Frequentist and Bayesian Quantum Phase Estimation

Yan Li [1,†] ![ORCID], **Luca Pezzè** [2,†], **Manuel Gessner** [2,†] ![ORCID], **Zhihong Ren** [1], **Weidong Li** [1,*]
and Augusto Smerzi [2]

[1] Institute of Theoretical Physics and Department of Physics, State Key Laboratory of Quantum Optics and
Quantum Optics Devices, Collaborative Innovation Center of Extreme Optics, Shanxi University,
Taiyuan 030006, China; li8989971@163.com (Y.L.); ren_zhihong1@163.com (Z.R.)
[2] QSTAR, INO-CNR and LENS, Largo Enrico Fermi 2, 50125 Firenze, Italy; luca.pezze@ino.it (L.P.);
manuel.gessner@ino.it (M.G.); augusto.smerzi@ino.it (A.S.)
* Correspondence: wdli@sxu.edu.cn
† These authors contributed equally to this work.

✓ check for updates

Abstract: Frequentist and Bayesian phase estimation strategies lead to conceptually different results on the state of knowledge about the true value of an unknown parameter. We compare the two frameworks and their sensitivity bounds to the estimation of an interferometric phase shift limited by quantum noise, considering both the cases of a fixed and a fluctuating parameter. We point out that frequentist precision bounds, such as the Cramér–Rao bound, for instance, do not apply to Bayesian strategies and vice versa. In particular, we show that the Bayesian variance can overcome the frequentist Cramér–Rao bound, which appears to be a paradoxical result if the conceptual difference between the two approaches are overlooked. Similarly, bounds for fluctuating parameters make no statement about the estimation of a fixed parameter.

1. Introduction

The estimation of a phase shift using interferometric techniques is at the core of metrology and sensing [1–3]. Applications range from the definition of the standard of time [4] to the detection of gravitational waves [5,6]. The general problem can be concisely stated as the search for optimal strategies to minimize the phase estimation uncertainty. The noise that limits the achievable phase sensitivity can have a "classical" or a "quantum" nature. Classical noise originates from the coupling of the interferometer with some external source of disturbance, like seismic vibrations, parasitic magnetic fields or from incoherent interactions within the interferometer. Such noise can, in principle, be arbitrarily reduced, e.g., by shielding the interferometer from external noise or by tuning interaction parameters to ensure a fully coherent time evolution. The second source of uncertainty has an irreducible quantum origin [7,8]. Quantum noise cannot be fully suppressed, even in the idealized case of the creation and manipulation of pure quantum states. Using classically-correlated probe states, it is possible to reach the so-called shot noise or standard quantum limit, which is the limiting factor for the current generation of interferometers and sensors [9–12]. Strategies involving probe states characterized by squeezed quadratures [13] or entanglement between particles [14–19] are able to overcome the shot noise, the ultimate quantum bound being the so-called Heisenberg limit. Quantum noise reduction in phase estimation has been demonstrated in several proof-of-principle experiments with atoms and photons [20,21].

There is a vast amount of literature dealing with the parameter estimation problem that has been mostly developed following two different approaches [22–24]: frequentist and Bayesian. Both approaches have been investigated in the context of quantum phase estimation [18,20,25–31] and

implemented/tested experimentally [32–36]. They build on conceptually different meanings attached to the word "probability" and their respective results provide conceptually different information on the estimated parameters and their uncertainties.

In the limit of a large number of repeated measurements, the sensitivity reached by the frequentist and Bayesian methods generally agree: this fact has very often induced the belief that the two paradigms can be interchangeably used in the phase estimation theory without acknowledging their irreconcilable nature. Overlooking these differences is not only conceptually inconsistent but can even create paradoxes, as, for instance, the existence of ultimate bounds in sensitivity proven in one paradigm that can be violated in the other.

In this manuscript, we directly compare the frequentist and the Bayesian parameter estimation theory. We study different sensitivity bounds obtained in the two frameworks and highlight the conceptual differences between the two. Besides the asymptotic regime of many repeated measurements, we also study bounds that are relevant for small samples. In particular, we show that the Bayesian variance can overcome the frequentist Cramér–Rao bound. The Cramér–Rao bound is a mathematical theorem providing the highest possible sensitivity in a phase estimation problem. The fact that the Bayesian sensitivity can be higher than the Cramér–Rao bound is therefore paradoxical. The paradox is solved by clarifying the conceptual differences between the frequentist and the Bayesian approaches, which therefore cannot be directly compared. Such difference should be considered when discussing theoretical and experimental figures of merit in interferometric phase estimation.

Our results are illustrated with a simple test model [37,38]. We consider N qubits with basis states $|0\rangle$ and $|1\rangle$, initially prepared in a (generalized) GHZ state $|\text{GHZ}\rangle = (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})/\sqrt{2}$, with all particles being either in $|1\rangle$ or in $|0\rangle$. The phase-encoding is a rotation of each qubit in the Bloch sphere $|0\rangle \to e^{-i\theta/2}|0\rangle$ and $|1\rangle \to e^{+i\theta/2}|1\rangle$, which transforms the $|\text{GHZ}\rangle$ state into $|\text{GHZ}(\theta)\rangle = (e^{-iN\theta/2}|0\rangle^{\otimes N} + e^{+iN\theta/2}|1\rangle^{\otimes N})/\sqrt{2}$. The phase is estimated by measuring the parity $(-1)^{N_0}$, where N_0 is the number of particles in the state $|0\rangle$ [37,39–41]. The parity measurement has two possible results $\mu = \pm 1$ that are conditioned by the "true value of the phase shift" θ_0 with probability $p(\pm 1|\theta_0) = (1 \pm \cos(N\theta_0))/2$. The probability to observe the sequence of results $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_m\}$ in m independent repetitions of the experiment (with same probe state and phase encoding transformation) is

$$p(\boldsymbol{\mu}|\theta_0) = \prod_{i=1}^{m} p(\mu_i|\theta_0) = \left(\frac{1 + \cos(N\theta_0)}{2}\right)^{m_+} \left(\frac{1 - \cos(N\theta_0)}{2}\right)^{m_-}, \tag{1}$$

where $m_\pm$ is the number of the observed results ± 1, respectively. Notice that $p(\boldsymbol{\mu}|\theta_0)$ is the conditional probability for the measurement outcome $\boldsymbol{\mu}$, given that the true value of the phase shift is θ_0 (which we consider to be unknown in the estimation protocol). Equation (1) provides the probability that will be used in the following sections for the case $N = 2$ and $\theta_0 \in [0, \pi/2]$. Sections 2 and 3 deal with the case where θ_0 has a fixed value and in Section 4 we discuss precision bounds for a fluctuating phase shift.

2. Frequentist Approach

In the frequentist paradigm, the phase (assumed having a fixed but unknown value θ_0) is estimated via an arbitrarily chosen function of the measurement results, $\theta_{\text{est}}(\boldsymbol{\mu})$, called the estimator. Typically, $\theta_{\text{est}}(\boldsymbol{\mu})$ is chosen by maximizing the likelihood of the observed data (see below). The estimator, being a function of random outcomes, is itself a random variable. It is characterized by a statistical distribution that has an objective, measurable character. The relative frequency with which the event θ_{est} occurs converges to a probability asymptotically with the number of repeated experimental trials.

2.1. Frequentist Risk Functions

Statistical fluctuations of the data reflect the *statistical uncertainty* of the estimation. This is quantified by the variance,

$$\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} = \sum_{\boldsymbol{\mu}} \left(\theta_{\text{est}}(\boldsymbol{\mu}) - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2 p(\boldsymbol{\mu}|\theta_0), \tag{2}$$

around the mean value $\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0} = \sum_{\boldsymbol{\mu}} \theta_{\text{est}}(\boldsymbol{\mu})p(\boldsymbol{\mu}|\theta_0)$, the sum extending over all possible measurement sequences (for fixed θ_0 and m). An important class is that of *locally unbiased* estimators, namely those satisfying $\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0} = \theta_0$ and $\frac{d\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta}}{d\theta}\big|_{\theta=\theta_0} = 1$ (see, for instance, [42]). An estimator is unbiased if and only if it is locally unbiased at every θ_0.

The quality of the estimator can also be quantified by the mean square error (MSE) [23]

$$\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0} = \sum_{\boldsymbol{\mu}} \left(\theta_{\text{est}}(\boldsymbol{\mu}) - \theta_0\right)^2 p(\boldsymbol{\mu}|\theta_0), \tag{3}$$

giving the deviation of θ_{est} from the true value of the phase shift θ_0. It is related to Equation (2) by the relation

$$\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0} = \left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} + \left(\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0} - \theta_0\right)^2. \tag{4}$$

In the frequentist approach, often the variance is not considered as a proper way to quantify the goodness of an estimator. For instance, an estimator that always gives the same value independently of the measurement outcomes is strongly biased: it has zero variance but a large MSE that does not scale with the number of repeated measurements. Notice that the MSE cannot be accessed from the experimentally available data since the true value θ_0 is unknown. In this sense, only the fluctuations of θ_{est} around its mean value, i.e., the variance $(\Delta^2\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0}$, have experimental relevance. For unbiased estimators, Equations (2) and (4) coincide. In general, since the bias term in Equation (4) is never negative, $\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0} \geq (\Delta^2\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0}$ and any lower bound on $(\Delta^2\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0}$ automatically provides a lower bound on $\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0}$ but not vice versa. In the following section, we therefore limit our attention to bounds on $(\Delta^2\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0}$. The distinction between the two quantities becomes more important in the case of a fluctuating phase shift θ_0, where the bias can affect the corresponding bounds in different ways. We will see this explicitly in Section 4.

2.2. Frequentist Bounds on Phase Sensitivity

2.2.1. Barankin Bound

The Barankin bound (BB) provides the tightest lower bound to the variance (2) [43]. It can be proven to be always (for any m) saturable, in principle, by a specific local (i.e., dependent of θ_0) estimator and measurement observable. Of course, since the estimator that saturates the BB depends on the true value of the parameter (which is unknown), the bound is of not much use in practice. Nevertheless, the BB plays a central role, from the theoretical point of view, as it provides a hierarchy of weaker bounds which can be used in practice with estimators that are asymptotically unbiased. The BB can be written as [44]

$$\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\text{BB}} \equiv \sup_{\theta_i, a_i, n} \frac{\left\{\sum_{i=1}^{n} a_i\left[\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_i} - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right]\right\}^2}{\sum_{\boldsymbol{\mu}} \left[\sum_{i=1}^{n} a_i L(\boldsymbol{\mu}|\theta_i, \theta_0)\right]^2 p(\boldsymbol{\mu}|\theta_0)}, \tag{5}$$

where $L(\boldsymbol{\mu}|\theta_i, \theta) = p(\boldsymbol{\mu}|\theta_i)/p(\boldsymbol{\mu}|\theta)$ is generally indicated as likelihood ratio and the supremum is taken over n parameters $a_i \in \mathbb{R}$, which are arbitrary real numbers, and θ_i, which are arbitrary phase values

in the parameter domain. For unbiased estimators, we can replace $\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_i} = \theta_i$ for all i and the BB becomes independent of the estimator:

$$\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\text{BB}}^{\text{ub}} \equiv \sup_{\theta_i,a_i,n} \frac{\left\{\sum_{i=1}^n a_i[\theta_i - \theta_0]\right\}^2}{\sum_{\boldsymbol{\mu}}\left[\sum_{i=1}^n a_i L(\boldsymbol{\mu}|\theta_i,\theta_0)\right]^2 p(\boldsymbol{\mu}|\theta_0)}. \tag{6}$$

A derivation of the BB is presented in Appendix A.

The explicit calculation of $\Delta^2\theta_{\text{BB}}$ is impractical in most applications due to the number of free variables that must be optimized. However, the BB provides a strict hierarchy of bounds of increasing complexity that can be of great practical importance. Restricting the number of variables in the optimization can provide local lower bounds that are much simpler to determine at the expense of not being saturable in general, namely, for an arbitrary number of measurements. Below, we demonstrate the following hierarchy of bounds:

$$\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\text{BB}} \geq \Delta^2\theta_{\text{EChRB}} \geq \Delta^2\theta_{\text{ChRB}} \geq \Delta^2\theta_{\text{CRLB}}, \tag{7}$$

where $\Delta^2\theta_{\text{CRLB}}$ is the Cramér–Rao lower bound (CRLB) [45,46] and $\Delta^2\theta_{\text{ChRB}}$ is the Hammersley–Chapman–Robbins bound (ChRB) [47,48]. We will also introduce a novel extended version of the ChRB, indicated as $\Delta^2\theta_{\text{EChRB}}$.

2.2.2. Cramér–Rao Lower Bound and Maximum Likelihood Estimator

The CRLB is the most common frequentist bound in parameter estimation. It is given by [45,46]:

$$\Delta^2\theta_{\text{CRLB}} = \frac{\left(\frac{d\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}}{d\theta_0}\right)^2}{mF(\theta_0)}. \tag{8}$$

The inequality $\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\text{CRLB}}$ is obtained by differentiating $\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}$ with respect to θ_0 and using a Cauchy–Schwarz inequality:

$$\left(\frac{d\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}}{d\theta_0}\right)^2 = \left(\sum_{\boldsymbol{\mu}}\left(\theta_{\text{est}}(\boldsymbol{\mu}) - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)\frac{dp(\boldsymbol{\mu}|\theta_0)}{d\theta_0}\right)^2 \leq mF(\theta_0)\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0}, \tag{9}$$

where we have used $\sum_{\boldsymbol{\mu}}\frac{dp(\boldsymbol{\mu}|\theta_0)}{d\theta_0} = 0$ and $\sum_{\boldsymbol{\mu}}\frac{1}{p(\boldsymbol{\mu}|\theta_0)}\left(\frac{\partial p(\boldsymbol{\mu}|\theta)}{\partial\theta}|_{\theta_0}\right)^2 = m\sum_{\mu}\frac{1}{p(\mu|\theta_0)}\left(\frac{\partial p(\mu|\theta)}{\partial\theta}|_{\theta_0}\right)^2$ valid for m independent measurements, and

$$F(\theta_0) = \sum_{\mu}\frac{1}{p(\mu|\theta_0)}\left(\frac{\partial p(\mu|\theta)}{\partial\theta}\bigg|_{\theta_0}\right)^2 \tag{10}$$

is the Fisher information. The equality $\left(\Delta^2\theta_{\text{est}}\right)_{\boldsymbol{\mu}|\theta_0} = \Delta^2\theta_{\text{CRLB}}$ is achieved if and only if

$$\theta_{\text{est}}(\boldsymbol{\mu}) - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0} = \lambda_{\theta_0}\frac{d\log p(\boldsymbol{\mu}|\theta_0)}{d\theta_0}, \tag{11}$$

with λ_{θ_0} a parameter independent of $\boldsymbol{\mu}$ (while it may depend on θ_0). Noticing that $\frac{d\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}}{d\theta_0} = \sum_{\boldsymbol{\mu}}\left(\theta_{\text{est}}(\boldsymbol{\mu}) - f(\theta_0)\right)\frac{dp(\boldsymbol{\mu}|\theta_0)}{d\theta_0}$, the CRLB can be straightforwardly generalized to any function $f(\theta_0)$ independent of $\boldsymbol{\mu}$. In particular, choosing $f(\theta_0) = \theta_0$, we can directly prove that $\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\text{CRLB}}$, which also depends on the bias.

Asymptotically in m, the saturation of Equation (8) is obtained for the maximum likelihood estimator (MLE) [22,23,49]. This is the value $\theta_{\mathrm{MLE}}(\boldsymbol{\mu})$ that maximizes the likelihood $p(\boldsymbol{\mu}|\theta_0)$ (as a function of the parameter θ_0) for the observed measurement sequence $\boldsymbol{\mu}$,

$$\theta_{\mathrm{MLE}}(\boldsymbol{\mu}) \equiv \arg\max_{\theta_0}\{p(\boldsymbol{\mu}|\theta_0)\}. \tag{12}$$

For a sufficiently large sample size m (in the central limit), independently of the probability distribution $p(\boldsymbol{\mu}|\theta_0)$, the MLE becomes normally distributed [18,22,23,49]:

$$p(\theta_{\mathrm{MLE}}|\theta_0) = \sqrt{\frac{mF(\theta_0)}{2\pi}}e^{-\frac{mF(\theta_0)}{2}(\theta_0-\theta_{\mathrm{MLE}})^2} \qquad (m\gg 1), \tag{13}$$

with mean given by the true value θ_0 and variance equal to the inverse of the Fisher information. The MLE is well defined provided that there is a unique maximum in the considered phase interval. In the case of Equation (1), this condition is fulfilled provided that one restrict the phase domain to $[0, \pi/(2N)]$ for instance.

In Figure 1, we plot the results of a maximum likelihood analysis for the example considered in this manuscript. In this case, the MLE is readily calculated and given by $\theta_{\mathrm{MLE}}(\boldsymbol{\mu}) = \frac{1}{2}\arccos(\frac{m_+-m_-}{m_++m_-})$, and the Fisher information is $F(\theta_0) = N^2$, independent of θ_0 (we recall that $N = 2$ in our example). In Figure 1a we plot the bias $\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0} - \theta_0$ (dots) as a function of m, for $\theta_0 = \pi/4$. Error bars are $\pm\Delta\theta_{\mathrm{CRLB}}$. Notice that $\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0} = \theta_0$ for every m. This does not mean that the estimator is locally unbiased: indeed, the derivative $d\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0$ [see panel (b)] is different from 1 for every value of m. We have $d\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0 \to 1$ asymptotically in m. In Figure 1b, we plot $mF(\theta_0)(\Delta^2\theta_{\mathrm{MLE}})_{\boldsymbol{\mu}|\theta_0}$ as a function of the number of independent measurements m (red dots). This quantity is compared to $mF(\theta_0)\Delta^2\theta_{\mathrm{CRLB}} = (d\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0)^2$ (red line). With increasing sample size m, $(\Delta^2\theta_{\mathrm{MLE}})_{\boldsymbol{\mu}|\theta_0} \to 1/(mF(\theta_0))$ corresponding to the CRLB for unbiased estimators.
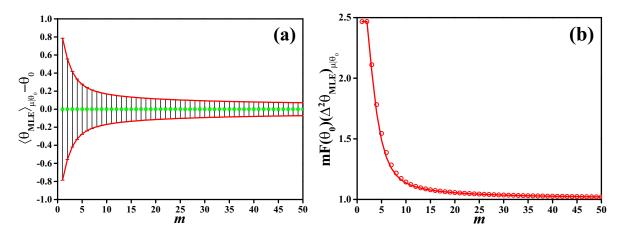

Figure 1. (a) Bias $\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0} - \theta_0$ (green dots) as function of m with error bars $(\Delta\theta_{\mathrm{MLE}})_{\boldsymbol{\mu}|\theta_0}$. The red lines are $\pm\Delta\theta_{\mathrm{CRLB}} = \pm|d\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0|/\sqrt{mF(\theta_0)}$; (b) variance of the maximum likelihood estimator multiplied by the Fisher information, $mF(\theta_0)(\Delta^2\theta_{\mathrm{MLE}})_{\boldsymbol{\mu}|\theta_0}$ (red circles), as a function of the sample size m. It is compared to the bias $(d\langle\theta_{\mathrm{MLE}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0)^2$ (red dashed line). We recall that $\theta_0 = \pi/4$ and $F(\theta_0) = 4$ here.

2.2.3. Hammersley–Chapman–Robbins Bound

The ChRB is obtained from Equation (5) by taking $n = 2, a_1 = 1, a_2 = -1, \theta_1 = \theta_0 + \lambda, \theta_2 = \theta_0$, and can be written as [47,48]

$$\Delta^2\theta_{\text{ChRB}} = \sup_\lambda \frac{\left(\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0+\lambda} - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2}{\sum_{\boldsymbol{\mu}}\frac{p(\boldsymbol{\mu}|\theta_0+\lambda)^2}{p(\boldsymbol{\mu}|\theta_0)} - 1}. \tag{14}$$

Clearly, restricting the number of parameters in the optimization in Equation (5) leads to a less strict bound. We thus have $\Delta^2\theta_{\text{BB}} \geq \Delta^2\theta_{\text{ChRB}}$. For unbiased estimators, we obtain

$$\Delta^2\theta_{\text{ChRB}}^{\text{ub}} = \sup_\lambda \frac{\lambda^2}{\sum_{\boldsymbol{\mu}}\frac{p(\boldsymbol{\mu}|\theta_0+\lambda)^2}{p(\boldsymbol{\mu}|\theta_0)} - 1}. \tag{15}$$

Furthermore, the supremum over λ on the right side of Equation (14) is always larger or equal to its limit $\lambda \to 0$:

$$\sup_\lambda \frac{\left(\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0+\lambda} - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2}{\sum_{\boldsymbol{\mu}}\frac{p(\boldsymbol{\mu}|\theta_0+\lambda)^2}{p(\boldsymbol{\mu}|\theta_0)} - 1} \geq \lim_{\lambda\to 0}\frac{\left(\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0+\lambda} - \langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2}{\sum_{\boldsymbol{\mu}}\frac{p(\boldsymbol{\mu}|\theta_0+\lambda)^2}{p(\boldsymbol{\mu}|\theta_0)} - 1} = \frac{\left(\frac{d\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}}{d\theta_0}\right)^2}{m\sum_{\boldsymbol{\mu}}\frac{1}{p(\boldsymbol{\mu}|\theta_0)}\left(\frac{dp(\boldsymbol{\mu}|\theta_0)}{d\theta_0}\right)^2}, \tag{16}$$

provided that the derivatives on the right-hand side exist. We thus recover the CRLB as a limiting case of the ChRB. The ChRB is always stricter than the CRLB and we obtain the last inequality in the chain (7). Notice that the CRLB requires the probability distribution $p(\boldsymbol{\mu}|\theta_0)$ to be differentiable [24]—a condition that can be dropped for the ChRB and the more general BB. Even if the distribution is regular, the above derivation shows that the ChRB, and more generally the BB, provide tighter error bounds than the CRLB. With increasing n, the BB becomes tighter and tighter and the CRLB represents the weakest bound in this hierarchy, which can be observed in Figure 2a. Next, we determine a stricter bound in this hierarchy.
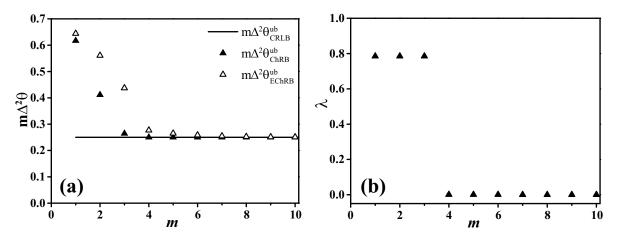

Figure 2. (**a**) comparison between unbiased frequentist bounds for the example considered in this manuscript, Equation (1): the CRLB $m\Delta^2\theta_{\text{CRLB}}^{\text{ub}} = 1/F(\theta_0)$ (black line), the Hammersley–Chapman–Robbins bound $m\Delta^2\theta_{\text{ChRB}}^{\text{ub}}$ (Equation (15), filled triangles) and the extended Hammersley–Chapman–Robbins bound $m\Delta^2\theta_{\text{EChRB}}^{\text{ub}}$ (Equation (18), empty triangles); (**b**) values of λ achieving the supremum in Equation (15), as a function of m.

2.2.4. Extended Hammersley–Chapman–Robbins Bound

We obtain the extended Hammersley–Chapman–Robbins bound (EChRB) as a special case of Equation (5), by taking $n = 3$, $a_1 = 1$, $a_2 = A$, $a_3 = -1$, $\theta_1 = \theta_0 + \lambda_1$, $\theta_2 = \theta_0 + \lambda_2$, and $\theta_3 = \theta_0$, giving

$$\Delta^2\theta_{\text{EChRB}} = \sup_{\lambda_1,\lambda_2,A} \frac{\left(\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0+\lambda_1} + A\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0+\lambda_2} - (1+A)\langle\theta_{\text{est}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2}{\sum_{\boldsymbol{\mu}}\frac{[p(\boldsymbol{\mu}|\theta_0+\lambda_1) - p(\boldsymbol{\mu}|\theta_0) + Ap(\boldsymbol{\mu}|\theta_0+\lambda_2)]^2}{p(\boldsymbol{\mu}|\theta_0)}}, \tag{17}$$

where the supremum is taken over all possible $\lambda_1, \lambda_2 \in \mathbb{N}$ and $A \in \mathbb{R}$. Since the ChRB is obtained from Equation (17) in the specific case $A = 0$, we have that $\Delta^2\theta_{\text{EChRB}} \geq \Delta^2\theta_{\text{ChRB}}$. For unbiased estimators, we obtain

$$\Delta^2\theta_{\text{EChRB}}^{\text{ub}} = \sup_{\lambda_1, \lambda_2, A} \frac{(\lambda_1 + A\lambda_2)^2}{\sum_{\boldsymbol{\mu}} \frac{[p(\boldsymbol{\mu}|\theta_0+\lambda_1) - p(\boldsymbol{\mu}|\theta_0) + Ap(\boldsymbol{\mu}|\theta_0+\lambda_2)]^2}{p(\boldsymbol{\mu}|\theta_0)}}. \tag{18}$$

In Figure 2a, we compare the different bounds for unbiased estimators and for the example considered in the manuscript: the CRLB (black line), the ChRB (filled triangles) and the EChRB (empty triangles), satisfying the chain of inequalities (7). In Figure 2b, we show the values of λ in Equation (15) for which the supremum is achieved in our case.

3. Bayesian Approach

The Bayesian approach makes use of the Bayes–Laplace theorem, which can be very simply stated and proved. The joint probability of two stochastic variables $\boldsymbol{\mu}$ and θ is symmetric: $p(\boldsymbol{\mu}, \theta) = p(\boldsymbol{\mu}|\theta)p(\theta) = p(\theta|\boldsymbol{\mu})p(\boldsymbol{\mu}) = p(\theta, \boldsymbol{\mu})$, where $p(\theta)$ and $p(\boldsymbol{\mu})$ are the marginal distributions, obtained by integrating the joint probability over one of the two variables, while $p(\boldsymbol{\mu}|\theta)$ and $p(\theta|\boldsymbol{\mu})$ are conditional distributions.

We recall that in a phase inference problem, the set of measurement results $\boldsymbol{\mu}$ is generated by a fixed and unknown value θ_0 according to the likelihood $p(\boldsymbol{\mu}|\theta_0)$. In the Bayesian approach to the estimation of θ_0, one introduces a random variable θ and uses the Bayes–Laplace theorem to define the conditional probability

$$p_{\text{post}}(\theta|\boldsymbol{\mu}) = \frac{p(\boldsymbol{\mu}|\theta)p_{\text{pri}}(\theta)}{p_{\text{mar}}(\boldsymbol{\mu})}. \tag{19}$$

The posterior probability $p_{\text{post}}(\theta|\boldsymbol{\mu})$ provides a degree of belief, or plausibility, that $\theta_0 = \theta$ (i.e., that θ is the true value of the phase), in the light of the measurement data $\boldsymbol{\mu}$ [50]. In Equation (19), the prior distribution $p_{\text{pri}}(\theta)$ expresses the a priori state of knowledge on θ, $p(\boldsymbol{\mu}|\theta)$ is the likelihood that is determined by the quantum mechanical measurement postulate, e.g., as in Equation (1), and the marginal probability $p_{\text{mar}}(\boldsymbol{\mu}) = \int_a^b d\theta\, p(\theta, \boldsymbol{\mu})$ is obtained through the normalization for the posterior, where a and b are boundaries of the phase domain. The posterior probability $p_{\text{post}}(\theta|\boldsymbol{\mu})$ describes the current knowledge about the random variable θ based on the available information, i.e., the measurement results $\boldsymbol{\mu}$.

3.1. Noninformative Prior

In the Bayesian approach, the information on θ provided by the posterior probability always depends on the prior distribution $p_{\text{pri}}(\theta)$. It is possible to account for the available a priori information on θ by choosing a prior distribution accordingly. However, if no a priori information is available, it is not obvious how to choose a "noninformative" prior [51]. The flat prior $p_{\text{pri}}(\theta) = const$ was first introduced by Laplace to express the absence of information on θ [51]. However, this prior would not be flat for other functions of θ and, in the complete absence of a priori information, it seems unreasonable that some information is available for different parametrizations of the problem. To see this, recall that a transformation of variables requires that $p_{\text{pri}}(\varphi) = p_{\text{pri}}(\theta)|df^{-1}(\varphi)/d\varphi|$ for any function $\varphi = f(\theta)$. Hence, if $p_{\text{pri}}(\theta)$ is flat, one obtains that $p_{\text{pri}}(\varphi) = |df^{-1}(\varphi)/d\varphi|$ is, in general, not flat.

Notice that $p_{\text{pri}}(\theta) \propto \sqrt{F(\theta)}$—called Jeffreys prior [52,53]—where $F(\theta)$ is the Fisher information (10), remains invariant under re-parametrization. For arbitrary transformations $\varphi = f(\theta)$, the Fisher information obeys the transformation property $F(\varphi) = F(\theta)(d\theta/d\varphi)^2 = F(\theta)(df^{-1}(\varphi)/d\varphi)^2$. Therefore, if $p_{\text{pri}}(\theta) \propto \sqrt{F(\theta)}$ and we perform the change of variable $\varphi = f(\theta)$, then the transformation property of the Fisher information ensures that $p_{\text{pri}}(\varphi) = p_{\text{pri}}(\theta)|df^{-1}(\varphi)/d\varphi| \propto \sqrt{F(\varphi)}$. Notice that, as in our case, the Fisher information $F(\theta)$

may actually be independent of θ. In this case, the invariance property does not imply that Jeffreys prior is flat for arbitrary re-parametrizations $\varphi = f(\theta)$, instead, $\sqrt{F(\varphi)} = |df^{-1}(\varphi)/d\varphi|$.

3.2. Posterior Bounds

From the posterior probability (19), we can provide an estimate $\theta_{\mathrm{BL}}(\mu)$ of θ_0. This can be the maximum a posteriori, $\theta_{\mathrm{BL}}(\mu) = \arg\max_\theta p_{\mathrm{post}}(\theta|\mu)$, which coincides with the maximum likelihood Equation (12) when the prior is flat, $p_{\mathrm{pri}}(\theta) = const$, or the mean of the distribution, $\theta_{\mathrm{BL}}(\mu) = \int_a^b d\theta\, \theta\, p_{\mathrm{post}}(\theta|\mu)$.

With the Bayesian approach, it is possible to provide a confidence interval around the estimator, given an arbitrary measurement sequence μ, even with a single measurement. The variance

$$\left(\Delta^2\theta_{\mathrm{BL}}(\mu)\right)_{\theta|\mu} = \int_a^b d\theta\, p_{\mathrm{post}}(\theta|\mu)\left(\theta - \theta_{\mathrm{BL}}(\mu)\right)^2 \tag{20}$$

can be taken as a measure of fluctuation of our degree of belief around $\theta_{\mathrm{BL}}(\mu)$. There is no such concept in the frequentist paradigm. The Bayesian posterior variance $\left(\Delta^2\theta_{\mathrm{BL}}(\mu)\right)_{\theta|\mu}$ and the frequentist variance $\left(\Delta^2\theta_{\mathrm{BL}}\right)_{\mu|\theta_0}$ have entirely different operational meanings. Equation (20) provides a degree of plausibility that $\theta_{\mathrm{BL}}(\mu) = \theta_0$, given the measurement results μ. There is no notion of bias in this case. On the other hand, the quantity $\left(\Delta^2\theta_{\mathrm{BL}}\right)_{\mu|\theta_0}$ measures the statistical fluctuations of $\theta_{\mathrm{BL}}(\mu)$ when repeating the sequence of m measurements infinitely many times.

Ghosh Bound

In the following, we derive a lower bound to Equation (20) first introduced by Ghosh [54]. Using $\int_a^b d\theta\, p_{\mathrm{post}}(\theta|\mu) = 1$, we have

$$\int_a^b d\theta\, (\theta - \theta_{\mathrm{BL}}(\mu)) \frac{dp_{\mathrm{post}}(\theta|\mu)}{d\theta} = p_{\mathrm{post}}(\theta|\mu)(\theta - \theta_{\mathrm{BL}}(\mu))\big|_a^b - \int_a^b d\theta\, p_{\mathrm{post}}(\theta|\mu) = f(\mu, a, b) - 1, \tag{21}$$

where $f(\mu, a, b) = b p_{\mathrm{post}}(b|\mu) - a p_{\mathrm{post}}(a|\mu) - \theta_{\mathrm{BL}}(\mu)(p_{\mathrm{post}}(b|\mu) - p_{\mathrm{post}}(a|\mu))$ depends on the value of the posterior distribution calculated at the boundaries. If $p_{\mathrm{pri}}(a) = p_{\mathrm{pri}}(b) = 0$, we have $f(\mu, a, b) = 0$. Analogously with the derivation of the (frequenstist) CRLB, we exploit the Cauchy–Schwarz inequality,

$$\left(\int_a^b d\theta\, \left(\frac{dp_{\mathrm{post}}(\theta|\mu)}{d\theta}\right)^2 \frac{1}{p_{\mathrm{post}}(\theta|\mu)}\right)\left(\int_a^b d\theta\, p_{\mathrm{post}}(\theta|\mu)(\theta - \theta_{\mathrm{BL}}(\mu))^2\right) \geq (f(\mu, a, b) - 1)^2,$$

leading to $\left(\Delta^2\theta_{\mathrm{BL}}(\mu)\right)_{\theta|\mu} \geq \Delta^2\theta_{\mathrm{GB}}(\mu)$, where [54]

$$\Delta^2\theta_{\mathrm{GB}}(\mu) = \frac{(f(\mu, a, b) - 1)^2}{\int_a^b d\theta\, \frac{1}{p_{\mathrm{post}}(\theta|\mu)}\left(\frac{dp_{\mathrm{post}}(\theta|\mu)}{d\theta}\right)^2}. \tag{22}$$

The above bound is a function of the specific measurement sequence μ and depends on $\int_a^b d\theta\, \frac{1}{p_{\mathrm{post}}(\theta|\mu)}\left(\frac{dp_{\mathrm{post}}(\theta|\mu)}{d\theta}\right)^2$ that we can identify as a "Fisher information of the posterior distribution". The Ghosh bound is saturated if and only if

$$\theta - \theta_{\mathrm{BL}}(\mu) = \lambda_\mu \frac{d\log p(\theta|\mu)}{d\theta}, \tag{23}$$

where λ_μ does not depend on θ while it may depend on μ.

3.3. Average Posterior Bounds

While Equation (20) depends on the specific $\boldsymbol{\mu}$, it is natural to consider its average over all possible measurement sequences at fixed θ_0 and m, weighted by the likelihood $p(\boldsymbol{\mu}|\theta_0)$:

$$\left(\Delta^2\theta_{\mathrm{BL}}\right)_{\boldsymbol{\mu},\theta|\theta_0} = \sum_{\boldsymbol{\mu}} \left(\Delta^2\theta_{\mathrm{BL}}(\boldsymbol{\mu})\right)_{\theta|\boldsymbol{\mu}} p(\boldsymbol{\mu}|\theta_0) = \sum_{\boldsymbol{\mu}} \int_a^b d\theta\, p(\theta,\boldsymbol{\mu}|\theta_0) \left(\theta - \theta_{\mathrm{BL}}(\boldsymbol{\mu})\right)^2, \qquad (24)$$

which we indicate as average Bayesian posterior variance, where $p(\theta,\boldsymbol{\mu}|\theta_0) = p_{\mathrm{post}}(\theta|\boldsymbol{\mu})p(\boldsymbol{\mu}|\theta_0)$.

We would be tempted to compare the average posterior sensitivity $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu},\theta|\theta_0}$ to the frequentist Cramér–Rao bound $\Delta^2\theta_{\mathrm{CRLB}}$. However, because of the different operational meanings of the frequentist and the Bayesian paradigms, there is no reason for Equation (24) to fulfill the Cramér–Rao bound: indeed, it does not, as we show below.

Likelihood-Averaged Ghosh Bound

A lower bound to Equation (24) is obtained by averaging the Ghosh bound Equation (22) over the likelihood function. We have $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu},\theta|\theta_0} \geq \Delta^2\theta_{\mathrm{aGB}}$, where [18]

$$\Delta^2\theta_{\mathrm{aGB}} = \sum_{\boldsymbol{\mu}} \frac{(f(\boldsymbol{\mu},a,b)-1)^2}{\int_a^b d\theta\, \frac{1}{p_{\mathrm{post}}(\theta|\boldsymbol{\mu})}\left(\frac{\partial p_{\mathrm{post}}(\theta|\boldsymbol{\mu})}{\partial\theta}\right)^2} p(\boldsymbol{\mu}|\theta_0). \qquad (25)$$

This likelihood-averaged Ghosh bound is independent of $\boldsymbol{\mu}$ because of the statistical average.

3.4. Numerical Comparison of Bayesian and Frequentist Phase Estimation

In the numerical calculations shown in Figure 3, we consider a Bayesian estimator given by $\theta_{\mathrm{BL}}(\boldsymbol{\mu}) = \int_a^b d\theta\, \theta\, p_{\mathrm{post}}(\theta|\boldsymbol{\mu})$ with prior distributions

$$p_{\mathrm{pri}}(\theta) = \frac{2}{\pi} \frac{e^{\alpha\sin(2\theta)^2}-1}{e^{\alpha/2}I_0(\alpha/2)-1}, \qquad (26)$$

where $I_0(\alpha)$ is the modified Bessel function of the first kind. This choice of prior distribution can continuously turn from a peaked function to a flat one when changing α, while being differentiable in the full phase interval. The more negative is α, the more $p_{\mathrm{pri}}(\theta)$ broadens in $[0, \pi/2]$. In particular, in the limit $\alpha \to -\infty$, the prior approaches the flat distribution, which in our case coincides with Jeffreys prior since the Fisher information is independent of θ. In the limit $\alpha = 0$, the prior is given by $\lim_{\alpha\to 0} p_{\mathrm{pri}}(\theta) = 4\sin(2\theta)^2/\pi$. For positive values of α, the larger α, the more peaked is $p_{\mathrm{pri}}(\theta)$ around $\theta_0 = \pi/4$. In particular $p_{\mathrm{pri}}(\theta) \approx e^{-4\alpha(\theta-\pi/4)^2}/\sqrt{\pi/4\alpha}$ for $\alpha \gg 1$. Equation (26) is normalized to one for $\theta \in [0, \frac{\pi}{2}]$. In the inset of the different panels of Figure 3, we plot $p_{\mathrm{pri}}(\theta)$ for $\alpha = -100$ [panel (a)], $\alpha = -10$ (b), $\alpha = 1$ (c) and $\alpha = 10$ (d).

In Figure 3, we plot, as a function of m, the posterior variance $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu},\theta|\theta_0}$ (blue circles) that, as expected, is always larger than the likelihood-averaged Ghosh bound Equation (25) (solid blue lines). For comparison, we also plot the frequentist variance $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu}|\theta_0} = \sum_{\boldsymbol{\mu}} \left(\theta_{\mathrm{BL}}(\boldsymbol{\mu}) - \langle\theta_{\mathrm{BL}}\rangle_{\boldsymbol{\mu}|\theta_0}\right)^2 p(\boldsymbol{\mu}|\theta_0)$ (red dots) around the mean value $\langle\theta_{\mathrm{BL}}\rangle_{\boldsymbol{\mu}|\theta_0} = \sum_{\boldsymbol{\mu}} \theta_{\mathrm{BL}}(\boldsymbol{\mu})p(\boldsymbol{\mu}|\theta_0)$ of the estimator. This quantity obeys the Cramér–Rao theorem $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu}|\theta_0} \geq \Delta^2\theta_{\mathrm{CRLB}}$ and the more general chain of inequalities (7). This is confirmed in the figure where we show $\Delta^2\theta_{\mathrm{CRLB}} = |d\langle\theta_{\mathrm{BL}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0|^2/(mF(\theta_0))$ (red line). Notice that, when the prior narrows around θ_0, the variance $(\Delta^2\theta_{\mathrm{BL}})_{\boldsymbol{\mu}|\theta_0}$ decreases, but, at the same time, the estimator becomes more and more biased, i.e., $|d\langle\theta_{\mathrm{BL}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0|$ decreases as well (note indeed that the red dashed line is proportional to $|d\langle\theta_{\mathrm{BL}}\rangle_{\boldsymbol{\mu}|\theta_0}/d\theta_0|^2$).
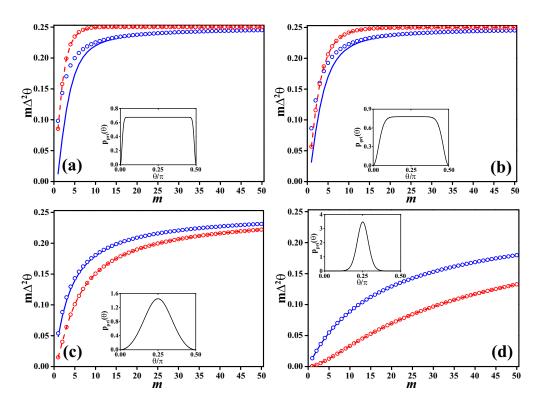
Figure 3. Comparisons of phase estimation variance as a function of the sample size for Bayesian and frequentist data analysis under different prior distributions, (**a**) $\alpha = -100$, (**b**) $\alpha = -10$, (**c**) $\alpha = 1$, (**d**) $\alpha = 10$. In all figures, Red circles (frequentist) are $m(\Delta^2\theta_{\mathrm{BL}})_{\mu|\theta_0}$, the red dashed line is the Cramér-Rao lower bound $m\Delta^2\theta_{\mathrm{CRLB}}$, Equation (8). Blue circles (Bayesian) are $m(\Delta^2\theta_{\mathrm{BL}})_{\mu,\theta|\theta_0}$, the blue solid line is the likelihood-averaged Ghosh bound $m\Delta^2\theta_{\mathrm{aGB}}$, Equation (25). The inset in each panel is $p_{\mathrm{pri}}(\theta)$, Equation (26), for the corresponding values of α.

Interestingly, in Figure 3, we clearly see that the Bayesian posterior variance $(\Delta^2\theta_{\mathrm{BL}})_{\mu,\theta|\theta_0}$ and the likelihood-averaged Ghosh bound may stay in some cases below the (frequentist) $\Delta^2\theta_{\mathrm{CRLB}}$ [see panels (a) and (b)], even if the prior is almost flat. The discrepancy with the CRLB is remarkable and can be quite large for small values of m. Still, there is no contradiction since $(\Delta^2\theta_{\mathrm{BL}})_{\mu,\theta|\theta_0}$ and $(\Delta^2\theta_{\mathrm{BL}})_{\mu|\theta_0}$ have different operational meanings and interpretations. They both respect their corresponding sensitivity bounds.

Asymptotically in the number of measurements m, the Ghosh bound as well as its likelihood average converge to the Cramér–Rao bound. Indeed, it is well known that in this limit the posterior probability becomes a Gaussian centered at the true value of the phase shift and with variance given by the inverse of the Fisher information,

$$p_{\mathrm{post}}(\theta|\mu) = \sqrt{\frac{mF(\theta_0)}{2\pi}}e^{-\frac{mF(\theta_0)}{2}(\theta-\theta_0)^2}, \qquad (m \gg 1), \qquad (27)$$

a result known as Laplace–Bernstein–von Mises theorem [18,23,55]. By replacing Equation (27) into Equation (22), we recover a posterior variance given by $1/(mF(\theta_0))$.

4. Bounds for Random Parameters

In this section, we derive bounds of phase sensitivity obtained when θ_0 is a random variable distributed according to $p(\theta_0)$. Operationally, this corresponds to the situation where θ_0 remains fixed (but unknown) when collecting a single sequence of m measurements μ. In between measurement sequences, θ_0 fluctuates according to $p(\theta_0)$.

4.1. Frequentist Risk Functions for Random Parameters

Let us first consider the frequentist estimation of a fluctuating parameter θ_0 with the estimator θ_{est}. The mean sensitivity obtained by averaging $(\Delta^2 \theta_{\text{est}})_{\mu|\theta_0}$, Equation (3), over $p(\theta_0)$ is

$$
\begin{aligned}
(\Delta^2 \theta_{\text{est}})_{\mu,\theta_0} &= \int_a^b d\theta_0 (\Delta^2 \theta_{\text{est}})_{\mu|\theta_0} p(\theta_0) \\
&= \sum_{\mu} \int_a^b d\theta_0 \, p(\mu|\theta_0) p(\theta_0) \left(\langle \theta_{\text{est}} \rangle_{\mu|\theta_0} - \theta_{\text{est}}(\mu) \right)^2 \\
&= \sum_{\mu} \int_a^b d\theta_0 \, p(\mu,\theta_0) \left(\langle \theta_{\text{est}} \rangle_{\mu|\theta_0} - \theta_{\text{est}}(\mu) \right)^2,
\end{aligned}
\tag{28}
$$

where μ and θ_0 are both random variables and we have used $p(\mu|\theta_0)p(\theta_0) = p(\mu,\theta_0)$.

An averaged risk function for the efficiency of the estimator is given by averaging the mean square error (3) over $p(\theta_0)$, leading to

$$
\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0} = \int d\theta_0 \text{MSE}(\theta_{\text{est}})_{\mu|\theta_0} p(\theta_0) = \int d\theta_0 \sum_{\mu} \left(\theta_{\text{est}}(\mu) - \theta_0 \right)^2 p(\mu,\theta_0).
\tag{29}
$$

Analogously to Equation (4), we can write

$$
\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0} = (\Delta^2 \theta_{\text{est}})_{\mu,\theta_0} + \int d\theta_0 \left(\langle \theta_{\text{est}} \rangle_{\mu|\theta_0} - \theta_0 \right)^2 p(\theta_0).
\tag{30}
$$

In the following, we derive lower bounds for both $(\Delta^2 \theta_{\text{est}})_{\mu,\theta_0}$ and $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0}$. Notice that bounds on $(\Delta^2 \theta_{\text{est}})_{\mu,\theta_0}$ hold also for $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0}$ due to $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0} \geq (\Delta^2 \theta_{\text{est}})_{\mu,\theta_0}$. Nevertheless, bounds on the average the mean square error are widely used (and are often called Bayesian bounds [56]) since they can be expressed independently of the bias.

4.2. Bounds on the Mean Square Error

We first consider bounds on $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0}$, Equation (29), for arbitrary estimators.

4.2.1. Van Trees Bound

It is possible to derive a general lower bound on the mean square error (29) based on the following assumptions:

1. $\frac{\partial p(\mu,\theta_0)}{\partial \theta_0}$ and $\frac{\partial^2 p(\mu,\theta_0)}{\partial \theta_0^2}$ are absolutely integrable with respect to μ and θ_0;

2. $p(a)\,\xi(a) - p(b)\,\xi(b) = 0$, where $\xi(\theta_0) = \sum_{\mu} \left(\theta_{\text{est}}(\mu) - \theta_0 \right) p(\mu|\theta_0)$.

Multiplying $\xi(\theta_0)$ by $p(\theta_0)$ and differentiating with respect to θ_0, we have

$$
\frac{\partial p(\theta_0)\xi(\theta_0)}{\partial \theta_0} = \sum_{\mu} \left(\theta_{\text{est}}(\mu) - \theta_0 \right) \frac{\partial p(\mu,\theta_0)}{\partial \theta_0} - p(\theta_0).
$$

Integrating over θ_0 in the range of $[a,b]$ and considering the above properties, we find

$$
\sum_{\mu} \int_a^b d\theta_0 \left(\theta_{\text{BL}}(\mu) - \theta_0 \right) \frac{\partial p(\mu,\theta_0)}{\partial \theta_0} = 1.
\tag{31}
$$

Finally, using the Cauchy–Schwarz inequality, we arrive at $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0} \geq \Delta^2 \theta_{\text{VTB}}$, where

$$
\Delta^2 \theta_{\text{VTB}} = \frac{1}{\sum_{\mu} \int_a^b d\theta_0 \frac{1}{p(\mu,\theta_0)} \left(\frac{\partial p(\mu,\theta_0)}{\partial \theta_0} \right)^2}
\tag{32}
$$

is generally indicated as Van Trees bound [24,56,57]. The equality holds if and only if

$$\theta_{\text{est}}(\boldsymbol{\mu}) - \theta_0 = \lambda \frac{d \log p(\boldsymbol{\mu}, \theta_0)}{d\theta_0}, \tag{33}$$

where λ does not depend on θ_0 and $\boldsymbol{\mu}$. It is easy to show that

$$\sum_{\boldsymbol{\mu}} \int_a^b d\theta_0 \frac{1}{p(\boldsymbol{\mu}, \theta_0)} \left(\frac{\partial p(\boldsymbol{\mu}, \theta_0)}{\partial \theta_0} \right)^2 = m \int_a^b d\theta_0 \, p(\theta_0) F(\theta_0) + \int_a^b d\theta_0 \frac{1}{p(\theta_0)} \left(\frac{\partial p(\theta_0)}{\partial \theta_0} \right)^2, \tag{34}$$

where the first term is the Fisher information $F(\theta_0)$, defined by Equation (10), averaged over $p(\theta_0)$, and the second term can be interpreted as a Fisher information of the prior [24]. Asymptotically in the number of measurements m and for regular distributions $p(\theta_0)$, the first term in Equation (34) dominates over the second one.

4.2.2. Ziv–Zakai Bound

A further bound on $\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu},\theta_0}$ can be derived by mapping the phase estimation problem to a continuous series of binary hypothesis testing problems. A detailed derivation of the Ziv–Zakai bound [24,58,59] is provided in Appendix B. The final result reads $\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu},\theta_0} \geq \Delta^2 \theta_{\text{ZZB}}$, where

$$\Delta^2 \theta_{\text{ZZB}} = \frac{1}{2} \int dh \, h \int d\theta_0 \, (p(\theta_0) + p(\theta_0 + h)) \, P_{\min}(\theta_0, \theta_0 + h), \tag{35}$$

and

$$P_{\min}(\theta_0, \theta_0 + h) = \frac{1}{2} \left(1 - \sum_{\boldsymbol{\mu}} \left| \frac{p(\theta_0) \, p(\boldsymbol{\mu}|\theta_0)}{p(\theta_0) + p(\theta_0 + h)} - \frac{p(\theta_0 + h) \, p(\boldsymbol{\mu}|\theta_0 + h)}{p(\theta_0) + p(\theta_0 + h)} \right| \right) \tag{36}$$

is the minimum error probability of the binary hypothesis testing problem. This bound has been adopted for quantum phase estimation in Ref. [26]. To this end, the probability $P_{\min}(\theta_0, \theta_0 + h)$ can be maximized over all possible quantum measurements, which leads to the trace distance [7]. As the optimal measurement may depend on θ_0 and h, the bound (35), which involves integration over all values of θ_0 and h, is usually not saturable. We remark that the trace distance also defines a saturable frequentist bound for a different risk function than the variance [60].

4.3. Bounds on the Average Estimator Variance

We now consider bounds on $(\Delta^2 \theta_{\text{est}})_{\boldsymbol{\mu},\theta_0}$, Equation (28), for arbitrary estimators.

4.3.1. Average CRLB

Taking the average over $p(\theta_0)$ of Equation (7), we obtain a chain of bounds for $(\Delta^2 \theta_{\text{est}})_{\boldsymbol{\mu},\theta_0}$. In particular, in its simplest form, we have $(\Delta^2 \theta_{\text{est}})_{\boldsymbol{\mu},\theta_0} \geq \Delta^2 \theta_{\text{aCRLB}}$, where

$$\Delta^2 \theta_{\text{aCRLB}} = \int_a^b d\theta_0 \frac{\left(\frac{d \langle \theta_{\text{est}} \rangle_{\boldsymbol{\mu}|\theta_0}}{d\theta_0} \right)^2}{m F(\theta_0)} p(\theta_0) \tag{37}$$

is the average CRLB.

4.3.2. Van Trees Bound for the Average Estimator Variance

We can derive a general lower bound for the variance (28) by following the derivation of the Van Trees bound, which was discussed in Section 4.2.1. In contrast to the standard Van Trees bound for the mean square error, here the bias enters explicitly. Defining $\xi(\theta_0) = \sum_{\boldsymbol{\mu}} \left(\theta_{\text{est}}(\boldsymbol{\mu}) - \langle \theta_{\text{est}} \rangle_{\boldsymbol{\mu}|\theta_0} \right) p(\boldsymbol{\mu}|\theta_0)$

and assuming the same requirements as in the derivation of the Van Trees bound for the MSE, we arrive at

$$\sum_{\mu} \int_a^b d\theta_0 \left(\theta_{\text{est}}(\mu) - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right) \frac{\partial p(\mu, \theta_0)}{\partial \theta_0} = \int_a^b d\theta_0 \frac{d\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}}{d\theta_0} p(\theta_0).$$

Finally, a Cauchy–Schwarz inequality gives $(\Delta^2\theta_{\text{est}})_{\mu,\theta_0} \geq \Delta^2\theta_{\text{fVTB}}$, where

$$\Delta^2\theta_{\text{fVTB}} = \frac{\left(\int_a^b d\theta_0 \frac{d\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}}{d\theta_0} p(\theta_0)\right)^2}{\sum_{\mu} \int_a^b d\theta_0 \frac{1}{p(\mu,\theta_0)} \left(\frac{\partial p(\mu,\theta_0)}{\partial \theta_0}\right)^2}, \tag{38}$$

with equality if and only if

$$\theta_{\text{est}}(\mu) - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0} = \lambda \frac{d\log p(\mu, \theta_0)}{d\theta_0}, \tag{39}$$

where λ is independent of θ_0 and μ.

We can compare Equation (38) with the average CRLB Equation (37). We find

$$\int_a^b d\theta_0 \frac{\left(\frac{d\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}}{d\theta_0}\right)^2}{mF(\theta_0)} p(\theta_0) \geq \frac{\left(\int_a^b d\theta_0 \frac{d\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}}{d\theta_0} p(\theta_0)\right)^2}{m\int_a^b d\theta_0 p(\theta_0) F(\theta_0)} \geq \frac{\left(\int_a^b d\theta_0 \left|\frac{d\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}}{d\theta_0}\right| p(\theta_0)\right)^2}{\sum_{\mu} \int_a^b d\theta_0 \frac{1}{p(\mu,\theta_0)} \left(\frac{\partial p(\mu,\theta_0)}{\partial \theta_0}\right)^2},$$

where in the first step we use Jensen's inequality, and the second step follows from Equation (34) which implies $m \int_a^b d\theta_0 p(\theta_0) F(\theta_0) \leq \sum_{\mu} \int_a^b d\theta_0 \frac{1}{p(\mu,\theta_0)} \left(\frac{\partial p(\mu,\theta_0)}{\partial \theta_0}\right)^2$ since $\int_a^b d\theta_0 \frac{1}{p(\theta_0)} \left(\frac{dp(\theta_0)}{d\theta_0}\right)^2 \geq 0$.

We thus arrive at

$$(\Delta^2\theta_{\text{est}})_{\mu,\theta_0} \geq \Delta^2\theta_{\text{aCRLB}} \geq \Delta^2\theta_{\text{fVTB}}, \tag{40}$$

which is valid for generic estimators.

4.4. Bayesian Framework for Random Parameters

The Bayesian posterior variance, $(\Delta^2\theta_{\text{BL}})_{\mu,\theta|\theta_0}$, Equation (24), averaged over $p(\theta_0)$ is

$$
\begin{aligned}
(\Delta^2\theta_{\text{BL}})_{\mu,\theta,\theta_0} &= \int_a^b d\theta_0 (\Delta^2\theta_{\text{BL}})_{\mu,\theta|\theta_0} p(\theta_0) \\
&= \sum_{\mu} \int_a^b d\theta \int_a^b d\theta_0 \, p_{\text{post}}(\theta|\mu) p(\mu|\theta_0) p(\theta_0) \left(\theta - \theta_{\text{BL}}(\mu)\right)^2 \\
&= \sum_{\mu} \int_a^b d\theta \, p_{\text{post}}(\theta|\mu) p(\mu) \left(\theta - \theta_{\text{BL}}(\mu)\right)^2,
\end{aligned} \tag{41}
$$

where $p(\mu) = \int_a^b d\theta_0 \, p(\mu|\theta_0) p(\theta_0)$ is the average probability to observe μ taking into account fluctuations of θ_0.

A bound on Equation (41) can be obtained by averaging Equation (25) over $p(\theta_0)$, or, equivalently, averaging the Ghosh bound, Equation (22), over $p(\mu)$. We obtain the average Ghosh bound for random parameters θ_0, $(\Delta^2\theta_{\text{BL}})_{\mu,\theta,\theta_0} \geq \Delta^2\theta_{\text{aGBr}}$, where

$$
\begin{aligned}
\Delta^2\theta_{\text{aGBr}} &= \int_a^b d\theta_0 \sum_{\mu} \frac{(f(\mu, a, b) - 1)^2}{\int_a^b d\theta \frac{1}{p_{\text{post}}(\theta|\mu)} \left(\frac{dp_{\text{post}}(\theta|\mu)}{d\theta}\right)^2} p(\mu|\theta_0) p(\theta_0) \\
&= \sum_{\mu} \frac{(f(\mu, a, b) - 1)^2}{\int_a^b d\theta \frac{1}{p_{\text{post}}(\theta|\mu)} \left(\frac{dp_{\text{post}}(\theta|\mu)}{d\theta}\right)^2} p(\mu).
\end{aligned} \tag{42}
$$

The bound holds for any prior $p_{\mathrm{pri}}(\theta)$ and is saturated if and only if, for every value of μ, there exists a λ_μ such that Equation (23) holds.

Bayesian Bounds

In Equation (41), the prior used to define the posterior $p_{\mathrm{post}}(\theta|\mu)$ via the Bayes–Laplace theorem is arbitrary. In general, such a prior $p_{\mathrm{pri}}(\theta)$ is different from the statistical distribution of θ_0, which can be unknown. If $p(\theta_0)$ is known, then one can use it as a prior in the Bayesian posterior probability, i.e., $p_{\mathrm{pri}}(\theta) = p(\theta_0)$. In this specific case, we have $p_{\mathrm{mar}}(\mu) = p(\mu)$, and thus $p_{\mathrm{post}}(\theta|\mu)p(\mu) = p_{\mathrm{post}}(\theta|\mu)p_{\mathrm{mar}}(\mu) = p(\mu, \theta)$. In other words, for this specific choice of prior, the physical joint probability $p(\mu, \theta_0)$ of random variables θ_0 and μ coincides with the Bayesian $p(\mu, \theta)$. Equation (41) thus simplifies to

$$(\Delta^2\theta_{\mathrm{BL}})_{\mu,\theta} = \sum_{\mu} \int_a^b d\theta \, p(\mu, \theta)\left(\theta - \theta_{\mathrm{BL}}(\mu)\right)^2. \tag{43}$$

Notice that this expression is mathematically equivalent to the frequentist average mean square error (29) if we replace θ with θ_0 and $\theta_{\mathrm{BL}}(\mu)$ with $\theta_{\mathrm{est}}(\mu)$. This means that precision bounds for Equation (29), e.g., the Van Trees and Ziv–Zakai bounds can also be applied to Equation (43). These bounds are indeed often referred to as "Bayesian bounds" (see Ref. [24]).

We emphasize that the average over the marginal distribution $p_{\mathrm{mar}}(\mu)$, which connects Equations (24) and (43), has operational meaning if we consider that θ_0 is a random variable distributed according to $p(\theta_0)$, and $p(\theta)$ is used as prior in the Bayes–Laplace theorem to define a posterior distribution. In this case, and under the condition $f(\mu, a, b) = 0$ (for instance if the prior distribution vanishes at the borders of the phase domain), using Jensen's inequality, we find

$$
\begin{aligned}
\Delta^2\theta_{\mathrm{aGBr}} &= \sum_{\mu} \frac{p(\mu)}{\int_a^b d\theta \, \frac{1}{p_{\mathrm{post}}(\theta|\mu)}\left(\frac{d p_{\mathrm{post}}(\theta|\mu)}{d\theta}\right)^2} \\
&\geq \frac{1}{\sum_{\mu} p(\mu) \int_a^b d\theta \, \frac{1}{p_{\mathrm{post}}(\theta|\mu)}\left(\frac{d p_{\mathrm{post}}(\theta|\mu)}{d\theta}\right)^2} \\
&= \frac{1}{\sum_{\mu} \int_a^b d\theta \, \frac{1}{p(\theta,\mu)}\left(\frac{\partial p(\theta,\mu)}{\partial\theta}\right)^2},
\end{aligned}
\tag{44}
$$

which coincides with the Van Trees bound discussed above. We thus find that the averaged Ghosh bound for random parameters (42) is sharper than the Van Trees bound (38):

$$(\Delta^2\theta_{\mathrm{BL}})_{\mu,\theta} \geq \Delta^2\theta_{\mathrm{aGBr}} \geq \Delta^2\theta_{\mathrm{VTB}}, \tag{45}$$

which is also confirmed by the numerical data shown in Figure 4.

In Figure 4, we compare $\left(\Delta^2\theta_{\mathrm{BL}}\right)_{\mu,\theta}$ with the various bounds discussed in this section. As $p(\theta_0)$, we consider the same prior (26) used in Figure 3. We observe that all bounds approach the Van Trees bound with increasing sharpness of the prior distribution. Asymptotically in the number of measurements m, all bounds converge to the Cramér–Rao bound.
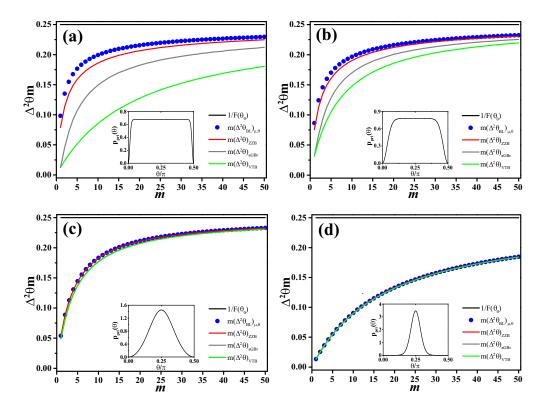
Figure 4. Comparisons of average posterior Bayesian variance, $m(\Delta^2\theta_{BL})_{\mu,\theta}$ (dots), as a function of the sample size m under different prior distributions, (**a**) $\alpha = -100$, (**b**) $\alpha = -10$, (**c**) $\alpha = 1$, (**d**) $\alpha = 10$. This variance is compared to to the average Ghosh bound for random parameters $m(\Delta^2\theta_{aGBr})$ (grey line), the Van Trees bound $m(\Delta^2\theta_{VTB})$ (green line), the Ziv–Zakai bound $m(\Delta^2\theta_{ZZB})$ (red line) and $1/F(\theta_0)$ (black horizontal line). The inset in each panel is the prior $p_{pri}(\theta)$, Equation (26), for the corresponding values of α.

5. Discussion and Conclusions

In this manuscript, we have clarified the differences between frequentist and Bayesian approaches to phase estimation. The two paradigms provide statistical results that have a different conceptual meaning and cannot be compared. We have also reviewed and discussed phase sensitivity bounds in the frequentist and Bayesian frameworks, when the true value of the phase shift θ_0 is fixed or fluctuates. These bounds are summarized in Table 1.

In the frequentist approach, for a fixed θ_0, the phase sensitivity is determined from the width of the probability distribution of the estimator. The physical content of the distribution is that, when repeating the estimation protocol, the obtained $\theta_{est}(\mu)$ will fall, with a certain confidence, in an interval around the mean value $\langle\theta_{est}\rangle_{\mu|\theta_0}$ (e.g., 68% of the times within a $2(\Delta\theta_{est})_{\mu|\theta_0}$ interval for a Gaussian distribution) that, for unbiased estimators, coincides with the true value of the phase shift.

In the Bayesian case, the posterior $p_{post}(\theta|\mu)$ provides a degree of plausibility that the phase shift θ equals the interferometer phase θ_0 when the data μ was obtained. This allows the Bayesian approach to provide statistical information for any number of measurements, even a single one. To be sure, this is not a sign of failure or superiority of one approach with respect to the other one, since the two frameworks manipulate conceptually different quantities. The experimentalist can choose to use one or both approaches, keeping in mind the necessity to clearly state the nature of the statistical significance of the reported results.

Table 1. Frequentist vs Bayesian bounds for fixed and random parameters.

	Paradigm	Risk Function	Bounds		Remarks
θ_0 fixed	Frequentist	$(\Delta^2\theta_{\text{est}})_{\mu\|\theta_0}$ $\text{MSE}(\theta_{\text{est}})_{\mu\|\theta_0}$	BB EChRB ChRB CRLB	Equation (5) Equation (17) Equation (14) Equation (8)	hierarchy of bounds, Equation (7)
	Bayesian	$(\Delta^2\theta_{\text{BL}})_{\mu\|\theta_0}$	GB	Equation (22)	function of μ
		$(\Delta^2\theta_{\text{BL}})_{\mu,\theta\|\theta_0}$	aGB	Equation (25)	average over likelihood $p(\mu\|\theta_0)$
θ_0 random	Frequentist	$(\Delta^2\theta_{\text{est}})_{\mu,\theta_0}$ $\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0}$	aCRLB fVTB VTB ZZB	Equation (37) Equation (38) Equation (32) Equation (35)	hierarchy of bounds, Equation (40) bounds are independent of the bias
	Bayesian	$(\Delta^2\theta_{\text{BL}})_{\mu,\theta,\theta_0}$	aGBr	Equation (42)	prior $p_{\text{pri}}(\theta)$ and fluctuations $p(\theta_0)$ arbitrary
		$(\Delta^2\theta_{\text{BL}})_{\mu,\theta}$	VTB ZZB	Equation (32) Equation (35)	prior $p_{\text{pri}}(\theta)$ and fluctuations $p(\theta_0)$ coincide hierarchy of bounds, Equation (45)

The two predictions converge asymptotically in the limit of a large number of measurements. This does not mean that in this limit the significance of the two approaches is interchangeable (it cannot be stated that in the limit of large repetition of the measurements, frequentist ad Bayesian provide the same results). In this respect, it is quite instructive to notice that the Bayesian 2σ confidence may be below that of the Cramér–Rao bound, as shown in Figure 3. This, at first sight, seems paradoxical, since the CRLB is a theorem about the minimum error achievable in parameter estimation theory. However, the CRLB is a frequentist bound and, again, the paradox is solved taking it into account that the frequentist and the Bayesian approaches provide information about different quantities.

Finally, a different class of estimation problems with different precision bounds is encountered if θ_0 is itself a random variable. In this case, the frequentist bounds for the mean-square error (Van Trees, Ziv–Zakai) become independent of the bias, while those on the estimator variance are still functions of the bias. The Van Trees and Ziv–Zakai bounds can be applied to the Bayesian paradigm if the average of the posterior variance over the marginal distribution is the relevant risk function. This is only meaningful if the prior $p_{\text{pri}}(\theta)$ that enters the Bayes–Laplace theorem coincides with the actual distribution $p(\theta_0)$ of the phase shift θ_0.

We conclude with a remark regarding the so-called Heisenberg limit, which is a saturable lower bound on the CRLB over arbitrary quantum states with a fixed number of particles. For instance, for a collection of N two-level systems, the CRLB can be further bounded by $\Delta\theta_{\text{est}} \geq 1/\sqrt{mF(\theta_0)} \geq 1/(\sqrt{m}N)$ [18,20]. This bound is often called the ultimate precision bound since no quantum state is able to achieve a tighter scaling than N. From the discussions presented in this article, it becomes apparent that Bayesian approaches (as discussed in Section 3) or precision bounds for random parameters (Section 4) are expected to lead to entirely different types of 'ultimate' lower bounds. Such bounds are interesting within the respective paradigm for which they are derived, but they cannot replace or improve the Heisenberg limit since they address fundamentally different scenarios that cannot be compared in general.

Author Contributions: Y.L., L.P., M.G., W.L. and A.S. conceived the study, performed theoretical calculations and drafted the article. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivation of the Barankin Bound

Let θ_{est} be an arbitrary estimator for θ. Its mean value

$$\langle\theta_{\text{est}}\rangle_{\mu|\theta} = \sum_{\mu}\theta_{\text{est}}(\mu)p(\mu|\theta) \tag{A1}$$

coincides with θ if and only if the estimator is unbiased (for arbitrary values of θ). In the following, we make no assumption about the bias of θ_{est} and therefore do not replace $\langle\theta_{\text{est}}\rangle_{\mu|\theta}$ by θ.

Introducing the likelihood ratio

$$L(\mu|\theta_i, \theta_0) = \frac{p(\mu|\theta_i)}{p(\mu|\theta_0)} \tag{A2}$$

under the condition $p(\mu|\theta_0) > 0$ for all μ, we obtain with Equation (A1) that

$$\sum_{\mu}\theta_{\text{est}}(\mu)L(\mu|\theta_i, \theta_0)p(\mu|\theta_0) = \langle\theta_{\text{est}}\rangle_{\mu|\theta_i}, \tag{A3}$$

for an arbitrary family of phase values $\theta_1, \ldots, \theta_n$ picked from the parameter domain. Furthermore, we have

$$\sum_{\mu}L(\mu|\theta_i, \theta_0)p(\mu|\theta_0) = \sum_{\mu}p(\mu|\theta_i) = 1 \tag{A4}$$

for all θ_i. Multiplying both sides of Equation (A4) with $\langle\theta_{\text{est}}\rangle_{\mu|\theta_0}$ and subtracting it from (A3) yields

$$\sum_{\mu}\left(\theta_{\text{est}}(\mu) - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right)L(\mu|\theta_i, \theta_0)p(\mu|\theta_0) = \langle\theta_{\text{est}}\rangle_{\mu|\theta_i} - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}. \tag{A5}$$

Let us now pick a family of n finite coefficients $a_1, \ldots, a_n$. From Equation (A5), we obtain

$$\sum_{\mu}\left(\theta_{\text{est}}(\mu) - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right)\left(\sum_{i=1}^{n}a_iL(\mu|\theta_i, \theta_0)\right)p(\mu|\theta_0) = \sum_{i=1}^{n}a_i\left(\langle\theta_{\text{est}}\rangle_{\mu|\theta_i} - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right). \tag{A6}$$

The Cauchy–Schwarz inequality now yields

$$\left(\sum_{i=1}^{n}a_i\left(\langle\theta_{\text{est}}\rangle_{\mu|\theta_i} - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right)\right)^2 \leq \left(\Delta^2\theta_{\text{est}}\right)_{\mu|\theta_0}\left(\sum_{\mu}\left(\sum_{i=1}^{n}a_iL(\mu|\theta_i, \theta_0)\right)^2p(\mu|\theta_0)\right), \tag{A7}$$

where

$$\left(\Delta^2\theta_{\text{est}}\right)_{\mu|\theta_0} = \sum_{\mu}\left(\theta_{\text{est}}(\mu) - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right)^2p(\mu|\theta_0) \tag{A8}$$

is the variance of the estimator θ_{est}. We thus obtain

$$\left(\Delta^2\theta_{\text{est}}\right)_{\mu|\theta_0} \geq \frac{\left(\sum_{i=1}^{n}a_i\left(\langle\theta_{\text{est}}\rangle_{\mu|\theta_i} - \langle\theta_{\text{est}}\rangle_{\mu|\theta_0}\right)\right)^2}{\sum_{\mu}\left(\sum_{i=1}^{n}a_iL(\mu|\theta_i, \theta_0)\right)^2p(\mu|\theta_0)}, \tag{A9}$$

for all n, a_i, and θ_i. The Barankin bound then follows by taking the supremum over these variables.

Appendix B. Derivation of the Ziv–Zakai Bound

Derivations of the Ziv–Zakai bound can be found in the literature (see, for instance, Refs. [24,58,59]). This Appendix follows these derivations closely and provides additional background, which may be useful for readers less familiar with the field of hypothesis testing.

Let $X \in [0, a]$ be a random variable with probability density $p(x)$. We can formally write $p(x) = -dP(X \geq x)/dx$, where $P(X \geq x) \equiv \int_x^a p(y)dy$ is the probability that X is larger or equal than x. We obtain from integration by parts

$$
\begin{aligned}
\langle X^2 \rangle = \int_0^a x^2 p(x)dx &= -\left[x^2 P(X \geq x) \right]_0^a + 2\int_0^a P(X \geq x)x\,dx \\
&= 2\int_0^a P(X \geq x)x\,dx \\
&= \frac{1}{2}\int_0^{2a} P\left(X \geq \frac{h}{2} \right) h\,dh,
\end{aligned}
\tag{A10}
$$

where we assume that a is finite [if $a \to \infty$ the above relation holds when $\lim_{a \to \infty} a^2 P(X \geq a) = 0$]. Finally, we can formally extend the above integral up to ∞ since $P(X \geq a) = 0$:

$$
\langle X^2 \rangle = \frac{1}{2}\int_0^\infty P\left(X \geq \frac{h}{2} \right) h\,dh.
\tag{A11}
$$

Following Ref. [59], we now take $\epsilon = \theta_{\text{est}}(\mu) - \theta_0$ and $X = |\epsilon|$. We thus have

$$
\text{MSE}(\theta_{\text{est}})_{\mu,\theta_0} = \langle |\epsilon|^2 \rangle = \frac{1}{2}\int_0^\infty P\left(|\epsilon| \geq \frac{h}{2} \right) h\,dh.
\tag{A12}
$$

We express the probability as

$$
\begin{aligned}
P\left(|\epsilon| \geq \frac{h}{2} \right) &= P\left(\epsilon > \frac{h}{2} \right) + P\left(\epsilon \leq -\frac{h}{2} \right) \\
&= P\left(\theta_{\text{est}}(\mu) - \theta_0 > \frac{h}{2} \right) + P\left(\theta_{\text{est}}(\mu) - \theta_0 \leq -\frac{h}{2} \right) \\
&= \int P\left(\theta_{\text{est}}(\mu) - \theta_0 > \frac{h}{2} \middle| \theta_0 \right) p(\theta_0)d\theta_0 + \int P\left(\theta_{\text{est}}(\mu) - \theta_0 \leq -\frac{h}{2} \middle| \theta_0 \right) p(\theta_0)d\theta_0.
\end{aligned}
$$

Next, we replace θ_0 with $\theta_0 + h$ in the second integral:

$$
\begin{aligned}
P\left(|\epsilon| \geq \frac{h}{2} \right) &= \int P\left(\theta_{\text{est}}(x) - \theta_0 > \frac{h}{2} \middle| \theta_0 \right) p(\theta_0)d\theta_0 + \int P\left(\theta_{\text{est}}(x) - \theta_0 \leq \frac{h}{2} \middle| \theta_0 + h \right) p(\theta_0 + h)d\theta_0 \\
&= \int (p(\varphi) + p(\varphi + h)) \left[\frac{p(\varphi)}{p(\varphi) + p(\varphi + h)} P\left(\theta_{\text{est}}(x) - \varphi > \frac{h}{2} \middle| \theta_0 = \varphi \right) + \right. \\
&\quad \left. + \frac{p(\varphi + h)}{p(\varphi) + p(\varphi + h)} P\left(\theta_{\text{est}}(x) - \varphi \leq \frac{h}{2} \middle| \theta_0 = \varphi + h \right) \right] d\varphi.
\end{aligned}
$$

We now take a closer look at the expression within the angular brackets and interpret it in the framework of hypothesis testing. Suppose that we try to discriminate between the two cases $\theta_0 = \varphi$ (hypothesis 1, denoted H_1) and $\theta_0 = \varphi + h$ (denoted H_2). We decide between the two hypothesis H_1 and H_2 on the basis of the measurement result x using the estimator $\theta_{\text{est}}(x)$. One possible strategy consists in choosing the hypothesis whose value is closest to the obtained estimator. Hence, if $\theta_{\text{est}}(x) \leq \varphi + h/2$, we assume H_1 to be correct and, otherwise, if $\theta_{\text{est}}(x) > \varphi + h/2$, we pick H_2.

Let us now determine the probability to make an erroneous decision using this strategy. There are two scenarios that will lead to a mistake. First, our strategy fails whenever $\theta_{\text{est}}(x) \leq \varphi + h/2$ when $\theta_0 = \varphi + h$. In this case, H_2 is true, but our strategy leads us to choose H_1. The probability for this to

happen, given that $\theta_0 = \varphi + h$, is $P(\theta_{est}(x) - \varphi \leq \frac{h}{2} | \theta_0 = \varphi + h)$. To obtain the probability error of our strategy, we need to multiply this with the probability with which θ_0 assumes the value $\varphi + h$, which is given by $p(H_2) = \frac{p(\varphi+h)}{p(\varphi)+p(\varphi+h)}$. Second, our strategy also fails if $\theta_{est}(x) > \varphi + h/2$ for $\theta_0 = \varphi$. This occurs with the conditional probability $P(\theta_{est}(x) - \varphi > \frac{h}{2} | \theta_0 = \varphi)$, and $\theta_0 = \varphi$ with probability $p(H_1) = \frac{p(\varphi)}{p(\varphi)+p(\varphi+h)}$. The total probability to make a mistake is consequently given by

$$
\begin{aligned}
P_{err}(\varphi, \varphi + h) &= P\left(\theta_{est}(x) - \varphi > \frac{h}{2} \middle| H_1\right) p(H_1) + P\left(\theta_{est}(x) - \varphi \leq \frac{h}{2} \middle| H_2\right) p(H_2) \\
&= \frac{p(\varphi)}{p(\varphi) + p(\varphi + h)} P\left(\theta_{est}(x) - \varphi > \frac{h}{2} \middle| \theta_0 = \varphi\right) + \\
&\quad + \frac{p(\varphi + h)}{p(\varphi) + p(\varphi + h)} P\left(\theta_{est}(x) - \varphi \leq \frac{h}{2} \middle| \theta_0 = \varphi + h\right),
\end{aligned}
\tag{A13}
$$

and we can rewrite Equation (A13) as

$$
P\left(|\epsilon| \geq \frac{h}{2}\right) = \int_{-\infty}^{\infty} (p(\varphi) + p(\varphi + h)) P_{err}(\varphi, \varphi + h) d\varphi.
\tag{A14}
$$

The strategy described above depends on the estimator θ_{est} and may not be optimal. In general, a binary hypothesis testing strategy can be characterized in terms of the separation of the possible values of x into the two disjoint subsets X_1 and X_2 which are used to choose hypothesis H_1 or H_2, respectively. That is, if $x \in X_1$ we pick H_1 and otherwise H_2. Since one of the two hypotheses must be true, we have

$$
\begin{aligned}
1 &= p(H_1) + p(H_2) \\
&= \int_{X_1} dx\, p(x|H_1) p(H_1) + \int_{X_2} dx\, p(x|H_1) p(H_1) + \int_{X_1} dx\, p(x|H_2) p(H_2) + \int_{X_2} dx\, p(x|H_2) p(H_2) \\
&= \int_{X_1} dx\, p(x|H_1) p(H_1) + \int_{X_2} dx\, p(x|H_2) p(H_2) + P_{err}^{X_1}(H_1, H_2),
\end{aligned}
\tag{A15}
$$

where the error made by such a strategy is given by

$$
\begin{aligned}
P_{err}^{X_1}(H_1, H_2) &= P(x \in X_2 | H_1) p(H_1) + P(x \in X_1 | H_2) p(H_2) \\
&= \int_{X_2} p(x|H_1) p(H_1) dx + \int_{X_1} p(x|H_2) p(H_2) dx \\
&= p(H_1) + \int_{X_1} [p(x|H_2) p(H_2) - p(x|H_1) p(H_1)]\, dx.
\end{aligned}
\tag{A16}
$$

This probability is minimized if $p(x|H_2) p(H_2) < p(x|H_1) p(H_1)$ for $x \in X_1$ and, consequently, $p(x|H_2) p(H_2) \geq p(x|H_1) p(H_1)$ for $x \in X_2$. This actually identifies an optimal strategy for hypothesis testing, known as the likelihood ratio test: if the likelihood ratio $p(x|H_1)/p(x|H_2)$ is larger than the threshold value $p(H_2)/p(H_1)$, we pick H_1, whereas, if it is smaller, we pick H_2. With this choice, the error probability is minimal and reads

$$
\begin{aligned}
P_{min}(H_1, H_2) &= \int_{X_2} [p(x|H_1) p(H_1) - p(x|H_2) p(H_2)]\, dx + \int_{X_1} [p(x|H_2) p(H_2) - p(x|H_1) p(H_1)]\, dx + \\
&\quad + \int_{X_1} p(x|H_1) p(H_1) dx + \int_{X_2} p(x|H_2) p(H_2) dx \\
&= \frac{1}{2} - \frac{1}{2} \int |p(x|H_1) p(H_1) - p(x|H_2) p(H_2)|\, dx,
\end{aligned}
\tag{A17}
$$

where we used Equation (A15).

Applied to our case, we obtain

$$P_{\min}(\varphi, \varphi + h) = \frac{1}{2}\left(1 - \sum_{\boldsymbol{\mu}}\left|\frac{p(\boldsymbol{\mu}|\theta_0 = \varphi)p(\varphi)}{p(\varphi) + p(\varphi + h)} - \frac{p(\boldsymbol{\mu}|\theta_0 = \varphi + h)p(\varphi + h)}{p(\varphi) + p(\varphi + h)}\right|\right). \tag{A18}$$

This result represents a lower bound on $P_{\text{err}}^{X_1}(\varphi, \varphi + h)$ for arbitrary choices of X_1. This includes the case discussed in Equation (A13). Thus, using

$$P_{\text{err}}(\varphi, \varphi + h) \geq P_{\min}(\varphi, \varphi + h) \tag{A19}$$

in Equation (A14) and inserting back into Equation (A12), we finally obtain the Ziv–Zakai bound for the mean square error:

$$\text{MSE}(\theta_{\text{est}})_{\boldsymbol{\mu},\theta_0} \geq \frac{1}{2}\int_0^\infty h\,dh\int d\theta_0(p(\theta_0) + p(\theta_0 + h))P_{\min}(\theta_0, \theta_0 + h). \tag{A20}$$

This bound can be further sharpened by introducing a valley-filling function [61], which is not considered here.

References

1. Zehnder, L. Ein neuer Interferenzrefraktor. *Zeitschrift für Instrumentenkunde* **1891**, *11*, 275. (In German)
2. Mach, L. Ueber einen Interferenzrefraktor. *Zeitschrift für Instrumentenkunde* **1892**, *12*, 89. (In German)
3. Ramsey, N.F. *Molecular Beams*; Oxford University Press: London, UK, 1963.
4. Wynands, R. Atomic Clocks. In *Lecture Notes in Physics*; Muga, G., Ruschhaupt, A., Campo, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 789.
5. Barish, B.C.; Weiss, R. LIGO and the Detection of Gravitational Waves. *Phys. Today* **1999**, *52*, 44–50. [CrossRef]
6. Pitkin, M.; Reid, S.; Rowan, S.; Hough, J. Gravitational Wave Detection by Interferometry (Ground and Space). *Living Rev. Relativ.* **2011**, *14*, 5. [CrossRef] [PubMed]
7. Helstrom, C.W. Quantum detection and estimation theory. *J. Stat. Phys.* **1969**, *1*, 231. [CrossRef]
8. Holevo, A.S. *Probabilistic and Statistical Aspects of Quantum Theory*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1982.
9. Ludlow, A.D.; Boyd, M.M.; Ye, J.; Peik, E.; Schmidt, P.O. Optical atomic clocks. *Rev. Mod. Phys.* **2015**, *87*, 637–701. [CrossRef]
10. Schnabel, R.; Mavalvala, N.; McClelland, D.E.; Lam, P.K. Quantum metrology for gravitational wave astronomy. *Nat. Commun.* **2010**, *1*, 121. [CrossRef] [PubMed]
11. Aasi, J.; Abadie, J.; Abbott, B.P.; Abbott, R.; Abbott, T.D.; Abernathy, M.R.; Adams, C.; Adams, T.; Addesso, P.; Adhikari, R.X. et al. Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nat. Photon.* **2010**, *7*, 613–619. [CrossRef]
12. Cronin, A.D.; Schmiedmayer, J.; Pritchard, D.E. Optics and interferometry with atoms and molecules. *Rev. Mod. Phys.* **2009**, *81*, 1051–1129. [CrossRef]
13. Caves, C.M. Quantum-mechanical noise in an interferometer. *Phys. Rev. D* **1981**, *23*, 1693–1708. [CrossRef]
14. Giovannetti, V.; Lloyd, S.; Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **2006**, *96*, 010401. [CrossRef] [PubMed]
15. Pezzè, L.; Smerzi, A. Entanglement, nonlinear dynamics, and the Heisenberg limit. *Phys. Rev. Lett.* **2009**, *102*, 100401. [CrossRef] [PubMed]
16. Hyllus, P.; Laskowski, W.; Krischek, R.; Schwemmer, C.; Wieczorek, W.; Weinfurter, H.; Pezzè, L.; Smerzi, A. Fisher information and multiparticle entanglement. *Phys. Rev. A* **2012**, *85*, 022321. [CrossRef]
17. Tóth, G. Multipartite entanglement and high-precision metrology. *Phys. Rev. A* **2012**, *85*, 022322. [CrossRef]
18. Pezzè, L.; Smerzi, A. Quantum theory of phase estimation. In *Atom Interferometry, Proceedings of the International School of Physics "Enrico Fermi", Italy, 15–20 July 2013*; Tino, G.M., Kasevich, M.A., Eds.; IOS Press: Sesto Fiorentino, Italy, 2014; *Course 188*, 691.

19. Tóth, G.; Apellaniz, I. Quantum metrology from a quantum information science perspective. *J. Phys. A Math. Theor.* **2014**, *47*, 424006. [CrossRef]

20. Giovannetti, V.; Lloyd, S.; Maccone, L. Advances in quantum metrology. *Nat. Photon.* **2011**, *5*, 222–229. [CrossRef]

21. Pezzè, L.; Smerzi, A.; Oberthaler, M.K.; Schimed, R.; Treutlein, P. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.* **2018**, in press.

22. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory, Volume I*; Prentice Hall: Upper Saddle River, NJ, USA, 1993.

23. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*; Springer: Berlin, Germany, 1998.

24. Van Trees, H.L.; Bell, K.L. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*; Wiley: New York, NY, USA, 2007.

25. Lane, A.S.; Braunstein, S.L.; Caves, C.M. Maximum-likelihood statistics of multiple quantum phase measurements. *Phys. Rev. A* **1993**, *47*, 1667. [CrossRef] [PubMed]

26. Tsang, M. Ziv–Zakai error bounds for quantum parameter estimation. *Phys. Rev. Lett.* **2012**, *108*, 230401. [CrossRef] [PubMed]

27. Lu, X.M.; Tsang, M. Quantum Weiss-Weinstein bounds for quantum metrology. *Quantum Sci. Technol.* **2016**, *1*, 015002. [CrossRef]

28. Hall, M.J.; Wiseman, H.M. Heisenberg-style bounds for arbitrary estimates of shift parameters including prior information. *New J. Phys.* **2012**, *14*, 033040. [CrossRef]

29. Giovannetti, V.; Maccone, L. Sub-Heisenberg estimation strategies are ineffective. *Phys. Rev. Lett.* **2012**, *108*, 210404. [CrossRef] [PubMed]

30. Pezzè, L. Sub-Heisenberg phase uncertainties. *Phys. Rev. A* **2013**, *88*, 060101(R). [CrossRef]

31. Pezzè, L.; Hyllus, P.; Smerzi, A. Phase-sensitivity bounds for two-mode interferometers. *Phys. Rev. A* **2015**, *91*, 032103. [CrossRef]

32. Hradil, Z.; Myška, R.; Peřina, J.; Zawisky, M.; Hasegawa, Y.; Rauch, H. Quantum phase in interferometry. *Phys. Rev. Lett.* **1996**, *76*, 4295. [CrossRef] [PubMed]

33. Pezzè, L.; Smerzi, A; Khoury, G.; Hodelin, J.F.; Bouwmeester, D. Phase detection at the quantum limit with multiphoton mach-zehnder interferometry. *Phys. Rev. Lett.* **2007**, *99*, 223602. [CrossRef] [PubMed]

34. Kacprowicz, M.; Demkowicz-Dobrzanski, R.; Wasilewski, W; Banaszek, K.; Walmsley, I. A. Experimental quantum-enhanced estimation of a lossy phase shift. *Nat. Photon.* **2010**, *4*, 357. [CrossRef]

35. Krischek, R.; Schwemmer, C.; Wieczorek, W.; Weinfurter, H.; Hyllus, P.; Pezzè, L.; Smerzi, A. Useful multiparticle entanglement and sub-shot-noise sensitivity in experimental phase estimation. *Phys. Rev. Lett.* **2011**, *107*, 080504. [CrossRef] [PubMed]

36. Xiang, G.Y.; Higgins, B.L.; Berry, D.W.; Wiseman, H.M.; Pryde, G.J. Entanglement-enhanced measurement of a completely unknown optical phase. *Nat. Photon.* **2011**, *5*, 43–47. [CrossRef]

37. Bollinger, J.J.; Itano, W.M.; Wineland, D.J.; Heinzen, D.J. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **1996**, *54*, R4649–R4652. [CrossRef] [PubMed]

38. Pezzè, L.; Smerzi, A. Sub shot-noise interferometric phase sensitivity with beryllium ions Schrödinger cat states. *Europhys. Lett.* **2007**, *78*, 30004. [CrossRef]

39. Gerry, C.C.; Mimih, J. The parity operator in quantum optical metrology. *Contemp. Phys.* **2010**, *51*, 497. [CrossRef]

40. Sackett, C.A.; Kielpinski, D.; King, B.E.; Langer, C.; Meyer, V.; Myatt, C.J.; Rowe, M.; Turchette, Q.A.; Itano, W.M.; et al. Experimental entanglement of four particles. *Nature* **2000**, *404*, 256–259. [CrossRef] [PubMed]

41. Monz, T.; Schindler, P.; Barreiro, J.T.; Chwalla, M.; Nigg, D.; Coish, W.; Harlander, M. Hänsel, W.; Hennrich, M.; Blatt, R. 14-Qubit Entanglement: Creation and Coherence. *Phys. Rev. Lett.* **2011**, *106*, 130506. [CrossRef] [PubMed]

42. Hayashi, M. *Asymptotic Theory of Quantum Statistical Inference, Selected Papers*; World Scientific Publishing: Singapore, 2005.

43. Barankin, E.W. Locally best unbiased estimates. *Ann. Math. Stat.* **1949**, *20*, 477. [CrossRef]

44. Mcaulay, R.J.; Hofstetter, E.M. Barankin bounds on parameter estimation. *IEEE Trans. Inf. Theory* **1971**, *17*, 669–676. [CrossRef]

45. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.

46. Rao, C.R. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. *Bull. Calcutta Math. Soc.* **1971**, *37*, 81–91.

47. Hammersley, J.M. On estimating restricted parameters. *J. R. Stat. Soc. Ser. B* **1950**, *12*, 192.

48. Chapman, D.G.; Robbins, H. Minimum variance estimation without regularity assumptions. *Ann. Math. Stat.* **1951**, *22*, 581. [CrossRef]

49. Pflanzagl, J.; Hamböker, R. *Parametric Statistical Theory*; De Gruyter: Berlin, Germany, 1994.

50. Sivia, D.S.; Skilling, J. *Data Analysis: A Bayesian Tutorial*; Oxford University Press: London, UK, 2006.

51. Robert, C.P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*; Springer: New York, NY, USA, 2007.

52. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A* **1946**, *186*, 453. [CrossRef]

53. Jeffreys, H. *Theory of Probability*; Oxford University Press: London, UK, 1961.

54. Ghosh, M. Cramér–Rao bounds for posterior variances. *Stat. Probabil. Lett.* **1993**, *17*, 173. [CrossRef]

55. Cam, L.L. *Asymptotic Methods in Statistical Decision Theory*; Springer: New York, NY, USA, 1986.

56. Van Trees, H.L. *Detection, Estimation, and Modulation Theory, Part I*; Wiley: New York, NY, USA, 1968.

57. Schutzenberger, M.P. A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation. *Bull. Am. Math. Soc.* **1957**, *63*, 142.

58. Ziv, J.; Zakai, M. Some lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theor.* **1969**, *15*, 386–391. [CrossRef]

59. Bell, K.L.; Steinberg, Y.; Ephraim, Y.; Van Trees, H.L. Extended Ziv–Zakai lower bound for vector parameter estimation. *IEEE Trans. Inf. Theor.* **1997**, *43*, 624–637. [CrossRef]

60. Gessner, M.; Smerzi, A. Statistical speed of quantum states: Generalized quantum Fisher information and Schatten speed. *Phys. Rev. A* **2018**, *97*, 022109. [CrossRef]

61. Bellini, S.; Tartara, G. Bounds on error in signal parameter estimation. *IEEE Trans. Commun.* **1974**, *22*, 340–342. [CrossRef]