}\right)^2 - \frac{2\mathrm{Re}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}\right) - \left(\sum_{s_l\in\mathbf{S}_M} \Pr(s_l\,|\,y_l,\theta_l,\xi)\frac{2\mathrm{Im}\left\{y_l s_l^* e^{-j\theta_l}\right\}}{\sigma_n^2}\right)^2.
\end{aligned} \tag{31}$$

In general, the expectation of (31) with respect to $\Pr(s_l \mid y_l, \theta_l, \xi)$ does not have any simple analytical solution. Hence, in practice, we have to evaluate

$$J_D^{(\mathrm{NDA})} = E_{\mathbf{y},\boldsymbol{\theta}|\xi=\tilde{\xi}}\left[-\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_l^2}\right] \tag{32}$$

by numerical integration.

C. *Computation of* $E_{\boldsymbol{\theta}|\tilde{\xi}}\left[\mathbf{F}(\tilde{\xi},\boldsymbol{\theta})\right]$ *for the CA scenario*

For the CA scenario, because of the code structure, the independence condition between symbols does not hold anymore. Using (11), one has that

$$\ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi) = \ln\left(\sum_{v=1}^{2^K} p(\mathbf{y}\mid\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)\, p(\mathbf{c}=\tilde{\mathbf{c}}_v)\right) = \ln\left(\sum_{v=1}^{2^K}\left(\prod_{l=1}^{L} p\left(y_l \mid s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right) p(\mathbf{c}=\tilde{\mathbf{c}}_v)\right). \tag{33}$$

From (70)-(72) in the Appendix, we show that the first and the second derivatives can respectively be expressed as

$$\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m} = \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\mid\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}, \tag{34}$$

for $m \neq n$

$$\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m\partial\theta_n} = \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\mid\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}\frac{\partial \ln p\left(y_n \mid s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial\theta_n} - \frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n}, \tag{35}$$

and for $m = n$

$$\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m^2} = \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\mid\mathbf{y},\boldsymbol{\theta},\xi)\left(\frac{\partial^2 \ln p\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m^2} + \left(\frac{\partial \ln p\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}\right)^2\right) - \left(\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\right)^2. \tag{36}$$

All these derivatives are too hard to use in practice because they all involve an exponential number of sums. We thus propose the following method to find a good approximation of these derivatives. From (34), the first derivative can be further expressed as follows

$$\begin{aligned}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m} &= \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\mid\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m} = \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v)\mid\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m} \\ &= \sum_{v=1}^{2^K}\sum_{i=1}^{M}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v)\mid\mathbf{y},\boldsymbol{\theta},\xi)\,\delta_{\mathbf{c}=\tilde{\mathbf{c}}_v}(s_m-\tilde{s}_i)\frac{\partial \ln p\left(y_m \mid s_m=\tilde{s}_i,\theta_m,\xi\right)}{\partial\theta_m},\end{aligned} \tag{37}$$

where

$$\delta_{\mathbf{c}=\tilde{\mathbf{c}}_v}(s_m-\tilde{s}_i) = \begin{cases}1, & \text{if } s_m=\tilde{s}_i \\ 0, & \text{otherwise.}\end{cases} \tag{38}$$

Thus

$$\begin{aligned}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m} &= \sum_{i=1}^{M}\sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{\partial \ln p\left(y_m\mid s_m=\tilde{s}_i,\theta_m,\xi\right)}{\partial\theta_m}\\ &= \sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_i^{*}e^{-j\theta_m}\right\}}{\sigma_n^2},\end{aligned} \tag{39}$$

where

$$\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)=\sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)=\sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v)\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\delta_{\mathbf{c}=\tilde{\mathbf{c}}_v}\left(s_m-\tilde{s}_i\right). \tag{40}$$

The term $\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$ in (39) is exactly the a posteriori probability of the code aided scenario.

Similarly to the first derivative (39), the second derivatives (35) for $n\neq m$ can be further expressed as

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n\partial\theta_m} &= \sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v)\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{\partial \ln p\left(y_m\mid s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}\frac{\partial \ln p\left(y_n\mid s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial\theta_n}-\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n}\\ &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v), s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{\partial \ln p\left(y_m\mid s_m=\tilde{s}_{i_1},\theta_m,\xi\right)}{\partial\theta_m}\frac{\partial \ln p\left(y_n\mid s_n=\tilde{s}_{i_2},\theta_n,\xi\right)}{\partial\theta_n}-\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n}\\ &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\Pr\left(s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_{i_1}^{*}e^{-j\theta_m}\right\}}{\sigma_n^2}\frac{2\,\mathrm{Im}\left\{y_n\tilde{s}_{i_2}^{*}e^{-j\theta_n}\right\}}{\sigma_n^2}-\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n},\end{aligned} \tag{41}$$

where

$$\begin{aligned}\Pr\left(s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right) &= \sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v), s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\\ &= \sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v)\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\delta_{\mathbf{c}=\tilde{\mathbf{c}}_v}\left(s_m-\tilde{s}_{i_1}\right)\delta_{\mathbf{c}=\tilde{\mathbf{c}}_v}\left(s_n-\tilde{s}_{i_2}\right).\end{aligned} \tag{42}$$

Like the term $\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$ in (39), the term $\Pr\left(s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$ in (41) is also an a posteriori probability of the code aided scenario and it measures the dependency between the symbols in position n and in position m. Actually the evaluation of $\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n\partial\theta_m}$ in (41) for any couple of positions really burdens the computation complexity. If the codeword has been sufficiently interleaved, the term $\Pr\left(s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$ can be approximately expressed as the product of two independent probabilities, *i.e.* $\Pr\left(s_m=\tilde{s}_{i_1}, s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\approx\Pr\left(s_m=\tilde{s}_{i_1}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\Pr\left(s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$. We can then further simplify (41) as

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n\partial\theta_m} &\approx \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\Pr\left(s_m=\tilde{s}_{i_1}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\Pr\left(s_n=\tilde{s}_{i_2}\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_{i_1}^{*}e^{-j\theta_m}\right\}}{\sigma_n^2}\frac{2\,\mathrm{Im}\left\{y_n\tilde{s}_{i_2}^{*}e^{-j\theta_n}\right\}}{\sigma_n^2}-\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n}\\ &= \left(\sum_{i_1=1}^{M}\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_{i_1}^{*}e^{-j\theta_m}\right\}}{\sigma_n^2}\Pr\left(s_m=\tilde{s}_{i_1}\mid\mathbf{r},\boldsymbol{\theta},\xi\right)\right)\left(\sum_{i_2=0}^{M}\frac{2\,\mathrm{Im}\left\{y_n\tilde{s}_{i_2}^{*}e^{-j\theta_n}\right\}}{\sigma_n^2}\Pr\left(s_n=\tilde{s}_{i_2}\mid\mathbf{r},\boldsymbol{\theta},\xi\right)\right)-\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n}\\ &= 0,\end{aligned} \tag{43}$$

where the last equality comes from (39).

The second derivative (36) for $m=n$ can be similarly derived as below

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m^2} &= \sum_{i=1}^{M}\sum_{v=1}^{2^K}\Pr\left(\mathbf{c}=\tilde{\mathbf{c}}_v, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v), s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\left(\frac{\partial^2 \ln p(y_m\mid s_m=\tilde{s}_i,\theta_m,\xi)}{\partial\theta_m^2}+\left(\frac{\partial \ln p(y_m\mid s_m=\tilde{s}_i,\theta_m,\xi)}{\partial\theta_m}\right)^2\right)-\left(\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\right)^2 \\ &= \sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\left(\frac{\partial^2 \ln p(y_m\mid s_m=\tilde{s}_i,\theta_m,\xi)}{\partial\theta_m^2}+\left(\frac{\partial \ln p(y_m\mid s_m=\tilde{s}_i,\theta_m,\xi)}{\partial\theta_m}\right)^2\right)-\left(\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_m}\right)^2 \\ &= -\sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Re}\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^2}+\sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\left(\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^2}\right)^2-\left(\sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^2}\right)^2.\end{aligned} \tag{44}$$

We thus obtain the expression of the second derivatives as follows when the code has been sufficiently interleaved

$$\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_n\partial\theta_m}=\begin{cases}0, & n\neq m\\ \sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\left(\frac{4\,\mathrm{Im}^2\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^4}-\frac{2\,\mathrm{Re}\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^2}\right)-\left(\sum_{i=1}^{M}\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)\frac{2\,\mathrm{Im}\left\{y_m\tilde{s}_i^* e^{-j\theta_m}\right\}}{\sigma_n^2}\right)^2, & n=m\end{cases} \tag{45}$$

Actually, the calculation of $\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$ is also NP hard (see (40)) and theoretically requires 2^K sums. In practice we just have to resort to some classical a posteriori probability (APP) decoding algorithms, namely the famous Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [37] or the belief propagation (BP) algorithm [38]-[40], in order to calculate the approximate value of $\Pr\left(s_m=\tilde{s}_i\mid\mathbf{y},\boldsymbol{\theta},\xi\right)$. In the rest of the paper, we assume that the code has been sufficiently interleaved so that (45) holds.Moreover, discarding the last zero line and column, the FIM then has an identity matrix form as

$$E_{\boldsymbol{\theta}}\left[\mathbf{F}(\boldsymbol{\theta})\right]=J_D^{(\mathrm{CA})}\mathbf{I}_L, \tag{46}$$

where $\mathbf{I}_L$ is the $L\times L$ identity matrix and $J_D^{(\mathrm{CA})}=E_{\mathbf{y},\boldsymbol{\theta}\mid\xi=\tilde{\xi}}\left[-\frac{\partial^2 \ln p(\mathbf{y}\mid\boldsymbol{\theta},\xi)}{\partial\theta_l^2}\right]$ can be obtained from (45) with a Monte Carlo simulation.

D. Computation of the HIM

As mentioned in (17), the HIM can be divided into four submatrices; due to the model of Section III, assuming that we have no a priori knowledge on the initial phase, one obtains that the upper left part $\mathbf{H}_{11}$, which corresponds to the random parameters $\boldsymbol{\theta}$, has a particular mathematical structure just like in [31]

$$\mathbf{H}_{11}=b\begin{bmatrix}A+1, & 1 & 0 & \cdots & 0\\ 1 & A & 1 & 0 & \vdots\\ 0 & \ddots & \ddots & \ddots & 0\\ \vdots & 0 & 1 & A & 1\\ 0 & \cdots & 0 & 1 & A+1\end{bmatrix}, \tag{47}$$

where

$$\begin{cases}A=-\sigma_w^2 J_D-2\\ b=-1/\sigma_w^2.\end{cases} \tag{48}$$

Using (14), $\mathbf{H}_{12}$ and $\mathbf{H}_{22}$ in (20) thus become:

$$\mathbf{H}_{12}=\mathbf{H}_{21}^{T}=\left[1/\sigma_w^2,\quad \mathbf{0}_{1\times(L-2)},\quad -1/\sigma_w^2\right]^T, \tag{49}$$

$$\mathbf{H}_{22}=(L-1)/\sigma_w^2. \tag{50}$$

E. The Bayesian Cramér-Rao bounds (BCRBs)

When there is no linear drift, the parameter vector $\mathbf{u}$ contains only random parameters $\boldsymbol{\theta}$, *i.e.* $\mathbf{u}=\mathbf{u}_r=\boldsymbol{\theta}$. In this scenario, the BCRB is the lower bound of the MSE. Moreover, the Bayesian information matrix (BIM) $\mathbf{B}_L$ is equal to the upper left sub-matrix of the HIM, *i.e.*

$$\mathbf{B}_L=\mathbf{H}_{11}. \tag{51}$$

The diagonal element $\left[\mathbf{B}_L^{-1}\right]_{l,l}$ of the inverse of matrix $\mathbf{B}_L$ is the off-line BCRB associated to the estimation of θ_l. Furthermore, the corresponding analytical expression associated with the off-line BCRB has already been obtained in Appendix I. D of [31]

$$\left[\mathbf{H}_{11}^{-1}\right]_{l,l}=\left[\mathbf{B}_L^{-1}\right]_{l,l}=\frac{1}{|\mathbf{B}_L|}\left(\rho_1^2(b+r_1)^2 r_1^{L-3}+\rho_2^2(b+r_2)^2 r_2^{L-3}-b^2\left(r_1^{L-2}r_2^{L-l-1}+r_1^{L-l-1}r_2^{L-2}\right)(A-2)^{-1}\right), \tag{52}$$

where

$$\begin{cases} r_1=1/\sigma_w^2+\left(1-\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}\right)J_D/2, \\ r_2=1/\sigma_w^2+\left(1+\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}\right)J_D/2, \end{cases} \tag{53}$$

$$\text{and}\ \begin{cases} \rho_1=\left(\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}-1-2\left(J_D\sigma_w^2\right)^{-1}\right)\Big/\left(2\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}\right), \\ \rho_2=\left(\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}+1+2\left(J_D\sigma_w^2\right)^{-1}\right)\Big/\left(2\sqrt{1+4\left(J_D\sigma_w^2\right)^{-1}}\right). \end{cases} \tag{54}$$

F. The hybrid Cramér-Rao bounds (HCRBs)

We now inverse the HIM to obtain the analytical expression of the HCRB. In the sequel, we shall need the expression of the elements of the first row of $\mathbf{H}_{11}^{-1}$. We proceed similarly as in [32]. From (47), thanks to the cofactor expression of the matrix inversion formula we have

$$\left[\mathbf{H}_{11}^{-1}\right]_{1,l}=\frac{b^{l-1}}{|\mathbf{H}_{11}|}\left(d_{L-l}+bd_{L-l-1}\right), \tag{55}$$

where d_l is the determinant of the following $l\times l$ matrix D_l

$$D_l = b\begin{bmatrix} A & 1 & 0 & \cdots & 0 \\ 1 & A & 1 & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & 1 & A & 1 \\ 0 & \cdots & 0 & 1 & A \end{bmatrix}, \tag{56}$$

and d_l satisfies the following recursive equation

$$d_l = Abd_{l-1} - b^2 d_{l-2}, \quad \text{with } d_0 = 1 \text{ and } d_1 = bA. \tag{57}$$

Similarly to Appendix I.A of [31], d_l can thus be rewritten as $d_l = \rho_1 r_1^l + \rho_2 r_2^l$, where r_1, r_2, ρ_1 and ρ_2 were given in (53) and (54). We then obtain from (55)

$$\left[\mathbf{H}_{11}^{-1}\right]_{1,l} = \frac{b^{l-1}}{|\mathbf{H}_{11}|}\left(\rho_1 r_1^{L-l-1}(r_1+b) + \rho_2 r_2^{L-l-1}(r_2+b)\right). \tag{58}$$

Thanks to the block-matrix inversion formula [3], we now can find the expression of the inverse of the HIM

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}_{11}^{-1} + \mathbf{V}_L & -\lambda^{-1}\mathbf{H}_{11}^{-1}\mathbf{H}_{12} \\ -\lambda^{-1}\mathbf{H}_{12}^T\mathbf{H}_{11}^{-1} & \lambda^{-1} \end{bmatrix}, \tag{59}$$

where

$$\lambda \triangleq \frac{L-1}{\sigma_w^2} - \mathbf{H}_{12}^T\mathbf{H}_{11}^{-1}\mathbf{H}_{12}, \tag{60}$$

$$\text{and } \mathbf{V}_L \triangleq \lambda^{-1}\mathbf{H}_{11}^{-1}\mathbf{H}_{12}\mathbf{H}_{12}^T\mathbf{H}_{11}^{-1}. \tag{61}$$

We start to compute λ corresponding to the inverse of the minimum bound on ξ. Due to the particular structures of matrices $\mathbf{H}_{11}$ and $\mathbf{H}_{12}$ (see (47),(49)), we get

$$\lambda = \frac{L-1}{\sigma_w^2} - \frac{2}{\sigma_w^4}\left(\left[\mathbf{H}_{11}^{-1}\right]_{1,1} - \left[\mathbf{H}_{11}^{-1}\right]_{1,L}\right), \tag{62}$$

where $\left[\mathbf{H}_{11}^{-1}\right]_{1,1}$ and $\left[\mathbf{H}_{11}^{-1}\right]_{1,L}$ are given by (58).

We now derive an analytical expression for the diagonal elements of the upper left part $\left(\mathbf{H}_{11}^{-1} + \mathbf{V}_L\right)$ corresponding to the minimum bound on $\boldsymbol{\theta}$ (see (59)). From the definition of $\mathbf{V}_L$ in (61), the diagonal elements $\left[\mathbf{V}_L\right]_{l,l}$ can be written as

$$\left[\mathbf{V}_L\right]_{l,l} = \frac{1}{\lambda\sigma_w^4}\left(\left[\mathbf{H}_{11}^{-1}\right]_{1,l} - \left[\mathbf{H}_{11}^{-1}\right]_{L,l}\right)^2 = \frac{1}{\lambda\sigma_w^4}\left(\left[\mathbf{H}_{11}^{-1}\right]_{1,l} - \left[\mathbf{H}_{11}^{-1}\right]_{1,L+1-l}\right)^2. \tag{63}$$

Inserting (58) into (63) and then into (59), we obtain the analytical expression of the upper diagonal elements $\left[\mathbf{H}^{-1}\right]_{l,l}$ *i.e.* the off-line HCRB associated to the estimation of θ_l $(1 \le l \le L)$

$$\left[\mathbf{H}^{-1}\right]_{l,l}=\frac{1}{\left|\mathbf{H}_{11}\right|}\left(\rho_1^2\left(b+r_1\right)^2 r_1^{L-3}+\rho_2^2\left(b+r_2\right)^2 r_2^{L-3}-b^2\left(r_1^{L-2}r_2^{L-l-1}+r_1^{L-l-1}r_2^{L-2}\right)\left(A-2\right)^{-1}\right)$$
$$+\frac{b^2}{\lambda\left|\mathbf{H}_{11}\right|^2}\left(b^{l-1}\left(\rho_1 r_1^{L-l-1}\left(b+r_1\right)+\rho_2 r_2^{L-l-1}\left(b+r_2\right)\right)+b^{L-l}\left(\rho_1 r_1^{l-2}\left(b+r_1\right)+\rho_2 r_2^{l-2}\left(b+r_2\right)\right)\right)^2. \quad (64)$$

Comparing to (52), we notice that the contribution of $\mathbf{V}_L$ in (64) is the additional uncertainty brought by the linear drift ξ.

G. High SNR Asymptote Bound

In the high-SNR range, there exists one constellation point $s_m^{\max}$ for the NDA case (resp. $\tilde{s}_i^{\max}$ for the CA case) which has a dominant a posteriori probability. Substituting $\Pr\left(s_m^{\max}\mid y_m,\theta_m,\xi\right)\approx 1$ into (31) (resp. $\Pr\left(s_m=\tilde{s}_l^{\max}\mid \mathbf{y},\boldsymbol{\theta},\xi\right)\approx 1$ into (45)), the second order derivatives can be respectively expressed as

$$\frac{\partial^2 \ln p\left(\mathbf{y}\mid\boldsymbol{\theta},\xi\right)}{\partial\theta_m^2}\approx\begin{cases}-\dfrac{2\,\mathrm{Re}\left\{y_m\left(s_m^{\max}\right)^* e^{-j\theta_m}\right\}}{\sigma_n^2}, & \text{NDA},\\ -\dfrac{2\,\mathrm{Re}\left\{y_m\left(\tilde{s}_i^{\max}\right)^* e^{-j\theta_m}\right\}}{\sigma_n^2}, & \text{CA}.\end{cases} \quad (65)$$

Thus both NDA and CA scenarios are equivalent to the DA scenario at high SNR and one has

$$J_{Dh}^{(\mathrm{NDA})}=J_{Dh}^{(\mathrm{CA})}=J_D^{(\mathrm{DA})}=\frac{2\sigma_s^2}{\sigma_n^2}=2\cdot\mathrm{SNR}. \quad (66)$$

We thus obtain the same result as in [31]; by introducing this value into the mathematical expression of the bound, we obtain similarly to section V of [31], an asymptotic bound which is under the bound itself. Note that a far more complex expression for $J_{Dl}^{(\mathrm{NDA})}$ can be obtained at low SNR, and one can combine the low and high SNR HCRB asymptotes to lead to an asymptotic lower bound without any Monte-Carlo simulation.

V. On-line Bounds

Up to this section, we have focused on the off-line scenario. We now show how the previous results can be directly used in the case of an on-line synchronization mode. In this mode, only the past and the current observations are available, *i.e.* $\mathbf{y}_1^l=\left[y_1,\cdots,y_l\right]^T$ where $l\le L$. Like in section III of [31], the on-line BCRB (resp. HCRB) associated to vector $\mathbf{y}_1^L$ is equal to entry (L,L) of the inverse of the BIM (resp. HIM); one thus readily obtains the analytical expression of the on-line BCRB (from (52)) and on-line HCRB (from (64)) associated to the estimation of θ_l ($l\ge 3$)

$$C_{\mathbf{B}_l}=\frac{1}{\left|\mathbf{B}_l\right|}\left(\rho_1^2\left(b+r_1\right)^2 r_1^{l-3}+\rho_2^2\left(b+r_2\right)^2 r_2^{l-3}-b^2\left(r_1^{l-2}r_2^{-1}+r_1^{-1}r_2^{l-2}\right)\left(A-2\right)^{-1}\right), \quad (67)$$

$$\text{and } C_{\mathbf{H}_l} = \frac{1}{\left|\mathbf{H}_{11}(l)\right|}\left(\rho_1^2\left(b+r_1\right)^2 r_1^{l-3} + \rho_2^2\left(b+r_2\right)^2 r_2^{l-3} - b^2\left(r_1^{l-2}r_2^{-1} + r_1^{-1}r_2^{l-2}\right)(A-2)^{-1}\right)$$
$$+ \frac{b^2}{\lambda\left|\mathbf{H}_{11}(l)\right|^2}\left(b^{l-1}\left(\rho_1 r_1^{-1}\left(b+r_1\right) + \rho_2 r_2^{-1}\left(b+r_2\right)\right) + \rho_1 r_1^{l-2}\left(b+r_1\right) + \rho_2 r_2^{l-2}\left(b+r_2\right)\right)^2, \tag{68}$$

where r_1, r_2, ρ_1 and ρ_2 are given by (53),(54), and $\mathbf{H}_{11}(l)$ is a $l \times l$ matrix with the same form as $\mathbf{H}_{11}$ of equation (47).

VI. Discussion

We now display simulation results with Gray mapping QAM symbols for different scenarios. For the code aided scenario, we use a recursive systematic rate $1/2$ turbo code (mother code rate $1/3$) whose generator polynomials are $G_1 = 37_{OCT}$ and $G_2 = 21_{OCT}$ with a S random interleaver [41]. Since turbo code can be described by means of a trellis, the marginal APPs can be efficiently computed by the famous BCJR algorithm [37].

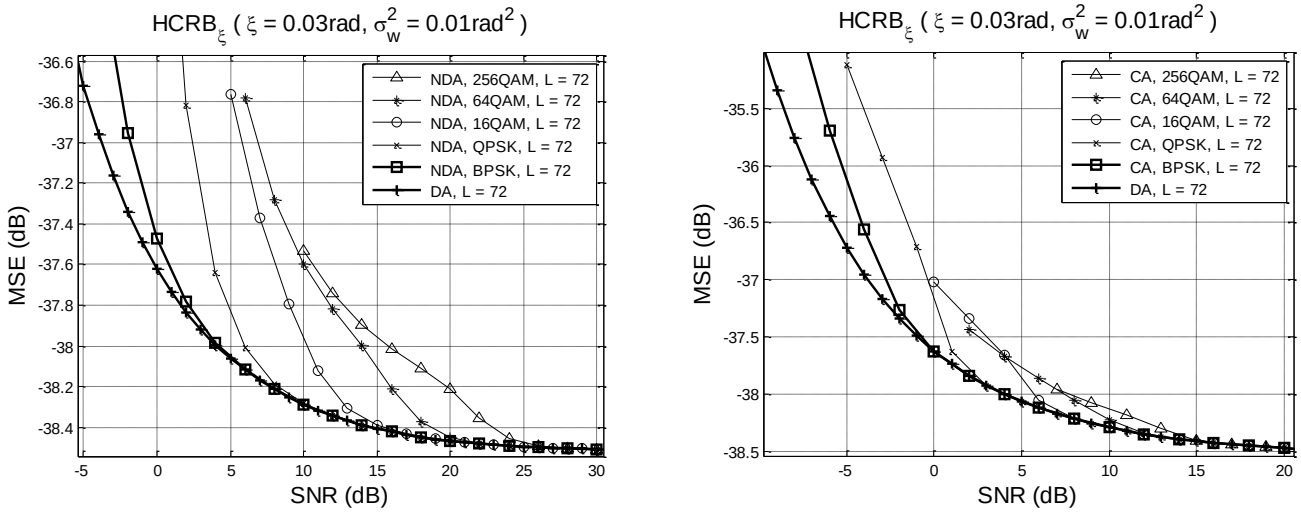


Fig. 1 NDA (left), CA (right) and DA HCRB_ξ for different constellations as a function of the SNR

We start with figure 1 displaying the behavior of the HCRB on ξ as a function of the SNR; we compare the performance of NDA (on the left), CA (on the right) and DA scenarios for various constellations. At high SNR, we first see that HCRB_ξ converges to its horizontal asymptote ($\sigma_w^2/(L-1)$) for here a symbol block length $L = 72$. The observation noise has then no influence on the estimation of the linear drift and clearly, for a larger block length or a smaller phase noise, the asymptote would be lower. Then at median SNR, HCRB_ξ leaves its asymptote and this happens at a larger SNR if a larger constellation is used. Note that the CA scenario allows to have better performance that the NDA aided scenario.

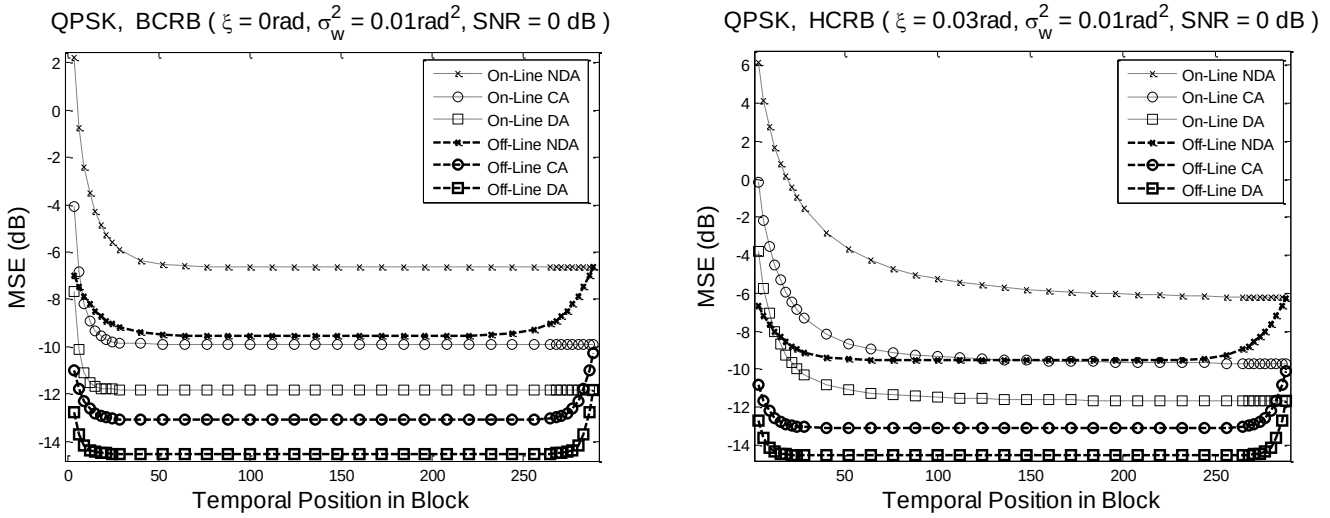


Fig. 2 QPSK CRB curves for the various block positions when the SNR = 0 dB

In figure 2, we now turn to the behavior of the dynamic phase bound and display some results for QPSK modulated signals as a function of the position in the block ($L = 288$). We first remark that in all the cases (BCRB/HCRB and DA/CA/NDA), the off-line scenario always gives a better result that the corresponding on-line result. To be more precise, the on-line curves always benefit from more and more observations until they saturate; at the last position the on-line bound corresponds exactly to the off-line bound. Contrarily to the on-line scenario, for the off-line scenario the best phase estimate is achieved at midblock, whereas they become poorer at the block border. This comes from the fact in the center position of the phase vector, one has more adjacent (past or future) and strongly correlated variables (see the phase model of (12)) than at the border of the block. Also logically, the DA scenario achieves a better performance than the CA aided scenario and even more than the NDA scenario. The BCRB provides the lower bound for this general block-phase estimation framework with a known linear drift ξ (or $\xi = 0$), whereas the HCRB provides the lower bound with an unknown linear drift ξ . The HCRB is always lower bounded by the BCRB for any position in the block but for such a block length, there is almost no loss of information due to the linear drift for the off-line case in the center position so that the HCRB almost corresponds to the BCRB.

Instead of looking at the different positions for a given SNR, we now display HCRB curves as function of the SNR for the phase at the central position of the block. Figure 3 displays results for the BPSK modulated signal**.** For codes that are described by means of a trellis, the marginal symbol APPs in (40) and (42) can be approximately computed from the trellis state APPs and state transition APPs, which in turn can be determined efficiently by the famous BCJR algorithm. Like already observed on figure 2, one can see on figure 3 that the off-line scenario allows better results than the on-line scenario and also that the NDA case is worse compared to the CA and to the DA cases. Here, it is important to note that one can expect a better performance with the NDA

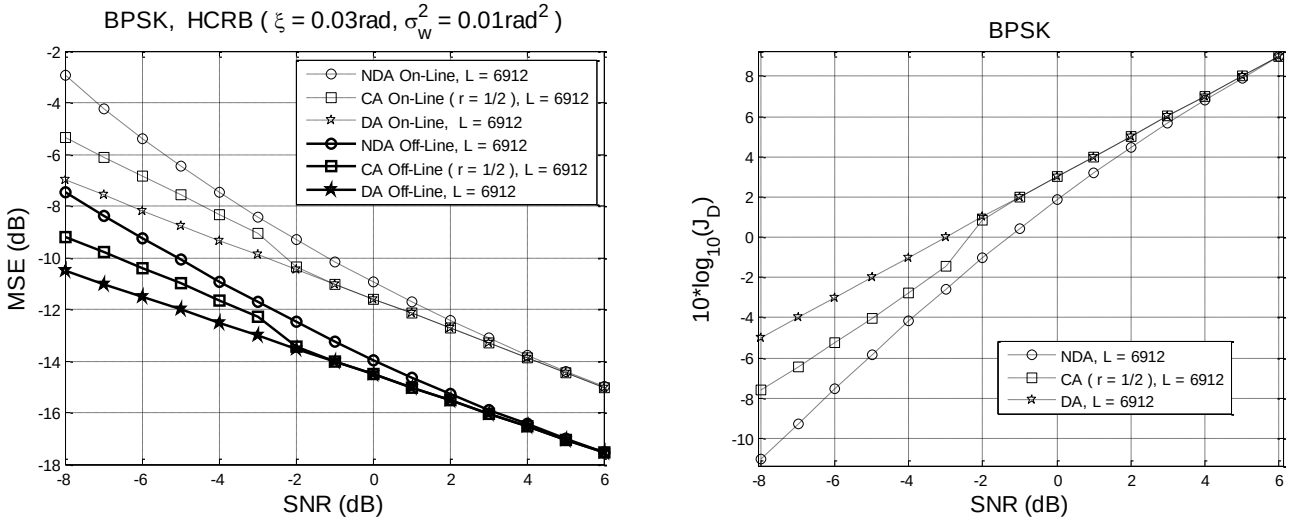


Fig. 3 BPSK NDA, turbo-CA and DA HCRB curves (left) and corresponding J_D (right) as functions of the SNR

off-line scenario rather than with the CA on-line scenario. Stated in other words, in general, one can expect more benefits from the estimation method (off-line/on-line) rather than the way of using data (DA, CA or NDA). This can also be observed for other modulations and other coding schemes (e.g. simple convolutional codes) and this is a general tendency; however the results are not always totally so simple as displayed in the following figure 4 with a 16QAM modulation and some additional parameters must be taken into account.

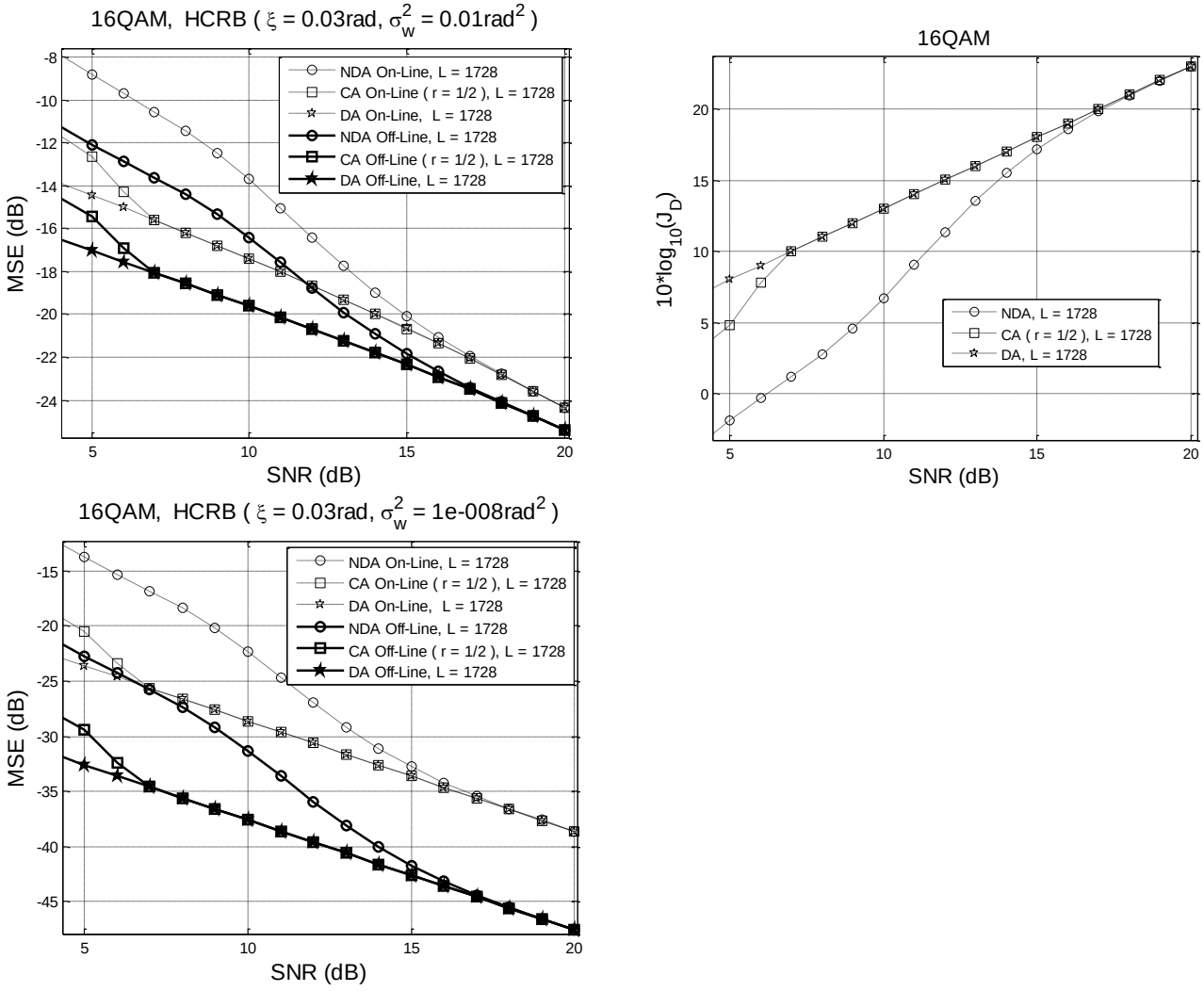


Fig. 4 16 QAM NDA, turbo-CA and DA HCRBs for two different phase variances (left) and corresponding J_D (right) as functions of the SNR

To get more insight into those curves we also display the various J_D as functions of the SNR on the right side of figure 3 and 4. From the asymptote analysis of Section IV.G, we know that at high SNR, J_D is just proportional to the SNR (see (66)) but watching the left part of these figures, we can see that at lower SNR , J_D behaves like a non linear amplifier of SNR and behaves differently according to the DA, CA, and NDA scenarios. This behaviour allows predicting exactly where the different bounds leave their asymptote. The J_D curves also give an indication whether one can expect a large deviation from the asymptote; but one also has to take into account the phase variance value to know the exact behaviour of the bounds (see figure 4 for two different phase variances).

Finally, figure 5 (resp. figure 6) displays for the off-line scenario, the NDA (resp. CA) HCRB on θ_k as a function of the SNR in the center position for various constellations. Like it was observed on previous figures, the on-line scenario would give exactly similar results that are not displayed here.

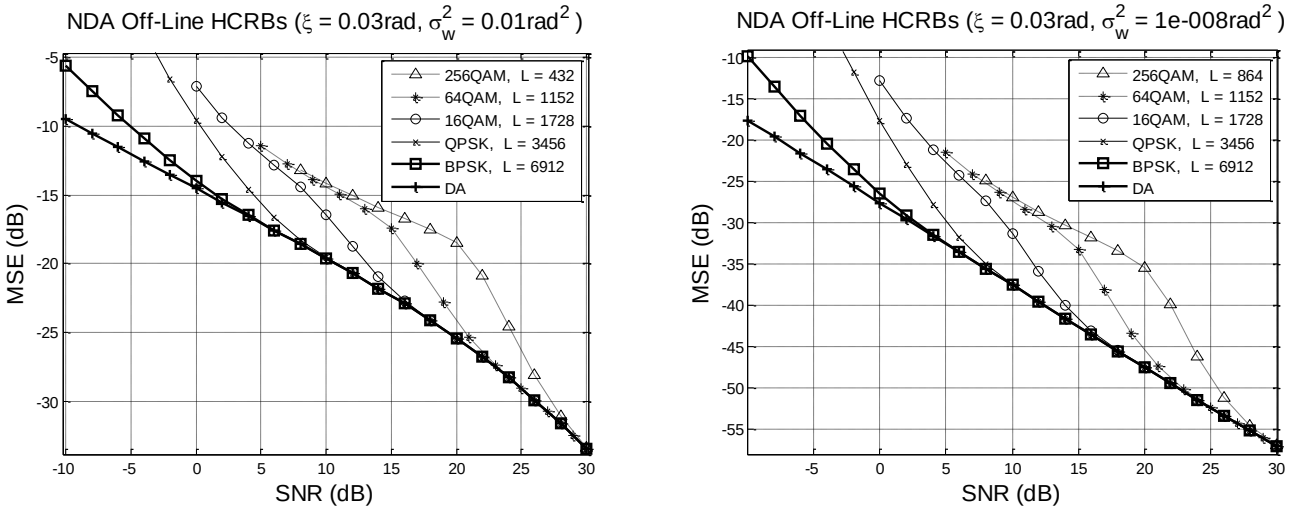


Fig. 5 NDA HCRB curves in the center position for different constellations and for two different phase variances

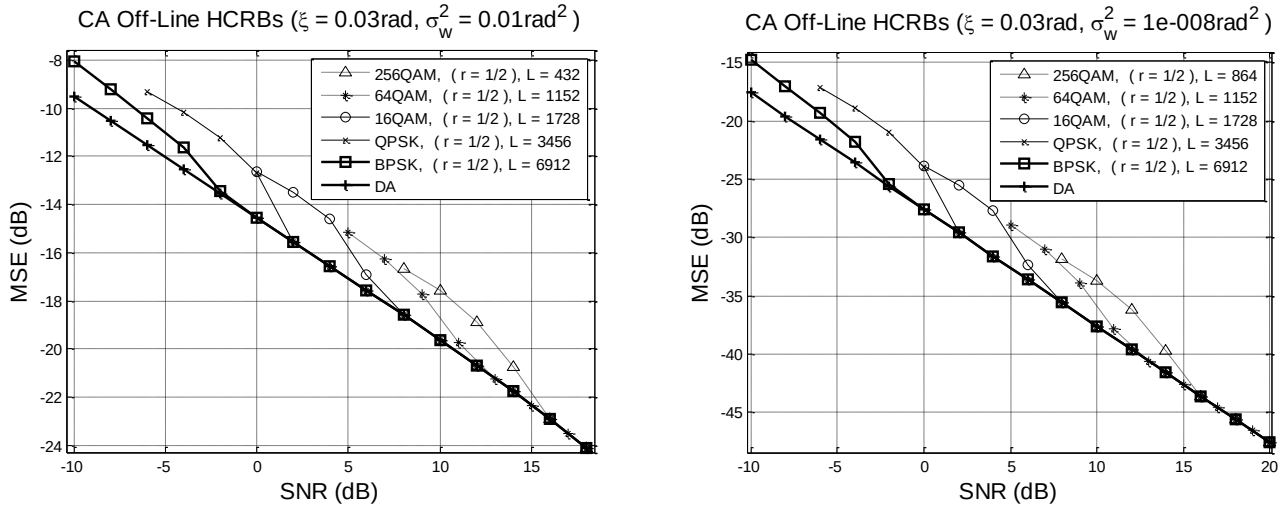


Fig. 6 CA HCRB curves in the center position for different constellations and for two different phase variances

- At high SNR, we notice that the various on-line and off-line CRBs logically merge and do not depend on the constellation. In this range of SNR, the received constellations are reliable enough to make a correct decision and the information provided by each observation y_l is preponderant over the a priori knowledge of $\boldsymbol{\theta}$. In this case it is sufficient to only take into account the present observation y_l in order to estimate θ_l and this is why the various constellation bounds merge. Furthermore, this is also the reason why the NDA and CA CRBs logically tend to the DA CRBs. As the a priori distribution of $\boldsymbol{\theta}$ has no influence, the estimation problem tends to a deterministic phase estimation problem where we estimate L independent phase θ_l with L independent observations.
- In mid-range SNRs whose values are different with the different constellations, the HCRBs leave their asymptotes. One observation is thus not sufficient to estimate the phase offset while a block of more observations should be used to improve

the estimation performance. This also explains why the NDA CRBs do not merge anymore with the DA CRBs. Moreover, we note that every time the density of the constellation is increased by a factor of 4, (*i.e.* the constellation size is multiplied by 4), the threshold where the NDA bounds leave the DA bound is increased by 6dB. Contrarily to the NDA CRBs, the CA CRBs continue to coincide with the DA CRBs till a SNR at least 6dB lower than the one of the NDA CRBs. Obviously, the performance prominently benefits from the decoding.

- At low SNRs, the CA CRBs do not merge anymore with the DA CRBs. The observation noise becomes preponderant compared to the phase noise and directly affects the estimation performance; also, the lack of knowledge on ξ further deteriorates the observed performance. Note that the NDA bound for the robust real BPSK scheme has not the same slope than the other complex constellations NDA bounds.

VII. Conclusion

In this paper, we have derived the on-line and off-line Bayesian Cramér-Rao bounds (BCRBs) and hybrid Cramér-Rao bounds (HCRBs) for the QAM dynamic phase estimation in data-aided (DA), non-data-aided (NDA) and code-aided (CA) scenarios. We were able to provide closed-form expressions for a realistic dynamic phase model and for realistic telecommunication systems involving turbo-codes and large constellations. We have shown that off-line bounds allow to have better performance than on-line bounds. Moreover, most of the time, NDA off-line bounds are eager to provide better performance than CA on-line bounds. We also saw that the often studied BPSK case does not tell the whole story and that there is some space for additional CA synchronization gain at low SNR for larger constellations. These bounds provide a powerful tool for the communication system design in stressing environments.

References

[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Royal Soc.*, vol. 222, pp. 309–368, 1922.

[2] H. L. V. Trees, *Detection, Estimation and Modulation Theory*. New York Wiley, 1968, vol. 1.

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

[4] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe "Code-aided turbo synchronization," *Proceedings of the IEEE*, vol. 95, pp. 1255-1271, June 2007.

[5] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers Synchronization, Channel Estimation and Signal Processing*, New York Wiley, 1998.

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

[7] C. Berrou, J. Hagenauer, M. Luise, C. Schlegel, and L. Vandendorpe "Turbo-Information Processing Algorithms, Implementations & Applications," *Editorial, IEEE Proceedings,* vol. 95, June 2007.

[8] C. Berrou, C. Langlais, and F. Seguin, "Turbo Processing," *New Directions in Statistical Signal Processing From Systems to Brain,* MIT Press Cambridge, 2006.

[9] D.C, Rife and R.R. Boorstyn, "Single-tone Parameter Estimation from Discrete-Time Observations," *IEEE Trans. Inf. Theory*, vol. IT-20, No. 5, pp. 591-597, Sept. 1974.

[10] J.A. Gansman, J.V. Krogmeier, and M.P. Fitz, "Single Frequency Estimation with Non-Uniform Sampling," *in Proc. of the 13th Asilomar Conference on Signals, Systems and Computers*, pp.878-882, Pacific Grove, CA, Nov. 1996.

[11] N. Noels, H. Steendam, and M. Moeneclaey, "Pilot-Symbol Assisted Carrier Synchronization Cramér-Rao Bound and Synchronizer Performance," *IEEE SCVT'03*, Eindhoven, The Netherlands, Paper 28, Nov. 13, 2003.

[12] W.G. Cowley, "Phase and frequency estimation for PSK packets Bounds and algorithms," *IEEE Trans. on Commun.*, vol. COM-44, pp. 26-28, Jan. 1996.

[13] N. Noels, H. Steendam, and M. Moeneclaey, "The True Cramér-Rao Bound for Phase-Independent Carrier Frequency Estimation from a PSK Signal," *IEEE GLOBECOM'02*, Taipei, Taiwan, R.O.C., Nov. 17-21, 2002.

[14] N. Noels, H. Steendam, M. Moeneclaey, and H. Bruneel, "Carrier Phase and Frequency Estimation for Pilot-Symbol Assisted Transmission Bounds and Algorithms," *IEEE Trans. on Signal Processing*, vol. 53, pp. 4578-4587, Dec. 2005.

[15] N. Noels, H. Steendam and M. Moeneclaey, "The Impact of the Observation Model on the Cramér-Rao Bound for Carrier and Frequency Synchronization," *in Proc. IEEE ICC 2003,* Anchorage, Alaska, Paper CT04-1, May 2003.

[16] M. Moeneclaey, "On the True and the Modified Cramér-Rao Bounds for the Estimation of a Scalar Parameter in the Presence of Nuisance Parameters," *IEEE Trans. on Comm.*, vol. 46, pp. 1536–1544, Nov. 1998.

[17] N. Noels and M. Moeneclaey, "True Cramér-Rao Bound for Estimating Synchronization Parameters from a Linearly Modulated Bandpass Signal with Unknown Data Symbols," *in Proc. 2nd IEEE Inter. Workshop CAMSAP 2007*, Saint-Thomas, US Virgin Islands, pp. 29-32, Dec. 12-14, 2007.

[18] F. Rice, B. Cowley, B. Moran, and M. Rice, "Cramér-Rao lower bounds for QAM phase and frequency estimation," *IEEE Trans. on Commun.*, vol. 49, pp 1582-1591, Sep. 2001.

[19] H. Steendam and M. Moeneclaey, "Low-SNR Limit of the Cramér-Rao Bound for Estimating the Carrier Phase and Frequency of a PAM, PSK or QAM Waveform," *IEEE Communications Letters*, vol. 5, pp. 215-217, May 2001.

[20] H. Steendam and M. Moeneclaey, "Performance Bounds in Synchronization for Low Signal to Noise Ratios," *in Proc. 7th International Workshop on Digital Signal Processing Techniques for Space Communications DSP2001,* Sesimbra (Portugal), paper 1.2, Oct. 2001.

[21] N. Noels, H. Steendam, and M. Moeneclaey, "The true Cramér-Rao Bound for Carrier Frequency Estimation from a PSK Signal," *IEEE Trans. Commun.*, vol. 52, pp. 834–844, May 2004.

[22] N. Noels, H. Steendam, and M. Moeneclaey, "True Cramér-Rao Bounds for Carrier and Symbol Synchronization," *EURASIP European Signal Processing Conference 2002, EUSIPCO'02*, Toulouse, France, Paper 890, Sep. 3-6, 2002 (invited paper).

[23] N. Noels, H. Steendam, and M. Moeneclaey, "The True Cramér-Rao Bound for Estimating the Carrier Phase of a Convolutionally Encoded PSK Signal," *9th Symposium on Vehicular Technology and Communications 2002, SCVT'02*, Louvain-la-Neuve, Belgium, pp. 9-14, Oct. 17, 2002.

[24] N. Noels, H. Steendam, and M. Moeneclaey, "The Cramér-Rao Bound for Phase Estimation from Coded Linearly Modulated Signals," *IEEE Communication Letters*, vol. 7, No. 5, pp. 207-209, May 2003.

[25] N. Noels, H. Steendam, and M. Moeneclaey, "Carrier and Clock recovery in (turbo) coded systems Cramér-Rao Bound and Synchronizer Performance," *EURASIP Journal on Applied Signal processing, Special Issue on Turbo Processing*, vol. 2005, pp. 972-980, May 2005.

[26] A. Barbieri, D. Bolletta, and G.Colavolpe, "On the Cramér-Rao Bound for Carrier Frequency Estimation in the Presence of Phase Noise", *IEEE Global Telecommunications Conference*, 2005, GLOBECOM'05. vol. 3, Nov.-2 Dec. 2005.

[27] P.O. Amblard, J.M. Brossier, and E. Moisan, "Phase Tracking What Do We Gain from Optimality? Particle Filtering versus Phase-Locked Loops," *Elsevier Signal Processing*, vol. 83, pp. 151–167, Oct. 2003.

[28] J. A. McNeill, "Jitter in ring oscillators," Ph.D. dissertation, Boston University, 1994.

[29] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase Noise in Oscillators a Unifying Theory and Numerical Methods for Characterization," *IEEE Trans. Circuits Syst.* I, vol. 47, pp. 655-674, May 2000.

[30] ETSI EN 302 307 "Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications," June 2004.

[31] S. Bay, C. Herzet, J. Barbot, J.M. Brossier, and B. Geller, "Analytic and Asymptotic Analysis of Bayesian Cramér–Rao Bound for Dynamical Phase Offset Estimation," *IEEE Trans. on Signal Processing*, vol. 56, pp. 61-70, Jan. 2008.

[32] S. Bay, B. Geller, A. Renaux, J.P. Barbot, and J.M. Brossier, "On the Hybrid Cramér-Rao bound and its Application to Dynamical Phase Estimation," *IEEE Signal Processing Letters*, vol. 15, pp. 453-456, 2008.

[33] J. Yang, B. Geller, and A. Wei, "Approximate Expressions for Cramér-Rao Bounds of Code Aided QAM Dynamical Phase Estimation," *in Proc. IEEE Inter. Conf. on Commun. 2009, ICC'09*, Dresden, 14-18 June 2009.

[34] J. Yang and B. Geller, "Near Optimum Low Complexity Smoothing Loops for Dynamical Phase Estimation - Application to BPSK Modulated Signals," IEEE Trans. on Signal Processing, pp. 3704-3711, Sept. 2009.

[35] J. Yang, B. Geller, C. Herzet, and J.M. Brossier, "Smoothing PLLs for QAM Dynamical Phase Estimation," *in Proc. IEEE Inter. Conf. on Commun. 2009, ICC'09*, Dresden, June 14-18, 2009.

[36] J. Yang, B. Geller, C. Herzet, and J.M. Brossier, "Near MAP Smoothing Loops for Code Aided QAM Dynamical Phase Estimation," *submitted to IEEE Trans. on Commun.*, 2010.

[37] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Info. Theory*, vol. IT-20(2), pp. 284-287, Mar, 1974.

[38] J. Pearl, "Reverend Bayes on Inference Engines a Distributed Hierarchical Approach," *Proc. American Association of Artificial Intelligence National Conference on AI*, pp.133-136, Pittsburgh, PA. 1982.

[39] J.H. Kim and J. Pearl, "A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems," *Proc. IJCAI-83*, Karlsruhe, Germany, pp. 190-193. 1983.

[40] J. Pearl, "Probabilistic Reasoning in Intelligent Systems Networks of Plausible Inference (Revised Second Printing)," San Francisco, CA Morgan Kaufmann, 1988.

[41] D. Divsalar and F. Pollara, "Turbo Codes for PCS Applications," *in Proc. ICC'95*, pp. 54-59, Seattle, WA, June 1995.

[42] ETSI EN 300 744, "Digital Video Broadcasting (DVB); Frame structure, channel coding and modulation for digital terrestrial television," v1.5.1, June, 2004.

APPENDIX

From (33), one can readily obtain the first derivative

$$\begin{aligned}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m} &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K}\left(\prod_{l=1,l\neq m}^{L} p\left(y_l \,|\, s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\frac{\partial p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\right)p(\mathbf{c}=\tilde{\mathbf{c}}_v)\\ &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K}\left(\prod_{l=1}^{L} p\left(y_l \,|\, s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right)p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{1}{p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}\frac{\partial p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\\ &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K} p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)\,p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}.\end{aligned} \tag{69}$$

Note that (69) can be further simplified as

$$\begin{aligned}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m} &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K} p(\mathbf{y},\mathbf{c}=\tilde{\mathbf{c}}_v\,|\,\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\\ &= \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\,|\,\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}.\end{aligned} \tag{70}$$

For $n\neq m$, taking the second derivative, one obtains

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m \partial \theta_n} &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K}\frac{\partial p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)}{\partial \theta_n}p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\\ &\quad -\frac{1}{\left(p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)\right)^2}\frac{\partial p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}\sum_{v=1}^{2^K} p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)\,p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\\ &= \frac{\sum_{v=1}^{2^K}\frac{\partial\left(\prod_{l=1}^{L} p\left(y_l \,|\, s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right)}{\partial \theta_n}p(\mathbf{c}=\mathbf{c}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\\ &\quad -\frac{\frac{\partial p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}\sum_{v=1}^{2^K} p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)\,p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\\ &= \frac{\sum_{v=1}^{2^K}\left(\prod_{l=1,l\neq n}^{L} p\left(y_l \,|\, s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\frac{\partial p\left(y_n \,|\, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial \theta_n}\right)p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)} - \frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}\\ &= \sum_{v=1}^{2^K}\frac{\left(\prod_{l=1}^{L} p\left(y_l \,|\, s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right)p(\mathbf{c}=\tilde{\mathbf{c}}_v)}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\frac{\partial \ln p\left(y_n \,|\, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial \theta_n} - \frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}\\ &= \sum_{v=1}^{2^K}\frac{p(\mathbf{y}\,|\,\mathbf{c}=\tilde{\mathbf{c}}_v,\boldsymbol{\theta},\xi)\,p(\mathbf{c}=\tilde{\mathbf{c}}_v)}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\frac{\partial \ln p\left(y_n \,|\, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial \theta_n} - \frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}\\ &= \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\,|\,\mathbf{y},\boldsymbol{\theta},\xi)\frac{\partial \ln p\left(y_m \,|\, s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial \theta_m}\frac{\partial \ln p\left(y_n \,|\, s_n=\tilde{s}_n(\tilde{\mathbf{c}}_v),\theta_n,\xi\right)}{\partial \theta_n} - \frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_n}.\end{aligned} \tag{71}$$

For $n = m$, the second derivative becomes

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial \theta_m^2} &= \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K}\frac{\partial\left(\prod_{l=1}^{L} p\left(y_l\,|\,s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right)}{\partial\theta_m} p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m\,|\,s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m} \\ &\quad + \frac{1}{p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}\sum_{v=1}^{2^K}\left(\prod_{l=1}^{L} p\left(y_l\,|\,s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right) p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial^2 \ln p\left(y_m\,|\,s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m^2}\right) \\ &\quad - \frac{\sum_{v=1}^{2^K}\left(\prod_{l=1}^{L} p\left(y_l\,|\,s_l=\tilde{s}_l(\tilde{\mathbf{c}}_v),\theta_l,\xi\right)\right) p(\mathbf{c}=\tilde{\mathbf{c}}_v)\frac{\partial \ln p\left(y_m\,|\,s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}\frac{\partial p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial\theta_m}}{\left(p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)\right)^2} \\ &= \sum_{v=1}^{2^K}\Pr(\mathbf{c}=\tilde{\mathbf{c}}_v\,|\,\mathbf{y},\boldsymbol{\theta},\xi)\left(\left(\frac{\partial \ln p\left(y_m\,|\,s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m}\right)^2 + \frac{\partial^2 \ln p\left(y_m\,|\,s_m=\tilde{s}_m(\tilde{\mathbf{c}}_v),\theta_m,\xi\right)}{\partial\theta_m^2}\right) - \left(\frac{\partial \ln p(\mathbf{y}\,|\,\boldsymbol{\theta},\xi)}{\partial\theta_m}\right)^2.\end{aligned} \tag{72}$$