Analytic and Asymptotic Analysis of Bayesian Cramér-Rao Bound for Dynamical Phase Offset Estimation

Stéphanie Bay, Cédric Herzet, Jean-Marc Brossier, Jean-Pierre Barbot, Benoit Geller

Analytic and Asymptotic Analysis of Bayesian Cramér-Rao Bound for Dynamical Phase Offset Estimation

S. Bay[(1)], C. Herzet[(1,2)], J.M. Brossier[(3)], J.P. Barbot[(1)], and B. Geller[(1)]
[(1)]ENS Cachan / Laboratoire SATIE, France
[(2)]UC Berkeley / Dept. of EECS, Berkeley, CA 94720
[(3)]INPG / Laboratoire LIS, Saint Martin d'Hères, France

***Abstract*— In this paper, we present a closed-form expression of a Bayesian Cramér-Rao lower bound for the estimation of a dynamical phase offset in a non-data-aided BPSK transmitting context. This kind of bound is derived considering two different scenarios: a first expression is obtained in an off-line context and then, a second expression in an on-line context logically follows. The SNR-asymptotic expressions of this bound drive us to introduce a new asymptotic bound, namely the Asymptotic Bayesian Cramér-Rao Bound. This bound is close to the classical Bayesian bound but is easier to evaluate.**

I. Introduction

During the last decade, synchronization has become one of the most challenging task of digital receivers. On the one hand, since the advent of very powerful error correcting codes (see *e.g.*, [1]–[3]), the synchronizers have to properly operate at Signal-to-Noise Ratios (SNRs) lower than ever before. On the other hand, the high data-rate requirements of future communication standards imply rapidly varying synchronization parameters at the receiver. In order to cope with such stressful environments, Maximum-Likelihood (ML) synchronization performance are often a desirable feature. Unfortunately, the ML estimator is often unfeasible in practice. Hence, sub-optimal synchronizers have been proposed, see *e.g.*, [4].

In order to assess the performance of such synchronizers, lower bounds on mean-squared estimation error are needed. In particular, the family of Cramér-Rao Bounds (CRBs) [5], [6] has been shown to give tight estimation lower bounds in a number of practical scenarios, (see *e.g.*, [7]–[11] for CRB contributions associated to synchronization problems). Rife *et al.* [7] and Cowley [8] have derived CRB closed-form expressions for constant phase-offset estimation in the so-called Data-Aided (DA) and Non-Data-Aided (NDA) scenarios, respectively. In [9], the authors proposed a semi-analytical method enabling to efficiently evaluate CRBs in Code-Aided (CA) scenarios. A Modified CRB (MCRB), easier to evaluate than the standard CRB, has been introduced in [12], [13]. The MCRB is usually looser than the CRB but it is equivalent to the CRB at high SNR [14]. More recently, the problem of deriving CRBs suited to time-varying parameters has been addressed throughout the Bayesian context. In [15], the authors propose a general framework for deriving analytical expression of *on-line* or also called *filtering* CRBs. In [16], the authors developed a numerical graph-based algorithm to evaluate the CRB in time-varying scenarios.

In this contribution, we address the open problem of deriving an analytical expression of the *off-line* or also called *smoothing* Bayesian CRB (BCRB) for time-varying phase estimation in NDA scenarios.

Explicit expressions of the BCRB and its modified version are provided in an *off-line* scenario. This bound helps us to evaluate and also to predict the estimator performance without any particular assumption or simulation restriction. Moreover, an *on-line* interpretation follows. The asymptotic cases at low and high SNR are presented. In particular it is shown that the Modified BCRB (MBCRB) is equal to the derived asymptote at high SNR. A new asymptote-based bound is then introduced simplifying the general expressions of the BCRB.

This paper is organized as follows. In Section II we set the system model. After recalling in Section III the different kinds of Cramér-Rao bounds, we derive in Section IV the *off-line* BCRB. In Section V the asymptotic cases are considered and we present the asymptote-based lower bound. Next, the closed-form expression of the BCRB in the *on-line* context is given (Section VI). Finally, the different results are illustrated and interpreted in Section VII.

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 $[A]_{m,n}$. The matrix transpose operator is indicated by a superscript ^{T} as in $\mathbf{A}^T$. $|.|$ denotes the modulus of a scalar quantity or the determinant of a matrix quantity. $\mathbf{a}_m^n$ represents the vector $[a_m \cdots a_n]^T$, where m and n are positive integers ($m < n$). $\Re\{a\}$ and $\Im\{a\}$ are respectively the real and the imaginary part of a. $E_{xy}[.]$ denotes the expectation over x and y. $\lfloor k$ is equal to the integer part of k. $\delta_{k,l}$ is the Kronecker symbol. $\nabla_{\boldsymbol{\theta}}$ and $\Delta_{\boldsymbol{\psi}}^{\boldsymbol{\theta}}$ represent the first and second-order partial derivatives operator *i.e.*, $\nabla_{\boldsymbol{\theta}} = \left[\frac{\partial}{\partial\theta_1} \cdots \frac{\partial}{\partial\theta_K}\right]^T$ and $\Delta_{\boldsymbol{\psi}}^{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\psi}}\nabla_{\boldsymbol{\theta}}^T$. The error function is defined as $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. The actual value of an unknown parameter is indicated by the

This work is partially supported by the European Community contract NO. 507325, NEWCOM.

superscript * as in $\theta^\star$.

II. System Model

We consider the transmission of a BPSK modulated sequence $\mathbf{a} = [a_1 \cdots a_K]^T$ over an Additive White Gaussian Noise (AWGN) channel affected by some carrier phase offsets stacked in a vector $\boldsymbol{\theta} = [\theta_1 \cdots \theta_K]^T$. Assuming that the received signal has been ideally filtered and sampled at the optimum sampling instant, the discrete-time baseband signal $\mathbf{y} = [y_1 \cdots y_K]^T$ is given by

$$y_k = a_k e^{j\theta_k} + n_k \quad \text{with } k = 1 \dots K, \tag{1}$$

where a_k is the k^{th} unknown transmitted BPSK symbol ($a_k = \pm 1$) and n_k is a zero-mean circular Gaussian noise with known variance σ_n^2. We suppose that the system operates in a NDA synchronization mode, *i.e.*, the transmitted symbols are independent and identically distributed (i.i.d.) with $p(a_k = a) = \frac{1}{2}$ where $a = \pm 1$. We assume that the energy per symbol is normalized, so that the noise variance is related to the E_s/N_0-ratio as $\frac{E_s}{N_0} = \frac{1}{\sigma_n^2}$.

We consider the case of a Wiener phase-offset evolution, *i.e.*,

$$\theta_k = \theta_{k-1} + w_k, \tag{2}$$

where w_k is an i.i.d. zero-mean Gaussian noise with known variance σ_w^2. This model is commonly used [17], [18] to describe the behavior of practical oscillators for which the frequency is randomly perturbated. As the phase is the integral of the frequency, the phase variance increases with time index k. This phenomenon is thus non stationary and one readily obtains that $var(\theta_k) = var(\theta_{k-1}) + \sigma_w^2$.

In order to recover the transmitted symbols, practical receivers have first to estimate the carrier phase offsets corrupting the received observations and perturbing the whole data recovery. Phase estimation can actually be considered following two main scenarios:

i) *Off-line synchronization*: the receiver waits until the whole observation frame, *i.e.*, $\mathbf{y} = [y_1 \cdots y_K]^T$, has been received. Then, it processes all the observations to compute the estimates of the carrier phase offsets $\boldsymbol{\theta} = [\theta_1 \cdots \theta_K]^T$.

ii) *On-line synchronization*: the receiver estimates θ_k upon the arrival of the k^{th} observation, *i.e.*, y_k. The phase estimate is then computed based on the current and previous observations only, *i.e.*, $\mathbf{y}_1^k = [y_1\ y_2 \cdots y_k]^T$.

In the sequel, the Cramér-Rao bound evaluation will be considered within the context of both the off-line and the on-line scenarios.

III. Cramér-Rao Bounds: Some Useful Expressions

A natural question which arises when designing estimators is the ultimate accuracy that can be achieved in the estimation operation. The family of CRBs (partially) answers this question by providing a lower bound on the Mean Square Error (MSE) achievable by any unbiased estimator. Different kinds of CRBs can be considered. In the following, we briefly recall general expressions of these bounds.

A. The Standard Cramér-Rao Bound and its Modified Version

Let $\hat{\boldsymbol{\theta}}(\mathbf{y})$ denote an unbiased estimator of $\boldsymbol{\theta}$. Then the estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ satisfies the following inequality

$$E_{\mathbf{y}|\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\left[(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}^\star)(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}^\star)^T\right] \geq \mathrm{SCRB}(\boldsymbol{\theta}^\star), \tag{3}$$

where SCRB is the Standard CRB (SCRB) and $\boldsymbol{\theta}^\star$ is the actual parameter value. The SCRB is equal to the inverse of the Fisher Information Matrix (FIM) $\mathbf{F}(\boldsymbol{\theta}^\star)$ defined by [6]

$$\mathbf{F}(\boldsymbol{\theta}^\star) = E_{\mathbf{y}|\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{y}|\boldsymbol{\theta})_{|\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right]. \tag{4}$$

The practical evaluation of the FIM for modulated signals is generally quite tedious because of the symbols $\mathbf{a}$ which are nuisance parameters. In order to circumvent this problem a "Modified" CRB has been proposed [12] and comes from the inversion of the following matrix:

$$\mathbf{G}(\boldsymbol{\theta}^\star) = E_{\mathbf{a}}\, E_{\mathbf{y}|\mathbf{a},\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{y}|\mathbf{a},\boldsymbol{\theta})_{|\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right]. \tag{5}$$

Unlike the SCRB, it is usually possible to derive a closed form expression of the MCRB. In particular, in [12], the analytical expression associated to a constant phase offset is derived:

$$\mathrm{MCRB} = \frac{1}{2K}\left(\frac{E_s}{N_0}\right)^{-1}, \tag{6}$$

where K is the number of observations. Note that although easier to evaluate than the SCRB, the MCRB is also generally looser. In the particular case of high-SNR transmissions, it has however been shown [14] that the MCRB corresponds to the SCRB. We can also notice that equation (6) coincides with the CRB provided in [7]. This bound is computed in the case of an unmodulated scenario or, equivalently, of a DA scenario.

B. The Bayesian Cramér-Rao Bound and its Modified Version

The standard and the modified CRBs are not suited to time-varying parameter estimation. In particular, they do not take into account the statistical dependence which may exist between phase offsets at different instants.

This dependence is naturally considered within the Bayesian framework: on the one hand, prior distribution $p(\boldsymbol{\theta})$ implicitly models the time dependence between stochastic phase offsets; on the other hand, Bayesian framework is intrinsically based on the knowledge of a vector prior distribution $p(\boldsymbol{\theta})$.

Within this context, a Bayesian CRB (BCRB) has been proposed in [19] such that

$$E_{\mathbf{y},\boldsymbol{\theta}}\left[(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^T\right] \geq \mathrm{BCRB}. \tag{7}$$

Unlike the SCRB and the MCRB, the BCRB does not depend on a particular value $\boldsymbol{\theta}^\star$. The BCRB is the inverse of the Bayesian Information Matrix (BIM), which from [19] can be written as

$$\mathbf{B} = E_{\boldsymbol{\theta}}\left[\mathbf{F}(\boldsymbol{\theta})\right] + E_{\boldsymbol{\theta}}\left[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta})\right], \tag{8}$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the FIM defined in (4). Each term of the BIM (8) plays a different role: the first term can be interpreted as the average information about $\boldsymbol{\theta}$ brought by the observations $\mathbf{y}$ whereas the second term can be regarded as the information

available from the prior knowledge of $\boldsymbol{\theta}$, *i.e.*, $p(\boldsymbol{\theta})$. This term allows us to take into account the time dependence between phase offsets at different instants.

Similarly to the MCRB, a modified version of the BCRB has been considered in [20]. This Modified BCRB (MBCRB) is the inverse of the following information matrix

$$\mathbf{C} = E_{\boldsymbol{\theta}}\big[\mathbf{G}(\boldsymbol{\theta})\big] + E_{\boldsymbol{\theta}}\big[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta})\big], \tag{9}$$

where $\mathbf{G}(\boldsymbol{\theta})$ is the modified Fisher information matrix defined in (5). The MBCRB is usually easier to compute than the BCRB but it is also looser [20].

IV. The Off-line BCRB and its Modified Version for Dynamical Phase Estimation

In this section, we derive an analytical expression of the BCRB associated to an off-line carrier-phase-offset estimation. We proceed as follows. We first obtain analytical expressions of the two terms contributing to the BIM (8). Then, the expression of the diagonal elements of the inverse of the BIM are derived. In addition, we also give an analytical expression of the MBCRB.

A. Computation of $E_{\boldsymbol{\theta}}[\mathbf{F}(\boldsymbol{\theta})]$

$E_{\theta}[\mathbf{F}(\theta)]$ corresponds to the first term in the right-hand side of equation (8). This term evaluation requires the computation of the Fisher information matrix $\mathbf{F}(\boldsymbol{\theta})$, which in turn requires the evaluation of the Hessian of the log-likelihood function $\log p(\mathbf{y}|\boldsymbol{\theta})$. Using the observation model defined in section II, one has that the log-likelihood function can be expanded as

$$\log\, p(\mathbf{y}|\boldsymbol{\theta}) = \log \sum_{\mathbf{a}} p(\mathbf{y}|\mathbf{a},\boldsymbol{\theta})\, p(\mathbf{a}). \tag{10}$$

Using the whiteness of the noise and the independence of the transmitted symbols, one then obtains that

$$\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\mathbf{y}|\boldsymbol{\theta}\right) = \sum_{k=1}^{K} \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(y_k|\theta_k\right). \tag{11}$$

It is important to note that each term of the sum (11) is a matrix with only one non-zero element at most, namely,

$$\left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(y_k|\theta_k\right)\right]_{k,k} = \frac{\partial^2}{\partial\theta_k^2} \log p\left(y_k|\theta_k\right). \tag{12}$$

As a direct consequence, the Hessian $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\mathbf{y}|\boldsymbol{\theta}\right)$ is a diagonal matrix with the k^{th} diagonal element given by equation (12). Moreover, because of the circularity of the observation noise, the expectation of equation (12) with respect to $p\left(y_k|\theta_k\right)$ does not depend on θ_k. One then obtains

$$E_{\boldsymbol{\theta}}\left[\mathbf{F}\left(\boldsymbol{\theta}\right)\right] = J_D \mathbf{I}_K, \tag{13}$$

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

$$J_D \triangleq E_{\mathbf{y},\boldsymbol{\theta}}\left[-\frac{\partial^2 \log p\left(y_k|\theta_k\right)}{\partial\theta_k^2}\right]. \tag{14}$$

B. Computation of $E_{\boldsymbol{\theta}}[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log(p(\boldsymbol{\theta}))]$

Due to the Wiener structure of the phase model (2), $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\boldsymbol{\theta}\right)$ can be expanded as

$$\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\boldsymbol{\theta}\right) = \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_1\right) + \sum_{k=2}^{K} \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_k|\theta_{k-1}\right). \tag{15}$$

Let us detail the expressions of each term contributing to equation (15). The first term is a matrix with only one non-zero element, namely, the entry $(1,1)$ which is equal to

$$\left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_1\right)\right]_{1,1} = \frac{\partial^2 \log p\left(\theta_1\right)}{\partial\theta_1^2}. \tag{16}$$

The other terms in (15) are matrices with only four non-zero elements, namely, the entries $(k-1,k-1)$, $(k-1,k)$, $(k,k-1)$ and (k,k). Due to the Gaussian nature of the noise, one finds

$$\begin{aligned}\left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_k|\theta_{k-1}\right)\right]_{k,k} &= \left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_k|\theta_{k-1}\right)\right]_{k-1,k-1} \\ &= \frac{-1}{\sigma_w^2}, \end{aligned}\tag{17}$$

$$\begin{aligned}\left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_k|\theta_{k-1}\right)\right]_{k,k-1} &= \left[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p\left(\theta_k|\theta_{k-1}\right)\right]_{k-1,k} \\ &= \frac{1}{\sigma_w^2}. \end{aligned}\tag{18}$$

Taking (16), (17) and (18) into account, one finally obtains

$$-E_{\boldsymbol{\theta}}[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta})] = \begin{pmatrix} \frac{1}{\sigma_w^2} - E_{\theta_1}\left[\frac{\partial^2}{\partial\theta_1^2}\log p(\theta_1)\right] & -\frac{1}{\sigma_w^2} & 0 & \dots & 0 \\ -\frac{1}{\sigma_w^2} & \frac{2}{\sigma_w^2} & -\frac{1}{\sigma_w^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & -\frac{1}{\sigma_w^2} & \frac{2}{\sigma_w^2} & -\frac{1}{\sigma_w^2} \\ 0 & \dots & 0 & -\frac{1}{\sigma_w^2} & \frac{1}{\sigma_w^2} \end{pmatrix} \tag{19}$$

In the sequel, for the sake of conciseness, we will set $E_{\theta_1}\left[\frac{\partial^2 \log p(\theta_1)}{\partial\theta_1^2}\right] = 0$. This corresponds to the case of a non-informative prior about θ_1 (see [21]). The reasoning which is made in the following could however be extended to the general case, at the expense of some more difficult derivations and would not modify our results interpretation.

C. Analytical Expression of the Off-Line BCRB

In this subsection, an analytical expression of the diagonal elements of the inverse of the BIM is derived. From (13) and (19) one readily obtains that the BIM has a particular mathematical structure, *i.e.*,

$$\mathbf{B}_K = b \begin{pmatrix} A+1 & 1 & & & \\ 1 & A & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & A & 1 \\ & & & 1 & A+1 \end{pmatrix}, \tag{20}$$

where A and b are defined as $A \triangleq -\sigma_w^2 J_D - 2$ and $b \triangleq -1/\sigma_w^2$. In particular, $\mathbf{B}_K$ is a symmetric sparse matrix having

an (almost) *periodical* structure. Based on these observations, we show in Appendix I that the k^{th} diagonal element of $\mathbf{B}_K^{-1}$ can be expressed as

$$\left[\mathbf{B}_K^{-1}\right]_{k,k} = \frac{1}{|\mathbf{B}_K|}\Big[\rho_1^2\,(b+r_1)^2\,r_1^{K-3} + \rho_2^2(b+r_2)^2 r_2^{K-3} - \frac{b^2}{A-2}(r_1^{k-2}r_2^{K-k-1} + r_1^{K-k-1}r_2^{k-2})\Big], \tag{21}$$

where

$$r_1 \triangleq \frac{1}{\sigma_w^2} + \frac{J_D}{2}\left(1 - \sqrt{1 + \frac{4}{J_D\sigma_w^2}}\right), \tag{22}$$

$$r_2 \triangleq \frac{1}{\sigma_w^2} + \frac{J_D}{2}\left(1 + \sqrt{1 + \frac{4}{J_D\sigma_w^2}}\right), \tag{23}$$

$$\rho_1 \triangleq \frac{\sqrt{1+\frac{4}{\sigma_w^2 J_D}} - 1 - \frac{2}{\sigma_w^2 J_D}}{2\sqrt{1+\frac{4}{\sigma_w^2 J_D}}}, \tag{24}$$

$$\rho_2 \triangleq \frac{\sqrt{1+\frac{4}{\sigma_w^2 J_D}} + 1 + \frac{2}{\sigma_w^2 J_D}}{2\sqrt{1+\frac{4}{\sigma_w^2 J_D}}}, \tag{25}$$

Note that (21) together with (22)-(25) provides an analytical expression of the off-line BCRB associated to the estimation of θ_k. This function of index k depends only on three parameters: the number of observations K, the phase noise variance σ_w^2 and the observation noise variance σ_n^2 (or equivalently J_D, see equation (14)). The evaluation of J_D is discussed in Section V.

D. The Off-line MBCRB

We now consider the MBCRB, (see equation (9)). The second term in the right-hand side of (9) has already been computed in section IV-B. The first term $E_{\boldsymbol{\theta}}\left[\mathbf{G}(\boldsymbol{\theta})\right]$ requires the evaluation of $\mathbf{G}(\boldsymbol{\theta})$ which corresponds to the modified Fisher information matrix defined in (5). Using the observation model, one has that $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{y}|\mathbf{a},\boldsymbol{\theta})$ is a diagonal matrix where

$$[\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{y}|\mathbf{a},\boldsymbol{\theta})]_{k,k} = \frac{\partial^2}{\partial\theta_k^2}\log p(y_k|a_k,\theta_k). \tag{26}$$

Due to the Gaussian distribution of the noise, one further finds that

$$\frac{\partial^2}{\partial\theta_k^2}\log p(y_k|a_k,\theta_k) = \frac{2}{\sigma_n^2}\Re\left(-a_k^* y_k e^{-j\theta_k}\right). \tag{27}$$

Consequently, one obtains that

$$[\mathbf{G}(\boldsymbol{\theta})]_{k,k} = E_{\mathbf{a}}\,E_{\mathbf{y}|\mathbf{a},\boldsymbol{\theta}}\left[-\frac{2}{\sigma_n^2}\Re\left(-a_k^* y_k e^{-j\theta_k}\right)\right] \tag{28}$$

$$= E_{\mathbf{a}}[\frac{2}{\sigma_n^2}|a_k|^2] \tag{29}$$

$$= \frac{2}{\sigma_n^2}. \tag{30}$$

One can observe that $[\mathbf{G}(\boldsymbol{\theta})]_{k,k}^{-1} = \frac{1}{2}\left(\frac{E_s}{N_0}\right)^{-1}$, which corresponds to the MCRB (6) for $K = 1$ observation. As straightforwardly $E_{\boldsymbol{\theta}}\left[\mathbf{G}(\boldsymbol{\theta})\right] = 2/\sigma_n^2\,\mathbf{I}_K$, the modified BIM is obtained exactly like the BIM of subsection IV-C with $J_M \triangleq \frac{2}{\sigma_n^2}$ playing the role of J_D in equation (20). Note, however, that the MBCRB is usually looser than the BCRB.

V. Evaluation of J_D and asymptote-based lower bound

In this section we first point out that, in the general case, the evaluation of J_D implies to resort to numerical integration. We then derive the high-SNR and low-SNR asymptotes of the BCRB, and emphasize that their evaluation is easy. Finally, we show that these asymptotes are themselves lower bounds on the MSE. In particular, we illustrate that the combination of the low and high SNR asymptotes leads to an alternative (tight) lower bound, whose evaluation is straightforward.

A. Evaluation of J_D

In this part, we calculate an expression of J_D defined in (14). First, using the Gaussian nature of the noise and the equiprobability of the data symbols, one finds that

$$\log p(y_k|\theta_k) = \log\left(\frac{1}{\pi\sigma_n^2}e^{-\frac{1+|y_k|^2}{\sigma_n^2}}\cosh\left(\frac{2}{\sigma_n^2}\Re\big(y_k e^{-j\theta_k}\big)\right)\right). \tag{31}$$

Taking the second derivative of (31), one easily obtains that

$$\frac{\partial^2 \log p(y_k|\theta_k)}{\partial\theta_k^2} = -\frac{2}{\sigma_n^2}\Re(x_k)\tanh\Big(\frac{2}{\sigma_n^2}\Re(x_k)\Big) + \frac{4}{\sigma_n^4}\Im^2(x_k)\left[1 - \tanh^2\left(\frac{2}{\sigma_n^2}\Re(x_k)\right)\right], \tag{32}$$

where $x_k \triangleq y_k\,e^{-j\theta_k}$. In the general case, the expectation of (32) with respect to $p(\mathbf{y}|\boldsymbol{\theta})$ does not have any simple analytical solution. Hence, in practice, we have to resort to either numerical integration methods or some approximations. In the following, we present both the high-SNR and the low-SNR approximations of the BCRB.

B. High-SNR BCRB asymptote

In this part, we investigate the BCRB behavior at high SNR. From the definition of the BIM (8), only the first term, *i.e.*, $E_{\boldsymbol{\theta}}[\mathbf{F}(\boldsymbol{\theta})]$, depends on the SNR. Moreover, we have from (13) that $E_{\boldsymbol{\theta}}[\mathbf{F}(\boldsymbol{\theta})]$ is fully characterized by J_D (14). Hence, in the remainder of this section, we focus on the behavior of J_D.

At high SNR (*i.e.*, $\sigma_n^2 \to 0$), the $\tanh$-function in (32) can be approximated as

$$\tanh\left[\frac{2}{\sigma_n^2}\Re(x_k)\right] \approx \text{sign}\Big(\Re(x_k)\Big). \tag{33}$$

Hence, replacing $\tanh$ in (32) by its approximation (33) and using the definition of J_D (14), one has (see Appendix III)

$$J_D \approx E_{y_k,\theta_k}\left[-\frac{2}{\sigma_n^2}\Re(x_k)\text{sign}\Big(\Re(x_k)\Big)\right] \tag{34}$$

$$= \frac{2}{\sigma_n\sqrt{\pi}}e^{-\frac{1}{\sigma_n^2}} + \frac{2}{\sigma_n^2}\text{erf}\left(\frac{1}{\sigma_n}\right). \tag{35}$$

So, taking the limit of (35) for σ_n^2 tending to zero, one has that

$$J_{Dh} \triangleq \lim_{\sigma_n^2 \to 0} J_D = \frac{2}{\sigma_n^2}. \tag{36}$$

We see therefore from (30) and (36), that the high-SNR asymptote of the BCRB (*i.e.*, J_{Dh}^{-1}) is equal to the MBCRB. This corroborates the result derived by Moeneclaey [14] in the non-Bayesian case for a scalar parameter.

C. Low-SNR BCRB asymptote

We now consider the low-SNR asymptote of the BCRB in the NDA BPSK context. Following the same reasoning as before, we equivalently focus on the behavior of J_D for σ_n^2 tending to infinity.

From (32), using the fact that $\tanh(z) \approx z$ around $z = 0$, we obtain

$$\frac{\partial^2 \log p\left(y_k|\theta_k\right)}{\partial \theta_k^2} \approx -\left(\frac{2}{\sigma_n^2}\Re(x_k)\right)^2 + \left(\frac{2}{\sigma_n^2}\Im(x_k)\right)^2. \tag{37}$$

Plugging (37) into (14), one readily obtains that the BCRB low-SNR asymptote is

$$J_{Dl} \triangleq \lim_{\sigma_n^2 \to \infty} J_D = \frac{4}{\sigma_n^4}. \tag{38}$$

In [22], the authors obtain a result similar to (38) in a non-Bayesian scenario for a scalar parameter.

D. Asymptote-based Lower Bound

In this subsection, we show that the combination of the low and high-SNR BCRB asymptotes, presented in the previous subsections, still leads to a lower bound on the MSE. In particular, we define the "Asymptotic" BCRB (ABCRB) as

$$\mathrm{ABCRB} \triangleq \Big(\min(J_{Dl}, J_{Dh})\,\mathbf{I}_K + E_{\boldsymbol{\theta}}[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta})]\Big)^{-1}, \tag{39}$$

which can be evaluated with equations (19), (36) and (38). We now show that

$$\mathrm{MBCRB} \leq \mathrm{ABCRB} \leq \mathrm{BCRB}. \tag{40}$$

The proof of the first inequality directly comes from the definition of the ABCRB and of the MBCRB. The second inequality is straightforward to show when $J_{Dh} \leq J_{Dl}$ because in such a case, we have

$$\begin{aligned} \mathrm{ABCRB} &= \Big(J_{Dh}\,\mathbf{I}_K + E_{\boldsymbol{\theta}}[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta})]\Big)^{-1}, \\ &= \mathrm{MBCRB}. \end{aligned} \tag{41}$$

Since we know that $\mathrm{BCRB} \geq \mathrm{MBCRB}$ (see [20]), (40) is proven for $J_{Dh} \leq J_{Dl}$. The proof for $J_{Dl} < J_{Dh}$ is more lengthy and is detailed in Appendix IV. The ABCRB is thus an easy-to-evaluate lower bound on the MSE. This Bayesian bound which does not require any Monte-Carlo simulation is compared to the BCRB in Section VII.

This kind of reasoning only depends on the observation model: the ABCRB is relevant in the case of a BPSK transmission. However one can adapt this asymptote-based lower bound on any phase model system.

VI. The on-line BCRB : Sequential Cramér-Rao Bound

Up to this section, we have focused on the off-line scenario. We now show how the previous results can be used in the case of an on-line synchronization mode. In this mode, the receiver updates the observation vector $\mathbf{y}_1^k$ in order to estimate θ_k : only the past and the current observations are available.

In order to evaluate the performance of this kind of estimator, a Posterior (on-line) Cramér-Rao Bound was derived in [15]. Tichavsky *et al.* provide a method for updating the BIM from the time index k to the time index $k+1$. This method was already applied in [23] on the same model described by equations (1) and (2): the lower bound for the last state parameter θ_k is given by the following recursive sequence

$$C_{k+1} = \frac{\sigma_w^2 + C_k}{J_D \sigma_w^2 + 1 + J_D C_k} \quad \text{and} \quad C_1 = \frac{1}{J_D}. \tag{42}$$

On the other hand, using our previous derivations, we can provide an alternative expression of the on-line BCRB. Indeed, the on-line BCRB associated to observation vector $\mathbf{y}_1^K$ is clearly equal to entry (K, K) of the inverse of the BIM, *i.e.*, $\left[\mathbf{B}_K^{-1}\right]_{K,K}$. Using expression (21), we therefore have

$$\begin{aligned} C_K &= \left[\mathbf{B}_K^{-1}\right]_{K,K} \\ &= \frac{1}{|\mathbf{B}_K|}\Big[\rho_1^2\left(b + r_1\right)^2 r_1^{K-3} + \rho_2^2 (b + r_2)^2 r_2^{K-3} \\ &\quad - \frac{b^2}{A-2}(r_1^{K-2} r_2^{-1} + r_1^{-1} r_2^{K-2})\Big]. \end{aligned} \tag{43}$$

The behavior of the off-line (21) and of the on-line (43) BCRBs will be studied and compared in the next section.

In addition, it is shown in Appendix II that this bound decreases with time index K to the following limit:

$$\lim_{K \to \infty} \left[\mathbf{B}_K^{-1}\right]_{K,K} = \frac{-\sigma_w^2 + \sqrt{\sigma_w^4 + 4\frac{\sigma_w^2}{J_D}}}{2}. \tag{44}$$

Asymptotically with the observation size, the estimation performance is limited both by the phase noise and the observation noise (due to J_D). Therefore, unlike the non-Bayesian CRB [8], the BCRB does not necessarily tend to zero when the number of observations goes to infinity.

VII. Discussion

In this section we bring to the fore the behavior of the previous bounds, namely the off-line and the on-line BCRBs and ABCRB, by considering different scenarios.

We first consider a transmission disturbed by an AWGN with variance $\sigma_n^2 = 4$ and phase noise with variance $\sigma_w^2 = 0.16\,\mathrm{rad}^2$. Figure 1 superimposes versus time index, the on-line BCRB (see equation (43)), its asymptote described by equation (44), and the off-line BCRBs for different block-observation lengths K (see equation (21)). We then obtain the lower bound for each phase offset θ_k according to the considered scenario.

In the off-line context, one can see that the best phase estimate is achieved at mid-block, whereas the estimates are likely to be poorer at the block boarder (the proof of this property is detailed in Appendix I-D). This stems from the fact

that in the center position of the phase vector $\boldsymbol{\theta}$ one has more adjacent (past or future) and strongly correlated variables than at the boarder of the phase vector. Stated more precisely, from equation (2), the correlation coefficient between θ_k and θ_{k+l} (where l is a positive integer) is equal to $\sqrt{\frac{E_{\boldsymbol{\theta}}\left[\theta_1^2\right]+(k-1)\,\sigma_w^2}{E_{\boldsymbol{\theta}}\left[\theta_1^2\right]+(k+l-1)\,\sigma_w^2}}$, and thus even if a boarder phase is estimated with the same number of observations than a mid-size position phase, the correlation coefficient at the boarder are globally poorer than at the center and the information cannot be exploited as well.

Concerning the on-line bound, at the beginning when the number of observations increases, the estimator takes into account more and more information and the estimation is improved; the bound thus decreases and converges to its asymptote: the estimation performance is then limited by the phase noise and the observation noise independently of the number of observations taken into account.

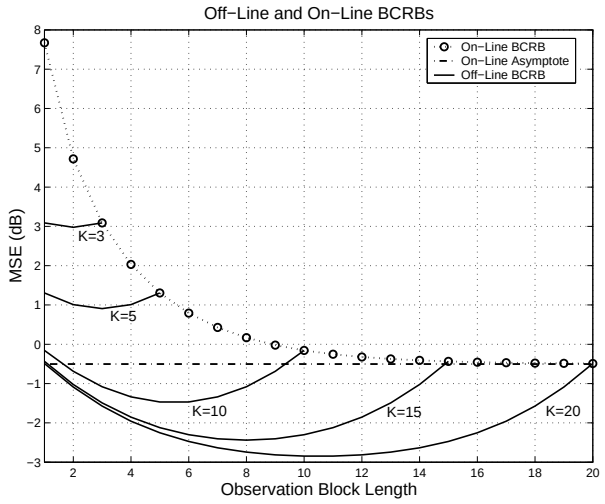

Fig. 1. BCRBs versus the number of observations for $\sigma_n^2 = 4$, $\sigma_w^2 = 0.16\,\mathrm{rad}^2$ (J_D is evaluated over 10^5 Monte-Carlo trials).

We now analyze the bound behavior versus the SNR over a block of $K = 20$ BPSK transmitted symbols. We illustrate some results for two distinct phase-noise variances: $\sigma_w^2 = 0.01\,\mathrm{rad}^2$ and $\sigma_w^2 = 0.16\,\mathrm{rad}^2$. Figure 2 superimposes for the two previous phase-noise variances, the BCRBs (see equation (21)) evaluated over 10^7 Monte-Carlo trials and the ABCRBs (see equation (39)) with $k = \lfloor \frac{K+1}{2}$ versus the SNR. One can recover on Figure 2 for $\sigma_w^2 = 0.16\,rad^2$ and SNR $= -6\,dB$, the minimum of the off-line BCRB displayed on Figure 1 for $K = 20$. We distinguish three SNR ranges.

- At high SNR, we notice that the ABCRBs in dashed lines and the corresponding BCRBs logically merge. Moreover, the ABCRB is close to the BCRB in the whole range of SNR for the two phase noise variances considered. At high SNR, both BCRBs converge to the deterministic MCRB evaluated for $K = 1$ observation [12], [13] *i.e.*, $MCRB = 1/J_{Dh}$. In this range of SNR, the information provided by each observation y_k is preponderant over the *a priori* knowledge of $\boldsymbol{\theta}$. Consequently in this case the observations are reliable enough to only take into account the present observation y_k in order to estimate θ_k and this is why the bounds converge to the well-known MCRB for any variance σ_w^2. As the *a priori* distribution of $\boldsymbol{\theta}$ has no influence, the Bayesian problem tends to a deterministic phase estimation problem where we estimate K independent phases θ_k with K independent observations.
- In mid-range SNR, when the phase variance σ_w^2 decreases, the BCRB gets closer to the deterministic MCRB evaluated for $K = 20$ observations. This is because in the limiting case for which the phase variance tends to zero, the time-varying phase estimation problem can be simplified to a deterministic constant-phase estimation problem. The MCRB is then a lower bound of the BCRBs for the same number of observations $K = 20$.
- At low SNR, the BCRB and the ABCRB coincide. In this SNR-range, the observation noise masks the *a priori* dynamic phase evolution. The Bayesian bound tends to the deterministic asymptote $\frac{1}{K\,J_{Dl}} = \frac{4}{K}\left(\frac{E_s}{N_0}\right)^{-1}$ given by [22]. One can note that the phase evolution is not stationary *i.e.*, $var\left(\theta_k\right) = var\left(\theta_{k-1}\right) + \sigma_w^2$, and consequently that uncertainty over the phase offset grows with time. Moreover as the phase model (2) is not restricted to a finite horizon, the MSE is not upper bounded. This is why there is no observed saturation as one could expect with a traditional Bayesian bound.

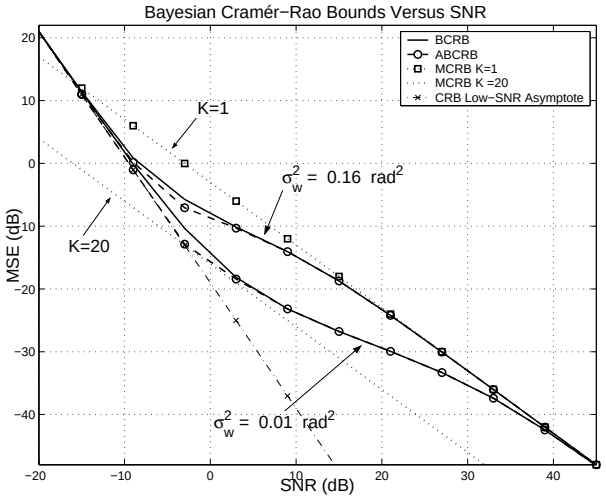

Fig. 2. Comparison between Bayesian CRBs and Asymptote-based CRBs for two different phase noise variances $\sigma_w^2 = 0.01\,\mathrm{rad}^2$ and $\sigma_w^2 = 0.16\,\mathrm{rad}^2$.

VIII. Conclusion

In this contribution, we have derived an analytical expression of a Bayesian Cramér-Rao Bound for the estimation of a realistic dynamical phase. We provide a closed-form expression in the case of a BPSK transmission disturbed by AWGN in a NDA context. The reasoning can be extended to any M-ary Phase Shift Keying transmissions. Furthermore, we can readily predict the bound behavior from its low and high-SNR asymptotes. We combine these asymptotes and we give a new Bayesian bound which is easy to evaluate at the price of being slightly lower than the BCRB. These bounds are useful when analyzing the performance of actual phase-

tracking estimators in stressing environments for both the off-line and on-line scenarios.

APPENDIX I
EXPRESSION OF DIAGONAL ELEMENTS OF THE BCRB

In this appendix, we detail the calculus to obtain the diagonal elements (21) of the inverse of the BIM (20). We first use the classical matrix-inversion formula

$$[\mathbf{B}_K^{-1}]_{k,k} = \frac{C_{k,k}}{|\mathbf{B}_K|}, \tag{45}$$

where $C_{k,k}$ is the cofactor of the element $[\mathbf{B}_K]_{k,k}$ and $|\mathbf{B}_K|$ is the determinant of matrix $\mathbf{B}_K$. In the sequel, we focus on the derivation of particular expressions of $C_{k,k}$ and $|\mathbf{B}_K|$. We first present some preliminary calculus (sections I-A to I-C) aiming at simplifying the expansion of $|\mathbf{B}_K|$ and $C_{k,k}$. Then, in section I-D, we derive a closed-form expression of the diagonal elements of $\mathbf{B}_K^{-1}$.

A. Preliminary calculus : Determinant d_k

In this subsection, we derive an analytical expression of the determinant d_k of a $k \times k$ matrix D_k defined as

$$D_k = b \begin{pmatrix} A & 1 & & & \\ 1 & A & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & A & 1 \\ & & & 1 & A \end{pmatrix} \tag{46}$$

Expanding d_k along the first and the second column, one obtains the recursive equation

$$d_k = Abd_{k-1} - b^2 d_{k-2} \text{ with } d_0 = 1 \text{ and } d_1 = bA. \tag{47}$$

$\{d_k\}_{k=0}^{\infty}$ is thus a linear recurrent sequence and its characteristic polynomial $p(x) = x^2 - Abx + b^2$ has real roots r_1 and r_2. Then, using the initial terms, an analytical expression of d_k is given by

$$d_k = \rho_1 (r_1)^k + \rho_2 (r_2)^k \qquad \text{for } k \in \mathbb{N}, \tag{48}$$

where

$$r_1 = \frac{b}{2}\left(A + \sqrt{A^2-4}\right), \qquad \rho_1 = \frac{\sqrt{A^2-4}+A}{2\sqrt{A^2-4}},$$
$$r_2 = \frac{b}{2}\left(A - \sqrt{A^2-4}\right), \qquad \rho_2 = \frac{\sqrt{A^2-4}-A}{2\sqrt{A^2-4}}.$$

B. Determinant $|\mathbf{B}_K|$

In this subsection, we derive an expression of $|\mathbf{B}_K|$ by using the preliminary result derived Appendix I-A. In a first step, expanding this determinant along the first column, we obtain the sum of two cofactors only. Then, expanding in turn these cofactors along the last column and using (47), we have

$$|\mathbf{B}_K| = (A+2)\, b\, d_{K-1}, \tag{49}$$

or with (48),

$$|\mathbf{B}_K| = (A+2)\, b \left(\rho_1 (r_1)^{K-1} + \rho_2 (r_2)^{K-1}\right). \tag{50}$$

C. Cofactor $C_{k,k}$

In order to calculate the cofactor $C_{k,k}$, one has to delete the k^{th} row and the k^{th} column of the matrix $\mathbf{B}_K$ and one obtains a two-block diagonal matrix. The upper (respectively lower) block is noted U_{B_K} (respectively L_{B_K}). The cofactor which is obtained from the determinant of the previous matrix is thus the product of two determinants:

$$C_{k,k} = det(U_{B_K}) det(L_{B_K}) \tag{51}$$
$$= (bd_{k-2} + d_{k-1})\,(bd_{K-k-1} + d_{K-k},)\,, \tag{52}$$

where d_k is calculated in Appendix I-A.

D. Expression of the diagonal elements of $\mathbf{B}_K^{-1}$

An analytical expression of the diagonal elements of the inverse of (20) is now derived. Then, the behavior of these elements along their index k and the matrix size $K \times K$ is studied.

Rewriting (45) and using (52), one has

$$\left[\mathbf{B}_K^{-1}\right]_{k,k} = \frac{(bd_{k-2} + d_{k-1})\,(bd_{K-k-1} + d_{K-k})}{|\mathbf{B}_K|}. \tag{53}$$

Then, using (48), $\left[\mathbf{B}_K^{-1}\right]_{k,k}$ can be expanded as follows

$$\left[\mathbf{B}_K^{-1}\right]_{k,k} = \frac{1}{|\mathbf{B}_K|}\Big[\rho_1^2 (b+r_1)^2 r_1^{K-3} + \rho_2^2 (b+r_2)^2 r_2^{K-3} - \frac{b^2}{A-2}(r_1^{k-2} r_2^{K-k-1} + r_1^{K-k-1} r_2^{k-2})\Big],$$

which is equivalent to (21).

For a fixed block length K from (21), note that $\left[\mathbf{B}_K^{-1}\right]_{k,k}$ depends on the index k through the expression $g_K(k)$ where

$$g_K(x) = r_1^{x-2} r_2^{K-x-1} + r_1^{K-x-1} r_2^{x-2}. \tag{54}$$

After analyzing this function, by classically studying the sign of $\frac{dg_K}{dx}$, one readily finds that

- the symmetry of $g_K(x)$ with respect to $\frac{K+1}{2}$ implies that
$$\left[\mathbf{B}_K^{-1}\right]_{k,k} = \left[\mathbf{B}_K^{-1}\right]_{l,l} \quad \text{with} \quad l = K+1-k,$$
- the minimal diagonal element is the mid-coefficient $\left[\mathbf{B}_K^{-1}\right]_{\frac{K+1}{2},\frac{K+1}{2}}$,
- the maximal diagonal elements are $\left[\mathbf{B}_K^{-1}\right]_{1,1}$ and $\left[\mathbf{B}_K^{-1}\right]_{K,K}$.

APPENDIX II
DERIVATION OF EXPRESSION (44)

Using (53), $\left[\mathbf{B}_K^{-1}\right]_{K,K}$ can be written as follows:

$$\left[\mathbf{B}_K^{-1}\right]_{K,K} = \frac{A+1}{b(A+2)} - \frac{1}{b^2(A+2)}\frac{d_K}{d_{K-1}}, \tag{55}$$

where d_K is defined in Appendix I-A. Since $\left[\mathbf{B}_K^{-1}\right]_{K,K}$ is only a function of $\frac{d_K}{d_{K-1}}$, we consider the fixed points of a sequence $\{u_n\}_{n=0}^{\infty}$ defined as

$$u_n \triangleq \frac{d_n}{bd_{n-1}} = \frac{Au_{n-1} - 1}{u_{n-1}}, \tag{56}$$

with $u_1 = A$. Clearly, this sequence is strictly increasing and converges to $u_\infty = \frac{1}{2}\left(A - \sqrt{A^2-4}\right)$. Combining this result with (55), we have that

$$\left[\mathbf{B}_K^{-1}\right]_{K,K} = \frac{A+1}{b\,(A+2)} - \frac{1}{b\,(A+2)} u_K, \tag{57}$$

is a strictly decreasing sequence with the following limit:

$$\lim_{K\to\infty}\left[\mathbf{B}_K^{-1}\right]_{K,K} = \frac{A+1}{b\,(A+2)} - \frac{\left(A-\sqrt{A^2-4}\right)}{2b\,(A+2)}. \tag{58}$$

Using the definitions of A and b in subsection IV-C, one obtains

$$\lim_{K\to\infty}\left[\mathbf{B}_K^{-1}\right]_{K,K} = \frac{-\sigma_w^2 + \sqrt{\sigma_w^4 + 4\frac{\sigma_w^2}{J_D}}}{2}. \tag{59}$$

Appendix III
High-SNR asymptote of J_D

Defining $F_H(\theta_k) \triangleq E_{y_k|\theta_k}\left[\frac{2}{\sigma_n^2}\Re(x_k)\text{sign}\Big(\Re(x_k)\Big)\right]$, (34) becomes

$$J_D \approx E_{\theta_k}\left[F_H(\boldsymbol{\theta})\right]. \tag{60}$$

Since the noise affecting the observation is circular, $x_k = a_k + \tilde{n}_k$ where $\tilde{n}_k = n_k e^{-j\theta_k}$ is a circular Gaussian noise with variance σ_n^2, therefore, we find

$$F_H(\theta_k) = E_{a_k,\tilde{n}_k^I|\theta_k}\left[\frac{2}{\sigma_n^2}\left(a_k+\tilde{n}_k^I\right)\text{sign}\Big(\left(a_k+\tilde{n}_k^I\right)\Big)\right],$$

where $\tilde{n}_k^I \triangleq \Re\{\tilde{n}_k\}$. Using first the definition of the expectation and then the definition of the error function erf$(\cdot)$

$$\begin{aligned} F_H(\theta_k) &= \sum_{a_k\in\{-1,1\}}\frac{1}{\sigma_n^2}\int_{-\infty}^{+\infty}\left(a_k+\tilde{n}_k^I\right) \\ &\qquad\qquad \times \text{sign}\left(a_k+\tilde{n}_k^I\right)p(\tilde{n}_k^I)d\tilde{n}_k^I \\ &= \frac{1}{\sigma_n^2}\left(2\int_{-1}^{+1}p(\tilde{n}_k^I)d\tilde{n}_k^I + 4\int_1^{+\infty}\tilde{n}_k^I p(\tilde{n}_k^I)d\tilde{n}_k^I\right), \\ &= \frac{2}{\sigma_n\sqrt{\pi}}e^{-\frac{1}{\sigma_n^2}} + \frac{2}{\sigma_n^2}\text{erf}\left(\frac{1}{\sigma_n}\right). \end{aligned} \tag{61}$$

Since (61) does not depend on θ_k, we finally obtain

$$J_D \approx E_{\theta_k}[F_H(\theta_k)] = \frac{2}{\sigma_n\sqrt{\pi}}e^{-\frac{1}{\sigma_n^2}} + \frac{2}{\sigma_n^2}\text{erf}\left(\frac{1}{\sigma_n}\right). \tag{62}$$

Appendix IV
Proof that BCRB $>$ ABCRB for $J_{Dl} < J_{Dh}$

To prove (40), we show that

$$J_D - J_{Dl} < 0, \tag{63}$$

Using the definition of J_D and J_{Dl}, *i.e.*, (14) and (38), we have

$$\begin{aligned} J_D - J_{Dl} =& E_{y_k,\theta_k}\left[\frac{4}{\sigma_n^4}\Im^2(x_k)\tanh^2\left(\frac{2}{\sigma_n^2}\Re(x_k)\right)\right. \\ &\left. + \frac{2}{\sigma_n^2}\Re(x_k)\left(\tanh\Big(\frac{2}{\sigma_n^2}\Re(x_k)\Big) - \frac{2}{\sigma_n^2}\Re(x_k)\right)\right]. \end{aligned} \tag{64}$$

For the sake of conciseness, we use the following notations: $X \triangleq \Re(x_k)$ and $Y \triangleq \Im(x_k)$. Taking then into account that the expectation in (64) is independent of θ_k, we have

$$\begin{aligned} J_D - J_{Dl} = E_{X,Y}&\left[\frac{4}{\sigma_n^4}Y^2\tanh^2\left(\frac{2}{\sigma_n^2}X\right)\right. \\ &\left. + \frac{2}{\sigma_n^2}X\left(\tanh\Big(\frac{2}{\sigma_n^2}X\Big) - \frac{2}{\sigma_n^2}X\right)\right]. \end{aligned} \tag{65}$$

Since the noise affecting the observations is circular, X and Y are independent. Moreover Y is a zero-mean random Gaussian variable with variance $\sigma_n^2/2$. Hence,

$$J_D - J_{Dl} = E_X\left[f(X)\right], \tag{66}$$

where $f(X) \triangleq \frac{2}{\sigma_n^2}\left[X\tanh\Big(\frac{2}{\sigma_n^2}X\Big) - \frac{2}{\sigma_n^2}X^2 + \tanh^2\left(\frac{2}{\sigma_n^2}X\right)\right]$ and where X is a random variable with the following probability density function (pdf)

$$p(X) = \frac{1}{\sigma_n\sqrt{\pi}}e^{\frac{(X-1)^2}{\sigma_n^2}}\,p\,(a=+1) + \frac{1}{\sigma_n\sqrt{\pi}}e^{\frac{(X+1)^2}{\sigma_n^2}}\,p\,(a=-1).$$

Then we rewrite (66) as

$$J_D - J_{Dl} = \frac{1}{2}\left[E_{X|a=+1}\left[f\,(X)\right] + E_{X|a=-1}\left[f\,(X)\right]\right]. \tag{67}$$

Noting that $f(X) = f(-X)$, we deduce from a change of variable ($X = -\tilde{X}$) that the two terms in the right-hand side of (67) are equal. This equation thus becomes

$$J_D - J_{Dl} = E_{X|a=+1}\left[f\,(X)\right]. \tag{68}$$

We now prove that $E_{X|a=+1}\left[f\,(X)\right]$ is negative. We use Stein's lemma [24]: suppose Z is a normally distributed random variable with expectation μ and variance v. Furthermore suppose g is a function for which the two expectations $E[(Z-\mu)g(Z)]$ and $E[g'(Z)]$ both exist. Then

$$vE_Z\left[g'(Z)\right] = E_Z\left[(Z-\mu)g(Z)\right]. \tag{69}$$

Applying this lemma to the random variable X with expectation $\mu = 1$ and variance $\sigma_n^2/2$ and to $g(X) = \tanh\left(\frac{2}{\sigma_n^2}X\right)$, we obtain

$$\begin{aligned} \frac{\sigma_n^2}{2}E_{X|a=+1}&\left[\frac{2}{\sigma_n^2}\left(1-\tanh^2\left(\frac{2}{\sigma_n^2}X\right)\right)\right] \\ &= E_{X|a=+1}\left[(X-1)\tanh\left(\frac{2}{\sigma_n^2}X\right)\right]. \end{aligned} \tag{70}$$

Moreover, we directly have

$$E_{X|a=+1}\left[\frac{4}{\sigma_n^4}X^2\right] = \frac{2}{\sigma_n^2} + \frac{4}{\sigma_n^4}. \tag{71}$$

Then, plugging equations (70) and (71) into equation (68)

$$J_D - J_{Dl} = E_{X|a=+1}\left[\frac{2}{\sigma_n^2}\tanh\left(\frac{2}{\sigma_n^2}X\right)\right] - \frac{4}{\sigma_n^4}. \tag{72}$$

Considering that X is a Gaussian variable with expectation $\mu = 1$, we now show that $E_{X|a=+1}\left[\tanh\left(\frac{2}{\sigma_n^2}X\right) - \frac{2}{\sigma_n^2}X\right] < 0$.

Suppose Z is a normally distributed random variable with expectation $\mu = 1$ and variance v. Denoting its pdf by $p(Z)$, we have for any $\lambda > 0$

$$E_Z\left[\tanh\left(\lambda Z\right) - \lambda Z\right] = \int_{-\infty}^{\infty} \left[\tanh\left(\lambda Z\right) - \lambda Z\right] p(Z) dZ \quad (73)$$

$$= \int_{0}^{\infty} \left[\tanh\left(\lambda Z\right) - \lambda Z\right] \left(p(Z) - p(-Z)\right) dZ. \quad (74)$$

As $p(Z) - p(-Z) > 0$ for any $Z > 0$, and as $\tanh(\lambda Z) < \lambda Z$ for any $Z > 0$ and $\lambda > 0$, we actually have $E_Z\left[\tanh\left(\lambda Z\right) - \lambda Z\right] < 0$, so that

$$E_Z\left[\tanh\left(\lambda Z\right)\right] < \lambda, \qquad \lambda > 0. \quad (75)$$

As X is a normally distributed random variable with expectation $\mu = 1$, applying this result with $\lambda = \frac{2}{\sigma_n^2}$, we obtain

$$E_{X|a=+1}\left[\frac{2}{\sigma_n^2}\tanh\left(\frac{2}{\sigma_n^2}X\right)\right] - \frac{4}{\sigma_n^4} < 0, \quad (76)$$

and finally from equations (72) and (76) we have proved equation (63)

$$J_D - J_{Dl} < 0, \qquad \sigma_n > 0. \quad (77)$$

REFERENCES

[1] R. Gallager, "Low-Density Parity-Check Codes," *IEEE Trans. Inform. Theory*, vol. 8, pp. 21–28, Jan. 1962.

[2] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261–1272, Oct. 1996.

[3] R. Pyndiah, "Near-optimum decoding of product codes: block turbo codes," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1003–1010, Aug. 1998.

[4] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York, NY, USA: John Wiley & Sons, Inc., 1997.

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

[6] H. Cramér, *Mathematical Methods of Statistics*. New York: Princeton University, Press, 1946.

[7] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. 20, pp. 591–598, 1974.

[8] W. G. Cowley, "Phase and frequency estimation for PSK packets: bounds and algorithms," *IEEE Trans. Commun.*, vol. 44, pp. 26–28, Jan. 1996.

[9] N. Noels, H. Steendam, and M. Moeneclaey, "The Cramér-Rao bound for phase estimation from coded linearly modulated signals," *IEEE Commun. Lett.*, vol. 7, pp. 207–209, May 2003.

[10] Y. Jiang, F. Sun, and J. Baras, "On the true Cramér-Rao lower bound for the DA joint estimation of carrier phase and timing offsets," in *IEEE International Conference on Communications*, New Orleans, USA, June 2000, pp. 331–335.

[11] B. M. F. Rice, B. Cowley and M. Rice, "Cramér-Rao Lower Bounds for QAM Phase and Frequency Estimation," *IEEE Trans. Commun.*, vol. 49, pp. 1582–1591, Sept. 2001.

[12] A. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramér-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, pp. 1391–1399, Fev./Mar./Apr 1994.

[13] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramér-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, pp. 52–60, Jan. 1998.

[14] 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. Commun.*, vol. 46, pp. 1536–1544, Nov. 1998.

[15] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Processing*, vol. 46, pp. 1386–1396, May 1998.

[16] J. Dauwels and S. Korl, "A numerical method to compute Cramér-Rao-type bounds for challenging estimation problems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toulouse, Fr, May 2006, pp. 717–720.

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

[18] 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.

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

[20] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér-Rao bounds," *Ann. Statistics*, vol. 15, pp. 1421–1438, 1987.

[21] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[22] H. Steendam and M. Moeneclaey, "Low-SNR limit of the Cramér-Rao bound for estimating the time delay of a PSK, QAM, or PAM waveform," *IEEE Commun. Lett.*, vol. 5, pp. 31–33, Jan. 2001.

[23] P. Amblard, J. 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.

[24] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.