Precision-guaranteed quantum metrology

Takanori Sugiyama[*]

Institute for Theoretical Physics, ETH Zurich, Wolfgang-Pauli-Strasse 27, CH-8093 Zurich, Switzerland

Quantum metrology is a general term for methods to precisely estimate the value of an unknown parameter by actively using quantum resources. In particular, some classes of entangled states can be used to significantly suppress the estimation error. Here we derive a formula for rigorously evaluating an upper bound for the estimation error in a general setting of quantum metrology with arbitrary finite data sets. Unlike in the standard approach, where lower bounds for the error are evaluated in an ideal setting with almost infinite data, our method rigorously guarantees the estimation precision in realistic settings with finite data. We also prove that our upper bound shows the Heisenberg limit scaling whenever the linearized uncertainty, which is a popular benchmark in the standard approach, shows it. As an example, we apply our result to a Ramsey interferometer, and numerically show that the upper bound can exhibit the quantum enhancement of precision for finite data.

PACS number(s): 06.20.Dk, 03.65.Wj, 02.50.Tt, 03.67.−a

I. INTRODUCTION

High-precision measurement is one of the most important techniques for developing science and technology. Quantum metrology is a general term for methods to precisely estimate the value of an unknown parameter by actively using quantum resources like entanglement and squeezing [1–3]. For example, when we use a separable state on an N-partite system in a Ramsey interferometer, an estimation error of phase $\delta\phi$ scales as $O(1/\sqrt{N})$ (the standard quantum limit, SQL). On the other hand, when we use an entangled state like a Greenberger-Horne-Zeilinger (GHZ) state with the same number of particles, $\delta\phi$ scales as $O(1/N)$ (the Heisenberg limit, HL). Such quantum enhancement of precision has been experimentally achieved in several quantum systems like quantum optics [4], ions [5], and atoms [6].

One of the main goal of quantum metrology theory is to derive a fundamental lower bound on the estimation error. So far many different benchmarks for the estimation error have been proposed and analyzed [3]. The most popular benchmark is the root mean squared error (RMSE), and there are two standard approaches for analyzing the RMSE. One is to apply a linear approximation of an estimation method to the RMSE. The approximated RMSE is called a linearized uncertainty (LU). The other is to analyze the classical and quantum Cramér-Rao bounds (CRBs), which are lower bounds on the RMSE for a class of estimation methods. LU and CRBs show the SQL scaling for separable states and the HL scaling for some entangled states.

From a theoretical viewpoint, LU and CRBs both are interesting and important quantities. From an experimental viewpoint, however, there are two problems with the use of these quantities.

First problem is about their ϕ dependency. LU and CRB are functions of the parameter to be estimated. The true value of the parameter is unknown in experiments, which is the reason why we try to estimate it. This means that we cannot know the exact values of LU and CRB in experiments, although it is possible to estimate their values from experimental data.

Second problem is about their invalidity for finite data. In experiments an amount of available data is finite. A linear approximation is used in the derivation of the LU, while many estimation methods used in quantum metrology like a maximum-likelihood estimator are nonlinear functions of data, and the nonlinearity is not negligible for finite data.

CRBs are lower bounds of the RMSE, and they are not attainable when the amount of data is finite [7]. Unattainable lower bounds on an estimation error cannot be used to guarantee an estimation precision.

Because the final goal of quantum metrology experiments is a highly precise estimation of an unknown parameter, it is best to *rigorously* guarantee an estimation precision, if possible. In order to do that, we need an upper bound on an estimation error satisfying two conditions: (1) be independent of the unknown parameter ϕ, and (2) be valid for finite data.

In this paper we derive an upper bound satisfying these two conditions for a general setting in quantum metrology. In Sec. II we explain the setting, notation, and our approach. In Sec. III we introduce an estimation method called a least squares estimator and give a theorem about the estimator. The upper bound shown in the theorem makes it possible to rigorously guarantee the estimation precision in experiments with finite data, which is not possible by the standard approach of quantum metrology theory. We sketch the proof, and the details are given in Appendix B. In Sec. IV we prove that the upper bound shows the scaling the same as the LU, which means that the upper bound shows the HL scaling whenever the LU shows it. As an example, we apply our method to a Ramsey interferometer with N atoms, and perform Monte Carlo simulations for $N = 1$–100. The numerical results indicate that the upper bound can exhibit the quantum enhancement of precision for finite data. In Sec. V we discuss how to treat known and unknown systematic errors in our approach. We summarize this paper in Sec. VI.

II. PRELIMINARIES

A. Procedures and assumptions

We consider the following procedure of quantum metrology (Fig. 1).

---

[*]Present address: Graduate School of Engineering Science, Osaka University, Japan; sugiyama@sigmath.es.osaka-u.ac.jp
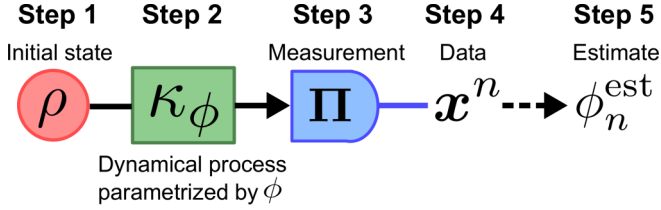
FIG. 1. (Color online) A procedure for quantum metrology: First, we prepare a known quantum state ρ on a probe system. Second, the state undergoes a dynamical process with an unknown parameter ϕ. Third, we perform a known measurement on the probe system and obtain an outcome. Fourth, we repeat the steps described above n times. Then we obtain data consisting of n outcomes. Finally, we calculate an estimate of ϕ from the data with an estimation method.

Step 1. Prepare a known quantum state ρ on a probe system.

Step 2. The state undergoes a dynamical process κ_ϕ with an unknown parameter ϕ. Our aim is to estimate $\phi \in \Phi := [\phi_{\min}, \phi_{\max}]$, where $\phi_{\min}$ and $\phi_{\max}$ are upper and lower values of possible ϕ and are assumed to be known.

Step 3. After the dynamical process, the state changes to a state $\rho_\phi = \kappa_\phi(\rho)$, which depends on ϕ. We perform a known measurement on the state and obtain a measurement outcome.

Step 4. Repeat steps 1 to 3 a number n of times [8]. Then we have data consisting of n outcomes, $x^n = \{x_1, \ldots, x_n\}$.

Step 5. Calculate an estimate of the parameter $\phi_n^{\text{est}}(x^n)$ from the data by a data processing method ϕ^{est}. This function from data to a real value is called an estimator.

The measurement performed in step 3 is described by a positive operator-valued measure (POVM) $\mathbf{\Pi} = \{\Pi_x\}_{x \in \mathcal{X}}$, which is not necessarily a projective measurement. We assume that the measurement outcomes are bounded, i.e., $-\infty < a \leqslant x \leqslant b < +\infty$. This assumption is valid in practice, because there are technical cutoffs on observable values of measurement outcomes in any experiment. The probability distribution is given by $p(x|\rho_\phi, \mathbf{\Pi}) = \text{Tr}[\rho_\phi \Pi_x]$. Let $\mathbb{E}[\mathbf{\Pi}|\rho_\phi]$ and $\mathbb{V}[\mathbf{\Pi}|\rho_\phi]$ denote the expectation and variance of the measurement outcome of $\mathbf{\Pi}$, respectively. The expectation is a function of ϕ, and let f denote the function, i.e., $f(\phi) := \mathbb{E}[\mathbf{\Pi}|\rho_\phi]$. We assume that f is injective and that the derivative $\frac{df}{d\phi}$ does not take zero for any $\phi \in \Phi$. Let $\mathcal{R}_f$ denote the range of f, i.e., $\mathcal{R}_f := \{f(\phi)|\phi \in \Phi\}$. Then $g := f^{-1}$ exists for $\mathcal{R}_f$, and we have $\phi = g(\mathbb{E}[\mathbf{\Pi}|\rho_\phi])$. So, if we know the value of the expectation, we can calculate the value of ϕ.

B. Estimators

For given data x^n, we define the sample mean $S_n := \frac{1}{n} \sum_{i=1}^n x_i$, where x_i is the ith outcome in the n outcomes. The sample mean converges to the expectation $\mathbb{E}[\mathbf{\Pi}|\rho_\phi]$ in the limit of n going to infinity (the law of large numbers). It might seem natural to consider a direct inversion estimator $\phi_n^{\text{DI}}(x^n) := g(S_n)$. In general, however, the direct inversion estimator does not work well, because the sample mean is a random variable and can be out of $\mathcal{R}_f$, which is caused by a statistical fluctuation originated from the finiteness of n. The inverse function $g = f^{-1}$ may not exist outside $\mathcal{R}_f$, and we

may not be able to calculate $\phi_n^{\text{DI}}(x^n)$ there. Even if g exists, $\phi_n^{\text{DI}}(x^n)$ can be out of Φ.

One solution to avoid this problem of ϕ^{DI} mentioned above is a maximum-likelihood estimator (MLE) $\phi_n^{\text{ML}}(x^n) := \text{argmin}_{\phi' \in \Phi} \prod_{i=1}^n p(x_i|\rho_{\phi'}, \mathbf{\Pi})$. Unlike ϕ_n^{DI}, an estimate of the MLE always exists and takes a value in Φ. The MLE has good statistical properties, for example, it attains the Cramér-Rao bound in the limit of n going to infinity [9]. The asymptotic ($n \sim \infty$) behavior of the MLE is well known in classical statistics [10], but a rigorous analysis for finite n is an open problem. Instead of the MLE, we consider a different estimator relatively easier to be analyzed.

C. Standard benchmarks

Let us choose an estimator ϕ^{est}. The estimates $\phi_n^{\text{est}}(x^n)$ depend on data and probabilistically fluctuate. This means that we can observe estimates deviated from the true parameter ϕ. This difference is called an estimation error of the estimator. To evaluate an estimation error is an important topic in quantum metrology.

As explained in Sec. I, the most popular benchmark is the root mean squared error (RMSE), which is defined as

$$(\delta\phi)_{\text{RMSE}} := \sqrt{\mathbb{E}\left[\left|\phi_n^{\text{est}} - \phi\right|^2\right]}. \tag{1}$$

Generally speaking, a direct analysis of the RMSE itself is difficult, since the RMSE is a function of the dynamics κ_ϕ and our choice of the initial state ρ, measurement $\mathbf{\Pi}$, and estimator ϕ^{est}. There are two approaches to reduce the degree of this difficulty for analyzing the RMSE.

The first approach is to approximate the RMSE for the DI estimator. When the number of repetition n is sufficiently large, we can approximate the RMSE of the DI estimator as follows:

$$(\delta\phi)_{\text{RMSE}}(\rho, \kappa_\phi, \mathbf{\Pi}, \phi^{\text{est}} = \phi^{\text{DI}}, n) \approx (\delta\phi)_{\text{LU}} := \frac{B_{\text{LU}}}{\sqrt{n}}, \tag{2}$$

where

$$B_{\text{LU}} := \frac{\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}}{\left|\frac{d}{d\phi}\mathbb{E}[\mathbf{\Pi}|\rho_\phi]\right|}. \tag{3}$$

In this approximation the nonlinearity of the DI estimator is ignored. In other words, the right-hand side of Eq. (2) is the RMSE of the linearized DI estimator, and it is called the linearized uncertainty (LU). The details of this approximation are given in Appendix A 2.

An advantage of the LU is that the analysis is easy, because it consists of the variance and derivative of expectation with respect to a single outcome. Estimates of B_{LU} such as

$$\frac{\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}}{\left|\frac{d}{d\phi}\mathbb{E}[\mathbf{\Pi}|\rho_\phi]\right|}\Bigg|_{\phi=\phi_n^{\text{DI}}(x^n)} \tag{4}$$

are calculated in some experiments [11–13].

The second approach is to analyze an asymptotic lower bound of the RMSE, which is independent of our choice of estimator. Let us introduce two classes of estimators in statistical estimation theory. When the expectation of an estimator with respect to n equals the true parameter, the estimator is called unbiased for n. When the derivative of

the expectation converges to one in the limit of n going to infinity, the estimator is called asymptotically unbiased. For any unbiased estimator with respect to n, the following inequality holds under certain normal conditions [9,14,15]:

$$(\delta\phi)_{\text{RMSE}}(\rho,\kappa_\phi,\mathbf{\Pi},\phi^{\text{est}},n) \geqslant 1/\sqrt{nF_{\text{C}}(\phi,\rho,\mathbf{\Pi})} \qquad (5)$$

$$\geqslant 1/\sqrt{nF_{\text{Q}}(\phi,\rho)}, \qquad (6)$$

where F_{C} and F_{Q} are quantities called the classical and quantum Fisher information, respectively. Equations (5) and (6) are called the classical and quantum Cramér-Rao inequality for finite n, respectively. Most of the estimators in quantum metrology, which include DI an MLE, are biased for finite n, which originated from the nonlinear parametrization of probability distributions. This means that the Cramér-Rao inequalities for finite n is not applicable for quantum metrology. However, most "natural" estimators in quantum metrology, which includes DI and MLE again, are asymptotically unbiased. For any asymptotically unbiased estimator, the following inequality holds under certain normal conditions [9,14,15]:

$$\lim_{n\to\infty} \sqrt{n}(\delta\phi)_{\text{RMSE}}(\rho,\kappa_\phi,\mathbf{\Pi},\phi^{\text{est}},n) \geqslant 1/\sqrt{F_{\text{C}}(\phi,\rho,\mathbf{\Pi})} \quad (7)$$

$$\geqslant 1/\sqrt{F_{\text{Q}}(\phi,\rho)}. \quad (8)$$

where Eqs. (7) and (8) are called classical and quantum Cramér-Rao inequalities for asymptotic n, respectively. The MLE attains the classical Cramér-Rao bound (CRB) for asymptotic n [9]. An advantage of the CRBs is their generality for our choice of estimator and measurement. The classical CRB is independent of estimators, and the quantum CRB is independent of measurement. So, classical and quantum CRBs are used for evaluating ultimate performances of a combination of ρ and $\mathbf{\Pi}$ or ρ, respectively.

From a theoretical viewpoint, LU and CRBs are interesting and important quantities. However, they are not suitable for rigorously evaluating an estimation error in experiments with finite n because of the following two reasons. (1) LU and CRBs are functions of the unknown parameter ϕ, and we cannot know their exact values in experiments. Of course we can estimate their values by calculating quantities like Eq. (4), but the calculated values are estimates that can be different from the exact value. (2) LU and CRBs are not valid for finite n as explained in this subsection. When n is sufficiently large, we may be able to validate the use of them, but it is unclear which n can be interpreted as sufficiently large. In order to rigorously evaluate an estimation error for finite n, we need another benchmark.

D. Confidence intervals

Roughly speaking, confidence intervals are intervals including the true parameter with high probability. We propose the size of an confidence interval as a new benchmark in quantum metrology. When an interval I is a function of data and is independent of ϕ, the function is called an interval estimator. We would like to find an interval estimator such that the interval estimates $I(\boldsymbol{x}^n)$ include ϕ with high probability. When $I(\boldsymbol{x}^n) \ni \phi$ holds with probability at least $1-\epsilon$ for any $\phi \in \Phi$, the interval estimator is called a confidence interval with $(1-\epsilon)$-confidence level. For example, $I = \Phi$ is

a confidence interval with 1-confidence level. This example is trivial and useless. We need a nontrivial and useful confidence interval. The following two properties are required for a "nontrivial" and "useful" confidence interval in quantum metrology experiments.

Property 1. Its size converges to zero in the limit of n going to infinity.

Property 2. Its size can show a quantum enhancement when we use a quantum resource in quantum metrology experiments.

In this paper we propose a new confidence interval and prove that it has the two properties mentioned above.

Before moving on to our results, let us note their difference from known results. Confidence interval and confidence level are well known concepts in classical statistics, and there are many statistical techniques to calculate them for finite data [16]. Most of these techniques are, however, based on the normal distribution approximation (NDA), and a confidence interval calculated with the NDA is called an approximate confidence interval. The NDA is valid when the number of measurement trials n is sufficiently large (the central limit theorem), but again, it is not clear which n can be considered as sufficiently large. Therefore, it is not rigorous to apply approximate confidence intervals for finite data in experiments. In contrast to an approximate confidence interval, our confidence interval is calculated without any assumption on probably distribution and is called an exact confidence interval.

An exact confidence region, which is a generalization of confidence interval to higher dimensional spaces, for quantum tomography was proposed in [17]. The estimation object in quantum tomography is quantum state, process, or measurement, which includes multiparameters. Some readers might think that the result in [17] would be applicable for quantum metrology because quantum metrology is an estimation problem of quantum process with a single parameter, but this is not correct. Quantum process tomography and quantum metrology are different problems from statistical viewpoints, and the result for quantum tomography obtained in [17] is not applicable for quantum metrology. The main difference is from the difference of their parametrization of probability distribution. In quantum tomography, the probability distribution of measurement outcome can be linearly parametrized by the estimation object. In quantum metrology, on the other hand, the estimation object is a single parameter, but the parametrization of probability distribution is nonlinear. In general, statistical estimation problems with linearly parametrized and nonlinearly parametrized probability distributions have different statistical properties. For example, the classical CRB for finite data is attainable in the linear case, but it is not attainable in the nonlinear case [18]. Actually, in quantum tomography, there exists an unbiased estimator that attains the equality of the classical Cramér-Rao inequality for any finite data, but in quantum metrology, there do not exist any unbiased estimators that attain the equality for finite data. This is caused by the difference of their parametrizations. Therefore, quantum process tomography and quantum metrology are different problems in statistical estimation. Additionally, the linearity of the parametrization in quantum tomography is used in the derivation of the exact confidence region in [17], and the result is not applicable for quantum metrology. In order to

derive an exact confidence interval that reflects quantumness of resources in quantum metrology, we need new mathematical techniques.

III. RESULTS

We consider the least squares (LS) estimator,

$$\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) := \underset{\phi' \in \Phi}{\operatorname{argmin}} |S_n - f(\phi')|. \qquad (9)$$

Same as the MLE, the LS estimates $\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n)$ always exist and take values in Φ. Let us define

$$S_n^{\mathrm{LS}} := \underset{r' \in \mathcal{R}_f}{\operatorname{argmin}} |r' - S_n|, \qquad (10)$$

$$V_{\max} := (b-a)^2/4, \qquad (11)$$

$$V_n := \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2, \qquad (12)$$

$$L := \max_{\phi' \in \Phi} \left| \left\{ \frac{df}{d\phi}(\phi') \right\}^{-3} \frac{d^2 f}{d\phi^2}(\phi') \right|. \qquad (13)$$

Note that V_n is the unbiased sample variance of the data and satisfies $\mathbb{E}[V_n] = \mathbb{V}[\boldsymbol{\Pi}|\rho_\phi]$. Using the quantities introduced above, we define three functions of data $\boldsymbol{x}^n$ and an user-specified constant ϵ:

$$\delta_1(\boldsymbol{x}^n,\epsilon) := \frac{1}{\left|\frac{df}{d\phi}(\phi_n^{\mathrm{LS}})\right|} \sqrt{\frac{2}{n} V_{\max} \ln \frac{2}{\epsilon}} + \frac{L}{n} V_{\max} \ln \frac{2}{\epsilon}, \quad (14)$$

$$\delta_2(\boldsymbol{x}^n,\epsilon) := \frac{1}{\left|\frac{df}{d\phi}(\phi_n^{\mathrm{LS}})\right|} \left\{ \sqrt{\frac{2}{n} V_n \ln \frac{4}{\epsilon}} + \frac{8(b-a)}{3(n-1)} \ln \frac{4}{\epsilon} \right\}$$
$$+ \frac{L}{2} \left\{ \sqrt{\frac{2}{n} V_n \ln \frac{4}{\epsilon}} + \frac{8(b-a)}{3(n-1)} \ln \frac{4}{\epsilon} \right\}^2, \quad (15)$$

$$\delta(\boldsymbol{x}^n,\epsilon) := \begin{cases} \delta_1(\boldsymbol{x}^n,\epsilon) & \text{if } n = 1, \\ \min\{\delta_1(\boldsymbol{x}^n,\epsilon), \delta_2(\boldsymbol{x}^n,\epsilon)\} & \text{if } n \geqslant 2. \end{cases} \quad (16)$$

The following theorem guarantees that the deviation of the LS estimates from the true parameter is upper bounded by δ with high probability.

Theorem 1. For any number of measurement trials $n \geqslant 1$, user-specified constant $\epsilon \in (0,1)$, and true parameter $\phi \in \Phi$,

$$\left| \phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) - \phi \right| \leqslant \delta(\boldsymbol{x}^n,\epsilon) \qquad (17)$$

holds with probability at least $1 - \epsilon$.

Theorem 1 means that the δ defined in Eq. (16) is a rigorous "error bar" on LS estimates $\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n)$ for arbitrary finite data. In other words, $\phi = \phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) \pm \delta(\boldsymbol{x}^n,\epsilon)$ holds with high probability $1 - \epsilon$, where we choose the value of ϵ as small as we like. The $\delta(\boldsymbol{x}^n,\epsilon)$ becomes larger as we choose smaller ϵ. This means that, if we require a higher confidence level for a fixed n, the error bar becomes larger for safe. If we want to keep the confidence level, we need to increase the number of measurement.

We sketch the proof of Theorem 1, with the details shown in Appendix B. The LS estimator is a nonlinear function of

the sample mean, which is the origin of the main difficulty for the analysis. We use the Taylor expansion up to second order with the remainder, and reduce the problem to an analysis on the deviation of the sample mean from the true expectation $|S_n - f(\phi)|$. In the reduction we use the contractivity of the LS estimator, i.e.,

$$\left| S_n^{\mathrm{LS}} - f(\phi) \right| \leqslant |S_n - f(\phi)|, \quad \forall \boldsymbol{x}^n, \quad \phi \in \Phi. \qquad (18)$$

The contractivity is one of the two main keys in this proof, and this is the reason why we choose the LS estimator. After the reduction we use two inequalities for evaluating $|S_n - f(\phi)|$. One is Hoeffding's inequality [19], which is well known in classical statistics. The other is the empirical Bernstein inequality [20], which is a new mathematical tool developed for finite data analysis in machine learning. The empirical Bernstein inequality is the second key in this proof. It enables us to show a relation to the linearized uncertainty explained later. By combining these inequalities, contractivity, and Taylor expansion, we obtain Theorem 1.

It is important that $\delta(\boldsymbol{x}^n,\epsilon)$ depends only on data $\boldsymbol{x}^n$ and user-specified constant ϵ, and that it is independent of the true parameter ϕ. [The probability distribution of $\delta(\boldsymbol{x}^n,\epsilon)$ depends on ϕ.] So, we can calculate $\delta(\boldsymbol{x}^n,\epsilon)$ without knowing ϕ. Let us introduce a data-dependent interval

$$I_\epsilon(\boldsymbol{x}^n) := \Phi \cap \left[\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) - \delta(\boldsymbol{x}^n,\epsilon), \phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) + \delta(\boldsymbol{x}^n,\epsilon) \right]. \quad (19)$$

Theorem 1 guarantees that this interval estimator I_ϵ is an exact confidence interval with $(1 - \epsilon)$-confidence level. For example, when we choose $\epsilon = 0.01$, we obtain a confidence interval $I_{\epsilon=0.01}(\boldsymbol{x}^n)$ that includes ϕ with probability at least 99%. What we do after step 4 in quantum metrology experiments is to choose a value of ϵ as we like and to calculate the LS estimate $\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n)$ and $\delta(\boldsymbol{x}^n,\epsilon)$ from data obtained. Then we have an estimate of the unknown parameter ϕ with a rigorous error bar.

IV. ANALYSIS

The main purpose of this paper is to propose an exact confidence interval satisfying properties 1 and 2 explained in Sec. II D. By definition of δ in Eq. (16), our new exact confidence interval I_ϵ satisfies property 1. In this section we theoretically and numerically prove that I_ϵ also satisfies property 2. In Sec. IV A we show relations to the LU and quantities calculated in experiments. Especially the relation to the LU indicates that δ shows a quantum enhancement for asymptotically large n whenever the LU shows it. In Sec. IV B we perform a numerical simulation of a Ramsey interferometer. The result indicates that, even for finite n, δ can show the quantum enhancement when a quantum resource is used in quantum metrology.

A. Relation to LU

First, we explain a relation between δ and the LU. By definition, δ_2 decreases as $O(1/\sqrt{n})$, and the coefficient of the dominant term is given by $\sqrt{2V_n \ln \frac{4}{\epsilon}}/|\frac{df}{d\phi}(\phi_n^{\mathrm{LS}})|$. This coefficient converges to $B_{\mathrm{LU}}\sqrt{2 \ln \frac{4}{\epsilon}}$ in the limit of n going to infinity because V_n and ϕ_n^{LS} converge to $\mathbb{V}[\boldsymbol{\Pi}|\rho_\phi]$ and ϕ,

respectively. So, we would expect that δ have the scaling the same as the LU with respect to n and N. Actually we can prove the following inequality:

$$\lim_{n \to \infty} \{\sqrt{n}\mathbb{E}[\delta(\boldsymbol{x}^n, \epsilon)]\} \leqslant B_{\mathrm{LU}}\sqrt{2\ln\frac{4}{\epsilon}}, \qquad (20)$$

where $\mathbb{E}$ denotes the expectation with respect to data $\boldsymbol{x}^n$. The proof is shown in Appendix C. The logic mentioned above and Eq. (20) guarantee that, on average, $\delta(\boldsymbol{x}^n, \epsilon)$ scales the same as the LU. The upper bound δ shows the HL scaling, whenever the LU shows it. This is important especially in noisy cases. The quantum enhancement of precision can be suppressed when the dynamical process κ_ϕ is noisy [21], and recently there are many proposals for recovering the quantum enhancement with respect to the LU [22] and CRB [23–25]. Equation (20) indicates that the recovery method with respect to LU also works well for δ.

Next, we explain a relation between δ and quantities calculated in experiments. As explained in Sec. II C, estimates of B_{LU}, such as Eq. (4), have been calculated in some quantum metrology experiments. Such a quantity also appears in δ. Let define an estimate of B_{LU} as

$$B_{\mathrm{LU}}^{\mathrm{est}} := \frac{\sqrt{V_n}}{\left|\frac{df}{d\phi}\left(\phi_n^{\mathrm{LS}}\right)\right|}. \qquad (21)$$

Then the dominant part of δ_2 is rewritten as $B_{\mathrm{LU}}^{\mathrm{est}}\sqrt{\frac{2}{n}\ln\frac{4}{\epsilon}}$. This means that δ_2 is a sum of an estimate of B_{LU} with a coefficient originated from the value of the confidence level and the higher order, i.e., $O(1/n)$, terms. The coefficient and higher order terms are corrections for guaranteeing the statistical rigorousness. Therefore, our result is not what is totally different from the conventional method in experiments, but it is an extension from such a rough method toward rigorously treating finite data.

B. Example: Ramsey interferometer

We apply our result to a Ramsey interferometer with N atoms. For a separable probe state of N atoms, the LU scales as the SQL scaling $O(1/\sqrt{N})$. On the other hand, for an entangled state like a GHZ state, the LU can scale as the HL scaling $O(1/N)$.

We consider two combinations of initial state and measurement. One is a combination of a separable state $[\frac{1}{\sqrt{2}}(|e\rangle + |g\rangle)]^{\otimes N}$ and the measurement of the total energy, and the other is that of a GHZ state $\frac{1}{\sqrt{2}}(|e\rangle^{\otimes N} + |g\rangle^{\otimes N})$ and the measurement of the parity, where $|e\rangle$ and $|g\rangle$ are excited and ground states of an atom, respectively. In the cases we have $B_{\mathrm{LU}} = 1/\sqrt{N}$ for the separable state and $B_{\mathrm{LU}} = 1/N$ for the GHZ state.

We performed Monte Carlo simulations for the cases with $N = 1$–100, $n = 1$–10000, $\phi_{\min} = 0$, $\phi_{\max} = \pi/400$, $\phi = \pi/4000$, and $\epsilon = 0.1$ (90%-confidence level). The details are given in Appendix D. In order to analyze typical behaviors of δ, we calculated expectations of δ and compared them to expectations of $|\phi_n^{\mathrm{LS}} - \phi|$. Figures 2(a) and 2(b) are the results. In both Figs. 2(a) and 2(b) the vertical axes are for expected deviations. Solid and dashed (black) lines are $\mathbb{E}[\delta(\boldsymbol{x}^n, \epsilon)]$ and $\mathbb{E}[|\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) - \phi|]$ for the separable state, respectively. Chained and dotted (red) lines are $\mathbb{E}[\delta(\boldsymbol{x}^n, \epsilon)]$ and $\mathbb{E}[|\phi_n^{\mathrm{LS}}(\boldsymbol{x}^n) - \phi|]$ for the GHZ state, respectively. The expectations were calculated by a Monte Carlo sampling with 5000 repetitions.

In Fig. 2(a) the horizontal axis is the number of atoms N. Plots in the panel express the scaling of the expected deviations with respect to N with a fixed number of measurement trials, $n = 3000$. The expectations of δ are larger than those of $|\phi_n^{\mathrm{LS}} - \phi|$, which is consistent with Theorem 1. Figure 2(a) also indicates that, up to $N = 100$, the expectation of δ for the GHZ state scales as the HL scaling, although that for the separable state scales as the SQL scaling.

In Fig. 2(b) the horizontal axis is the number of measurement trials n. Plots in the panel express the scaling of the expected deviations with respect to n with a fixed number of
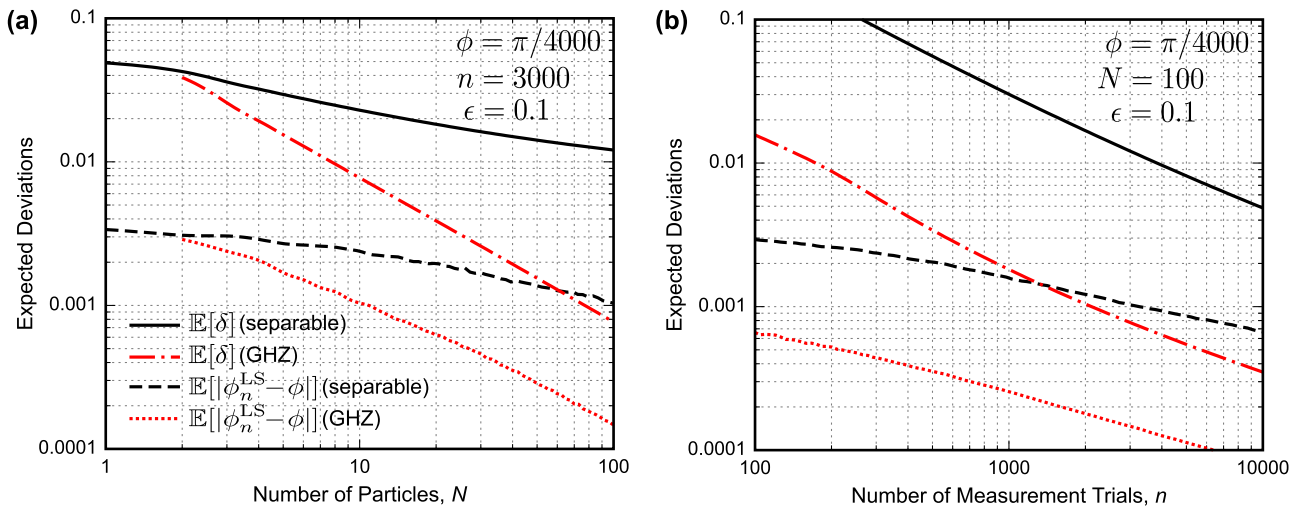

FIG. 2. (Color online) Numerical result on expected deviations ($\mathbb{E}[\delta]$ with $\epsilon = 0.1$, and $\mathbb{E}[|\phi_n^{\mathrm{LS}} - \phi|]$) in a Ramsey interferometer using a separable or GHZ initial state of N atoms. Expectations were calculated by a Monte Carlo sampling with 5000 repetitions. (a) N dependency with $n = 3000$, and (b) n dependency with $N = 100$. Both panels indicate that δ for the GHZ state shows the HL scaling up to $N = 100$ for finite n.

atoms $N = 100$. The expectations of δ for both states scale as $O(1/\sqrt{n})$, and δ for the GHZ state is, on average, $10(=\sqrt{N}$ in the panel) times smaller than δ for the separable state.

In conclusion of the numerical simulations, the expectations of δ with 90%-confidence level are larger than the expectations of the actual deviations for both separable and entangled states, which is consistent with Theorem 1. Furthermore, Fig. 2 indicates that, compared to the separable state, the entangled state gives smaller deviation of estimates and smaller error bar δ. Equation (20) guarantees that δ shows the HL scaling for asymptotically large n whenever the LU shows the scaling, and Fig. 2 indicates that δ can also show the quantum enhancement of precision for finite n. Note that the Ramsey interferometer is mathematically equivalent to a Mach-Zehnder interferometer [26], which means that δ can show the quantum enhancement of precision in an optical interferometer with $N00N$ states.

V. DISCUSSION

In this section we discuss how to apply our result to some cases in which unknown systematic errors exist.

A. Partially unknown systematic errors

In Theorem 1 it is assumed that we perfectly know ρ, $\mathbf{\Pi}$, and the functional form of κ_ϕ. This assumption may not be valid when there exists a systematic error in experiments.

In the standard approach of quantum metrology theory, a model for the systematic error is introduced, and it is assumed that the model correctly characterizes the error and that we know the value of a noise parameter in the model [21–25,27–30]. Theorem 1 is applicable for such a perfectly known systematic error. Even if the model is correct, however, the value that we think of as the noise parameter may be different from the true value in an experiment. Theorem 1 and the standard approach are not directly applicable for such a partially unknown systematic error. However, we can obtain an exact confidence interval for quantum metrology with a partially unknown systematic error, by modifying Theorem 1 based on the worst case of the noise parameter.

Here let us consider the case that the noise is partially unknown, i.e., the noise model is correct, but we do not know the value of a noise parameter in the model. Suppose that there is an imperfection of the preparation of the initial state and that it is characterized by a noise model with a parameter η_1. The time evolution is characterized by the true parameter of interest ϕ and noise parameter η_2. There is an imperfection in the measurement apparatus, and it is characterized by a noise model with a parameter η_3 (η_1, η_2, and η_3 can be multiparameters). Suppose that the noise parameters $\boldsymbol{\eta} := (\eta_1, \eta_2, \eta_3)$ are unknown, but that we know a region E including the true noise parameters, i.e., $\boldsymbol{\eta} \in E$. In this case, the probability distribution of the measurement outcome is given by

$$p(x|\phi, \boldsymbol{\eta}) = \text{Tr}\left[\kappa_{\phi, \eta_2}(\rho_{\eta_1}) \Pi_{x, \eta_3}\right]. \tag{22}$$

We know the function form of the probability distribution, but we do not know the true values of ϕ and $\boldsymbol{\eta}$. Then the functional forms of f and g depend on the values of $\boldsymbol{\eta}$, and the value of δ

depends on $\boldsymbol{\eta}$ as well. To clarify this noise dependency of ϕ_n^{LS} and δ, let us use new notations $\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta})$ and $\delta(x^n, \epsilon, \boldsymbol{\eta})$.

Let $\boldsymbol{\eta}' := (\eta_1', \eta_2', \eta_3')$ denote the values that we think of as the true values of $\boldsymbol{\eta}$. In general, $\boldsymbol{\eta}'$ and $\boldsymbol{\eta}$ are different. We want to evaluate the difference between $\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}')$, which is a LS estimate calculated from data and incorrect noise parameter and ϕ. We have

$$\left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi\right|$$

$$\leqslant \left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi_n^{\text{LS}}(x^n, \boldsymbol{\eta})\right| + \left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}) - \phi\right| \tag{23}$$

$$\leqslant \left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi_n^{\text{LS}}(x^n, \boldsymbol{\eta})\right| + \delta(x^n, \epsilon, \boldsymbol{\eta}) \tag{24}$$

$$\leqslant \max_{\boldsymbol{\eta} \in E} \left\{\left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi_n^{\text{LS}}(x^n, \boldsymbol{\eta})\right| + \delta(x^n, \epsilon, \boldsymbol{\eta})\right\}, \tag{25}$$

where Eq. (24) holds with probability at least $1 - \epsilon$. Let us define

$$\tilde{\delta}(x^n, \epsilon, \boldsymbol{\eta}') := \max_{\boldsymbol{\eta} \in E} \left\{\left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi_n^{\text{LS}}(x^n, \boldsymbol{\eta})\right| + \delta(x^n, \epsilon, \boldsymbol{\eta})\right\}. \tag{26}$$

We obtain the following theorem.

Lemma 1. For any number of measurement trials $n \geqslant 1$ and user-specified constant $0 < \epsilon < 1$,

$$\left|\phi_n^{\text{LS}}(x^n, \boldsymbol{\eta}') - \phi\right| \leqslant \tilde{\delta}(x^n, \epsilon, \boldsymbol{\eta}') \tag{27}$$

holds with probability at least $1 - \epsilon$.
Lemma 1 provides an exact confidence interval for quantum metrology with partially unknown noise.

The first term in the right-hand side of Eq. (26) is the effect of the partially unknown noise. This is a systematic error. The second term in the right-hand side of Eq. (26) corresponds to the statistical error. When a noise is partially unknown and we choose an incorrect value for the noise parameter, any estimator ϕ^{est} cannot converge to the true parameter ϕ. So, when n goes to infinity, δ converges to 0, but $\tilde{\delta}$ does not. To avoid this problem in the case that the noise is partially unknown, we need to estimate the parameter of interest ϕ and noise parameters $\boldsymbol{\eta}$ both. This simultaneous estimation of ϕ and $\boldsymbol{\eta}$ is a theoretically interesting and practically important problem, but it is out of the main topic of this paper.

B. Physical vs statistical models

Here we explain a possible method for treating unknown statistical errors, which is different from the way described in the previous subsection. An experimental setup of quantum metrology is characterized by an initial state ρ, a dynamical process κ_ϕ, and a measurement $\mathbf{\Pi}$. Let us call a set $(\rho, \kappa_\phi, \mathbf{\Pi})$ a *physical model* of the experiment.

Let us call the function form of the expectation f a *statistical model* for the experiment. Recall that the calculation of the LS estimate and δ requires only f. So, if we know the statistical model, we can use Theorem 1 even if we do not perfectly know the physical model.

Our strategy is as follows:

Step 1′. We perform a pre-experiment before starting a quantum metrology experiment for an unknown ϕ. We set a known value of ϕ and perform quantum metrology experiments.

Step 2′. We repeat the pre-experiment for many different known values of ϕ.

Step 3′. We estimate the statistical model f from the data obtained in the pre-experiments.

If the numbers of known ϕs and measurement trials for each known ϕ are sufficiently large, we have a precise estimate of the statistical model f, and we can use the estimate of f instead of the true f in Theorem 1. The method consisting of steps from 1′ to 3′ is exactly the same as experiments for observing interference fringes in quantum metrology [13,31,32]. The precision of estimating f depends on the way of sampling ϕs and the choice of estimator for f. To establish a method for rigorously evaluating a total precision of interference fringe observation and quantum metrology for estimating unknown ϕ after the pre-experiments is an open problem, which is important for practical quantum metrology.

VI. SUMMARY

We considered a general setting of quantum metrology, proposing a least squares estimator and deriving an explicit formula of an exact confidence interval for the estimator with an arbitrary finite number of measurement trials. The explicit formula makes it possible to calculate a *rigorous* error bar δ on the least squares estimates in experiments. We showed that the error bar δ scales the same as the linearized uncertainty, which is a popular benchmark in the standard approach of quantum metrology, for asymptotically large number n of measurement trials. This means that δ asymptotically shows the Heisenberg limit scaling whenever the linearized uncertainty shows the scaling. As an example, we applied our results to a Ramsey interferometer with N atoms and performed Monte Carlo simulations for $N = 1$–100 and $n = 1$–10 000. The numerical result indicates that, when a GHZ state is used as an initial state, δ shows the Heisenberg limit scaling for finite n. It means that δ can also exhibit the quantum entrancement of precision for finite n. To the best of our knowledge, this is the first result that makes it possible to rigorously guarantee an estimation precision in quantum metrology with finite data, and we hope it finds application in the analysis of experimental data.

ACKNOWLEDGMENTS

APPENDIXES

We explain the details of our results. In Appendix A we give a summary of assumptions and the derivation of the linearized uncertainty. In Appendix B we give the proof of Theorem 1. In Appendix C we give the proof of Eq. (20). In Appendix D we explain the details of the Ramsey interferometer and Monte Carlo simulation mentioned in Sec. IV B.

APPENDIX A: NOTATIONS AND ASSUMPTIONS

In this Appendix, for convenience we give a summary of assumptions. We also explain a relation between the RMSE and LU.

1. List of assumptions

Theorem 1 holds under the following four assumptions.

(A1) We know ρ, $\mathbf{\Pi}$, and the functional form of κ_ϕ, i.e., we know f.

(A2) The measurement outcomes are bounded, i.e., $-\infty < a \leqslant x \leqslant b < +\infty$, $\forall x \in \mathcal{X}$.

(A3) f is injective for $\phi \in \Phi$.

(A4) The derivative of f is always nonzero on Φ, i.e., $|\frac{df}{d\phi}| \neq 0$, $\forall \phi \in \Phi$.

Assumption A1 is the standard assumption not only in quantum metrology, but also in statistical parameter estimation. Assumption A2 is necessary for the use of Hoeffding's inequality (Lemma 2) and empirical Bernstein inequality (Lemma 3) in the proof of Theorem 1. Unbounded outcomes can exist theoretically, but outcomes are always bounded in experiments since there is a technical limit, or cutoff, on an observable range of measurement outcomes. So, assumption A2 is natural in experiments. Assumption A3 is necessary for the uniqueness of the least squares estimates for any data, and assumption A4 is necessary for avoiding the divergence of δ_1 and δ_2. In Appendix A 2 we explain that assumptions A3 and A4 are required in the use of the linearized uncertainty, which means that A3 and A4 are implicitly assumed in the standard approach using the LU.

2. Linearized uncertainty and assumptions

We explain a relation between the RMSE and LU [Eq. (A8)], which clarifies the role of assumptions A3 and A4 for the LU. The RMSE of an estimator ϕ^{est} is defined by

$$(\delta\phi)_{\mathrm{RMSE}}(\phi^{\mathrm{est}},n|\phi) := \sqrt{\mathbb{E}\big[\big(\phi_n^{\mathrm{est}}(\boldsymbol{x}^n) - \phi\big)^2\big]}. \quad (A1)$$

Let us choose the DI estimator ϕ^{DI} as the estimator. The DI estimates do not necessarily exist for any data. Assumption A3 guarantees the existence of the DI estimates only for $S_n \in \mathcal{R}_f$. When S_n is out of $\mathcal{R}_f$, the DI estimate may not exist. A3 is a necessary condition for the existence, but it is not a sufficient condition. However, let us ignore this fact, i.e., we assume that DI estimates exist for any data. By definition,

$$f(\phi_n^{\mathrm{DI}}) = S_n \quad (A2)$$

holds. We have

$$\mathbb{V}[\mathbf{\Pi}|\rho_\phi] = n\mathbb{E}[(S_n - \mathbb{E}[\mathbf{\Pi}|\rho_\phi])^2] \tag{A3}$$

$$= n\mathbb{E}\big[\{f(\phi_n^{\mathrm{DI}}) - f(\phi)\}^2\big]. \tag{A4}$$

We apply the Taylor expansion to f,

$$f(\phi_n^{\mathrm{DI}}) = f(\phi) + \frac{df}{d\phi}(\phi)(\phi_n^{\mathrm{DI}} - \phi) + O(|\phi_n^{\mathrm{DI}} - \phi|^2), \tag{A5}$$

and suppose that n is sufficiently large that the nonlinear terms in the Taylor expansion $O(|\phi_n^{\mathrm{DI}} - \phi|^2)$ is negligible. Then

$$\mathbb{V}[\mathbf{\Pi}|\rho_\phi] = n\left(\frac{df}{d\phi}\right)^2 \mathbb{E}\big[(\phi_n^{\mathrm{DI}} - \phi)^2\big] \tag{A6}$$

$$\approx n\left(\frac{df}{d\phi}\right)^2 (\delta\phi)_{\mathrm{RMSE}}(\phi^{\mathrm{DI}}, n|\phi)^2 \tag{A7}$$

holds. Since $\frac{df}{d\phi} \neq 0$ holds from assumption A4, we obtain

$$(\delta\phi)_{\mathrm{RMSE}}(\phi^{\mathrm{DI}}, n|\phi) \approx \frac{1}{\sqrt{n}} \frac{\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}}{|\frac{d}{d\phi}\mathbb{E}[\mathbf{\Pi}|\rho_\phi]|} \tag{A8}$$

$$= (\delta\phi)_{\mathrm{LU}}. \tag{A9}$$

Equation (A8) means that the linearized uncertainty is an approximated RMSE of the DI estimator, which is derived by ignoring the existence problem of the estimator and the nonlinearity of f. This is the reason why we call $(\delta\phi)_{\mathrm{LU}}$ a *linearized* uncertainty.

In the derivation of Eq. (A8), the following two conditions are required in addition to assumptions A3 and A4.

(C1) DI estimates exist for $S_n \notin \mathcal{R}_f$.

(C2) The number of measurement trials n is sufficiently large that the nonlinearity of f around ϕ is negligible.

In the standard approach using the LU, assumptions A1, A3, A4, and conditions C1 and C2 are implicitly assumed. On the other hand, Theorem 1 does not require C1 and C2. Especially the disuse of C2 is important to analyze finite data, because it is unclear which n can be considered as "sufficiently" large in C2.

APPENDIX B: PROOF OF THEOREM 1

In this Appendix we show the proof of Theorem 1. We derive an upper bound of $\mathbb{P}[|\phi_n^{\mathrm{LS}} - \phi| > \delta] = 1 - \mathbb{P}[|\phi_n^{\mathrm{LS}} - \phi| \leqslant \delta]$. It is difficult to directly analyze this quantity, because ϕ_n^{LS} is a nonlinear function of S_n and is a biased estimator. On the other hand, the following two lemmas hold for S_n.

Lemma 2 (Hoeffding's inequality [19]). Let X be a random variable with $X \in [a,b]$ and $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables satisfying $X_i = X(i = 1, \ldots, n)$, respectively. Then for any $0 < \epsilon < 1$ and $n \geqslant 1$,

$$\mathbb{P}\left[|S_n - \mathbb{E}[X]| > \sqrt{\frac{2}{n}V_{\max}\ln\frac{2}{\epsilon}}\right] \leqslant \epsilon \tag{B1}$$

holds.

Lemma 3 (Empirical Bernstein inequality [20]). Let X be a random variable with $a \leqslant X \leqslant b$ and $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables satisfying $X_i = X(i =$

1, ...,n), respectively. Then for any $0 < \epsilon < 1$ and $n \geqslant 2$,

$$\mathbb{P}\left[|S_n - \mathbb{E}[X]| > \sqrt{\frac{2}{n}V_n(X^n)\ln\frac{4}{\epsilon}} + \frac{8(b-a)}{3(n-1)}\ln\frac{4}{\epsilon}\right] \leqslant \epsilon \tag{B2}$$

holds [33], where

$$V_n(X^n) := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2. \tag{B3}$$

Note that $V_{\max}$ in Hoeffding's inequality is independent of data, and that $V_n(X^n)$ in the empirical Bernstein inequality is dependent of data.

First, we reduce the analysis of $|\phi_n^{\mathrm{LS}}(x^n) - \phi|$ to that of $|S_n - \mathbb{E}[\mathbf{\Pi}|\rho_\phi]|$. Let r denote the argument of g. Using the Taylor expansion of $g(r)$ around S_n^{LS} up to second order with the remainder in the Lagrange form, we obtain the following inequality:

$$|\phi_n^{\mathrm{LS}}(x^n) - \phi| = |g(S_n^{\mathrm{LS}}) - g(r)| \tag{B4}$$

$$= \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})(r - S_n^{\mathrm{LS}}) + \frac{1}{2}\frac{d^2g}{dr^2}(r')(r - S_n^{\mathrm{LS}})^2\right| \tag{B5}$$

$$\leqslant \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right||r - S_n^{\mathrm{LS}}| + \frac{1}{2}\left|\frac{d^2g}{dr^2}(r')\right||r - S_n^{\mathrm{LS}}|^2, \tag{B6}$$

where r' is some real number between r and S_n^{LS}. By combining Eq. (B6) with $|\frac{d^2g}{dr^2}(r')| = |\{\frac{df}{d\phi}(\phi')\}^{-3}\frac{d^2f}{d\phi^2}(\phi')| \leqslant L$ and the contractivity, we obtain

$$|\phi_n^{\mathrm{LS}}(x^n) - \phi| \leqslant \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right||S_n - r| + \frac{1}{2}L|S_n - r|^2. \tag{B7}$$

Then $\delta < |\phi_n^{\mathrm{LS}} - \phi|$ implies

$$\delta < \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right||S_n - r| + \frac{1}{2}L|S_n - r|^2. \tag{B8}$$

By solving this quadratic inequality with $\delta > 0$, we can show that Eq. (B8) is equivalent to

$$|S_n - r| > \frac{1}{L}\left\{\sqrt{\left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|^2 + 2\delta L} - \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|\right\}. \tag{B9}$$

By substituting $\delta = \delta_1$ and $\delta = \delta_2$ into Eq. (B9), we obtain

$$\frac{1}{L}\left\{\sqrt{\left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|^2 + 2\delta_1 L} - \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|\right\} = \sqrt{\frac{2}{n}V_{\max}\ln\frac{2}{\epsilon}}, \tag{B10}$$

$$\frac{1}{L}\left\{\sqrt{\left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|^2 + 2\delta_2 L} - \left|\frac{dg}{dr}(S_n^{\mathrm{LS}})\right|\right\}$$
$$= \sqrt{\frac{2}{n}V_n(x^n)\ln\frac{4}{\epsilon}} + \frac{8(b-a)}{3(n-1)}\ln\frac{4}{\epsilon}, \tag{B11}$$

where we used the equalities $\frac{dg}{dr} = (\frac{df}{d\phi})^{-1}$ from assumption A3 and $\frac{df}{d\phi}(\phi_n^{\mathrm{LS}}) \neq 0$ from assumption A4 [note that $\phi_n^{\mathrm{LS}}(x^n) \in \Phi$

holds for any x^n]. From Lemmas 2 and 3 we obtain

$$
\mathbb{P}\big[|\phi_n^{\mathrm{LS}} - \phi| > \delta\big] \leqslant \mathbb{P}\Bigg[|S_n - r| > \min\Bigg\{\sqrt{\frac{2}{n}V_{\max}\ln\frac{2}{\epsilon}},
$$

$$
\times\sqrt{\frac{2}{n}V_n(x^n)\ln\frac{4}{\epsilon}} + \frac{8(b-a)}{3(n-1)}\ln\frac{4}{\epsilon}\Bigg\}\Bigg] \qquad \text{(B12)}
$$

$$
\leqslant \epsilon. \qquad \text{(B13)}
$$

∎

APPENDIX C: PROOF OF EQ. (20) IN THE PAPER

Here we show the proof of Eq. (20) in the paper:

$$
\lim_{n\to\infty}\{\sqrt{n}\mathbb{E}[\delta(x^n,\epsilon)]\}
$$

$$
\leqslant \lim_{n\to\infty}\{\sqrt{n}\mathbb{E}[\delta_2(x^n,\epsilon)]\} \qquad \text{(C1)}
$$

$$
= \lim_{n\to\infty}\Bigg\{\mathbb{E}\bigg[\bigg|\frac{dg}{dr}\big(S_n^{\mathrm{LS}}\big)\bigg|\sqrt{V_n}\bigg]\Bigg\}\sqrt{2\ln\frac{4}{\epsilon}} \qquad \text{(C2)}
$$

$$
\leqslant \lim_{n\to\infty}\Bigg\{\sqrt{\mathbb{E}\bigg[\bigg|\frac{dg}{dr}\big(S_n^{\mathrm{LS}}\big)\bigg|^2\bigg]}\sqrt{\mathbb{E}[V_n]}\Bigg\}\sqrt{2\ln\frac{4}{\epsilon}} \qquad \text{(C3)}
$$

$$
= \sqrt{\lim_{n\to\infty}\mathbb{E}\bigg[\bigg|\frac{dg}{dr}\big(S_n^{\mathrm{LS}}\big)\bigg|^2\bigg]}\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}\sqrt{2\ln\frac{4}{\epsilon}}, \qquad \text{(C4)}
$$

where we used the Cauchy-Schwarz inequality and the equality $\mathbb{E}[V_n] = \mathbb{V}[\mathbf{\Pi}|\rho_\phi]$.

From the Taylor expansion we have

$$
\frac{dg}{dr}\big(S_n^{\mathrm{LS}}\big) = \frac{dg}{dr}(r) + O\big(|S_n^{\mathrm{LS}} - r|\big). \qquad \text{(C5)}
$$

At the limit of n to infinity, $|S_n^{\mathrm{LS}} - r|$ converges to 0 because of the contractivity $|S_n^{\mathrm{LS}} - r| \leqslant |S_n - r|$, and the law of large numbers. Then

$$
\lim_{n\to\infty}\mathbb{E}\bigg[\bigg|\frac{dg}{dr}\big(S_n^{\mathrm{LS}}\big)\bigg|^2\bigg] = \bigg|\frac{dg}{dr}(r)\bigg|^2 \qquad \text{(C6)}
$$

holds, and we obtain

$$
\lim_{n\to\infty}\{\sqrt{n}\mathbb{E}[\delta(x^n,\epsilon)]\} \leqslant \bigg|\frac{dg}{dr}(r)\bigg|\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}\sqrt{2\ln\frac{4}{\epsilon}} \qquad \text{(C7)}
$$

$$
= \frac{\sqrt{\mathbb{V}[\mathbf{\Pi}|\rho_\phi]}}{\big|\frac{d\mathbb{E}[\mathbf{\Pi}|\rho_\phi]}{d\phi}\big|}\sqrt{2\ln\frac{4}{\epsilon}}. \qquad \text{(C8)}
$$

∎

APPENDIX D: DETAILS OF RAMSEY INTERFEROMETER SIMULATION

In this Appendix we explain the details of a Ramsey interferometer and the Monte Carlo simulation. When we use a separable state of N atoms for the initial state, the LU scales as the SQL scaling $O(1/\sqrt{N})$. On the other hand, when we use an entangled state, the LU can scale as the HL scaling $O(1/N)$.

The procedure of the Ramsey interferometer is as follows.

1. Prepare an initial state $|\phi\rangle$ of N atoms. Each atom is a two-level system.

2. Each atom independently undergoes a free evolution, $\exp(i\frac{\phi}{2}\sigma_3)$.

3. After the evolution, we perform a $\frac{\pi}{2}$ pulse along an axis, $\cos\phi_0\,\sigma_1 + \sin\phi_0\,\sigma_2$, where ϕ_0 is a reference phase to be user tuned.

4. Perform a projective measurement of an observable A.

5. Repeat 1 to 4 a number n of times.

We consider the following two combinations of the initial state $|\psi\rangle$ and measured observable A.

(1) A product state and energy measurement. Let us choose a product state

$$
|\phi\rangle = \bigg[\frac{1}{\sqrt{2}}(|e\rangle + |g\rangle)\bigg]^{\otimes N} \qquad \text{(D1)}
$$

as the initial state, where $|e\rangle$ and $|g\rangle$ are the excited and ground states, respectively. We observe the total energy

$$
J_3 := \sum_{j=1}^{N}\sigma_3^{(j)}, \qquad \text{(D2)}
$$

where $\sigma_3^{(j)} := I^{\otimes(j-1)} \otimes \sigma_3 \otimes I^{\otimes(N-j)}$. The set of possible measurement outcomes is $\mathcal{X} = \{-N, -(N-1), \ldots, N-1, N\}$. In this combination, the probability distribution is given by

$$
p(x|\rho_\phi,\mathbf{\Pi}) = \frac{N!}{[(N+x)/2]![(N-x)/2]!}
$$

$$
\times \bigg\{\frac{1 + \sin(\phi - \phi_0)}{2}\bigg\}^{\frac{N+x}{2}}
$$

$$
\times \bigg\{\frac{1 - \sin(\phi - \phi_0)}{2}\bigg\}^{\frac{N-x}{2}}, \qquad \text{(D3)}
$$

and we obtain the following equalities:

$$
f(\phi) = N\sin(\phi - \phi_0), \qquad \text{(D4)}
$$

$$
\frac{df}{d\phi} = N\cos(\phi - \phi_0), \qquad \text{(D5)}
$$

$$
\mathbb{V}[\mathbf{\Pi}|\rho_\phi] = N\cos^2(\phi - \phi_0), \qquad \text{(D6)}
$$

$$
V_{\max} = N^2, \qquad \text{(D7)}
$$

$$
B_{\mathrm{LU}} = \frac{1}{\sqrt{N}}, \qquad \text{(D8)}
$$

$$
\bigg\{\frac{df}{d\phi}\bigg\}^{-3}\frac{d^2 f}{d\phi^2} = -\frac{1}{N^2}\frac{\sin(\phi - \phi_0)}{\cos^3(\phi - \phi_0)}. \qquad \text{(D9)}
$$

From Eq. (D4), f is a periodic function with period 2π. In order to satisfy assumptions A3 and A4, the size of Φ must be at most smaller than π.

(2) A GHZ state and parity measurement. Let us choose a GHZ state,

$$
|\psi\rangle = \frac{1}{\sqrt{2}}(|e\rangle^{\otimes N} + |g\rangle^{\otimes N}) \qquad \text{(D10)}
$$

as the initial state. We observe the parity,

$$P = (+1)^{N_g}(-1)^{N_e}, \qquad (D11)$$

where N_g and N_e are the particle number operators for $|g\rangle$ and $|e\rangle$, respectively. The set of possible measurement outcomes is $\mathcal{X} = \{-1, +1\}$. In this combination, the probability distribution is given by

$$p(x|\rho_\phi, \mathbf{\Pi}) = \frac{1}{2}\left\{ 1 + x \cos N\left(\phi - \phi_0 + \frac{\pi}{2}\right)\right\}. \quad (D12)$$

In this case we have the following equalities:

$$f(\phi) = \cos N\left(\phi - \phi_0 + \frac{\pi}{2}\right), \qquad (D13)$$

$$\frac{df}{d\phi} = -N \sin N\left(\phi - \phi_0 + \frac{\pi}{2}\right), \qquad (D14)$$

$$\mathbb{V}[\mathbf{\Pi}|\rho_\phi] = \sin^2 N\left(\phi - \phi_0 + \frac{\pi}{2}\right), \qquad (D15)$$

$$V_{\max} = 1, \qquad (D16)$$

$$B_{\mathrm{LU}} = \frac{1}{N}, \qquad (D17)$$

$$\left\{\frac{df}{d\phi}\right\}^{-3}\frac{d^2f}{d\phi^2} = \frac{1}{N}\frac{\cos N(\phi - \phi_0 + \frac{\pi}{2})}{\sin^3 N(\phi - \phi_0 + \frac{\pi}{2})}. \quad (D18)$$

From Eq. (D13), f is a periodic function with period $2\pi/N$. In order to satisfy assumptions A3 and A4, the size of Φ must be at most smaller than π/N.

Note that J_3 and P are commuting, and that these are reduced from the measurement of particle numbers for each energy level in an experiment [31]. Mathematically these are different observables, but the same measurement apparatus is used for both in the experiment. So, in cases (1) and (2), their initial states are different, their observables are different, and their POVMs are the same. In general, the LU is larger than or equivalent to the CRB, but in cases (1) and (2) their LUs coincide with their CRBs, respectively. This is the reason why we choose different observables.

We performed a Monte Carlo simulation for cases (1) and (2) with the following parameters: $\phi_{\min} = 0$, $\phi_{\max} = \pi/400$, $\phi = \pi/4000$, and $\epsilon = 0.1$ (90%-confidence level). The reference phases are chosen as $\phi_0 = -\pi/8$ for case (1) and $\phi_0 = \pi/2 - \pi/10N$. We show the result in Fig. 3, where B_{LU} and $\phi_{\max} - \phi_{\min}$ are added to Fig. 2. In Figs. 3(a) and 3(b) vertical axes are for expected deviations, $\mathbb{E}[\delta(x^n, \epsilon)]$ and $\mathbb{E}[|\phi_n^{\mathrm{LS}}(x^n) - \phi|]$. Solid and dashed (black) lines are $\mathbb{E}[\delta(x^n, \epsilon)]$ and $\mathbb{E}[|\phi_n^{\mathrm{LS}}(x^n) - \phi|]$ for the separable state, respectively. Chained and dotted (red) lines are $\mathbb{E}[\delta(x^n, \epsilon)]$ and $\mathbb{E}[|\phi_n^{\mathrm{LS}}(x^n) - \phi|]$ for the GHZ state, respectively. The expectations were calculated by a Monte Carlo sampling with 5000 repetitions. Each horizontal axis in Figs. 3(a) and 3(b) is for the number of atoms N and the number of measurement trials n, respectively. As explained in Sec. IV B, these panels indicate that δ for the GHZ state shows the Heisenberg limit scaling $O(1/N)$ for finite n. The scalings of $\mathbb{E}[\delta(x^n, \epsilon)]$ with respect to N and n are independent of ϵ, and the quantum enhancement on δ appears not only for $\epsilon = 0.1$ but also for other values of ϵ,

Solid blue and orange lines in Fig. 3 are $\frac{B_{\mathrm{LU}}}{\sqrt{n}}\sqrt{2\ln\frac{4}{\epsilon}}$ for cases (1) and (2), respectively. Equation (20) guarantees that $\mathbb{E}[\delta(x^n, \epsilon)]$ becomes smaller than the new lines in the limit of n going to infinity. Figure 3(b) indicates that two lines for $\mathbb{E}[\delta(x^n, \epsilon)]$ (solid black and chained red lines) become closer to the blue and orange lines as n becomes larger, respectively. However, lines for $\mathbb{E}[\delta(x^n, \epsilon)]$ are still larger than $\frac{B_{\mathrm{LU}}}{\sqrt{n}}\sqrt{2\ln\frac{4}{\epsilon}}$ at $n = 10\,000$. This means that $n = 10\,000$ cannot be considered as a "sufficiently" large number.
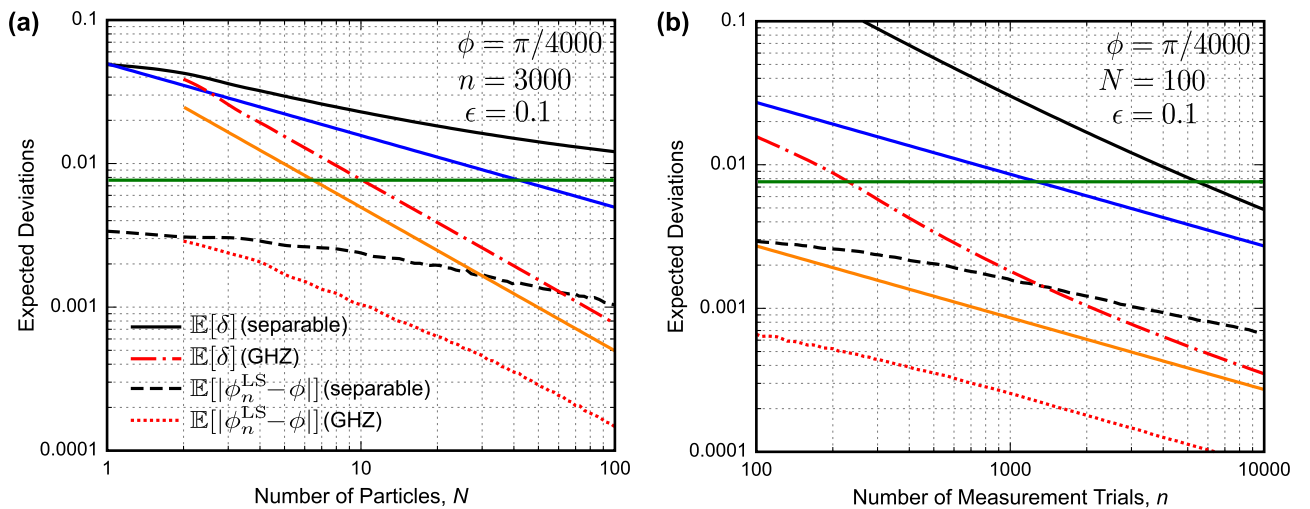

FIG. 3. (Color online) Numerical result on expected deviations ($\mathbb{E}[\delta]$ with $\epsilon = 0.1$, and $\mathbb{E}[|\phi_n^{\mathrm{LS}} - \phi|]$) in a Ramsey interferometer using a separable or GHZ initial state of N atoms. Expectations were calculated by a Monte Carlo sampling with 5000 repetitions. Solid blue and orange lines are $\frac{B_{\mathrm{LU}}}{\sqrt{n}}\sqrt{2\ln\frac{4}{\epsilon}}$ for the separable and GHZ states, respectively. Solid green lines are for $\phi_{\max} - \phi_{\min}$, which is a trivial 100%-confidence interval. (a) N dependency with $n = 3000$, and (b) n dependency with $N = 100$. Both panels indicate that δ for the GHZ state shows the HL scaling $O(1/N)$ for finite n.
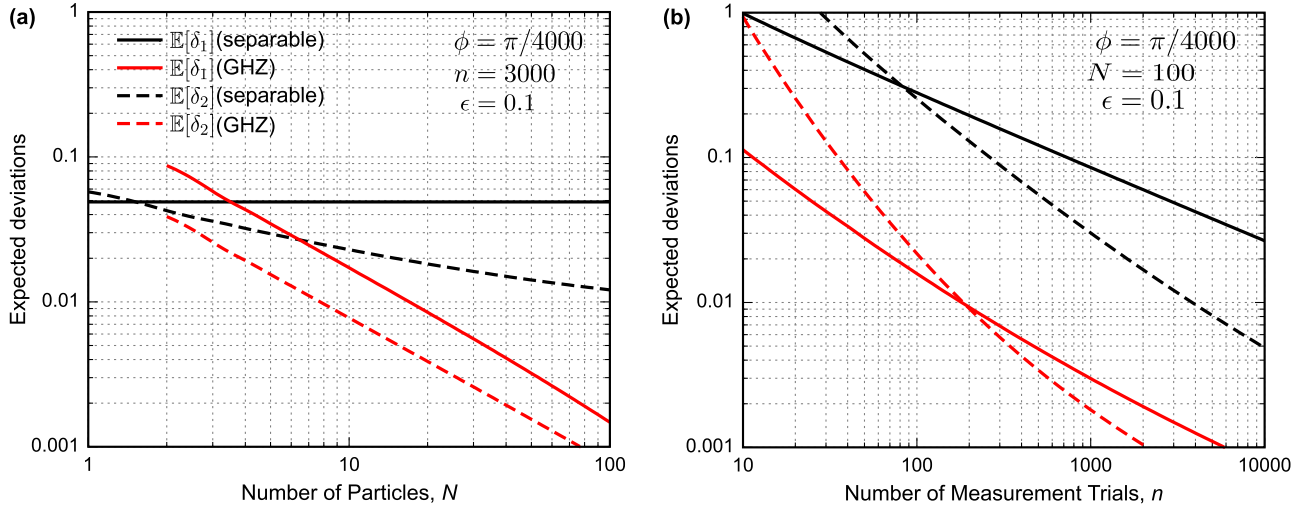
FIG. 4. (Color online) Numerical result on expected deviations ($\mathbb{E}[\delta_1]$ and $\mathbb{E}[\delta_2]$ with $\epsilon = 0.1$) in a Ramsey interferometer using a separable or GHZ initial state of N atoms. Expectations were calculated by a Monte Carlo sampling with 5000 repetitions. (a) N dependency with $n = 3000$, and (b) n dependency with $N = 100$.

Solid green lines in Fig. 3 are for $\phi_{\max} - \phi_{\min} = \pi/400 \approx 0.008$. Because $\phi_n^{\mathrm{LS}}(x^n)$ is always included in Φ for any data,

$$\left| \phi_n^{\mathrm{LS}} - \phi \right| \leqslant \phi_{\max} - \phi_{\min} \qquad (\mathrm{D19})$$

holds with probability 1. So, $\phi_{\max} - \phi_{\min}$ gives a trivial 100%-confidence interval. In Fig. 3(a), $\mathbb{E}[\delta]$ for the separable state (solid black line) is larger than $\phi_{\max} - \phi_{\min}$ (the solid green line) for all N between 1 and 100. This means that, on average, $n = 3000$ is not enough for obtaining a nontrivial 90%-confidence interval in case (1), while the number is enough in case (2) with $N \geqslant 10$. In Fig. 3(b), $\mathbb{E}[\delta]$ for the separable state (solid black line) becomes smaller than $\phi_{\max} - \phi_{\min}$ (solid green line) for $n \geqslant 6000$. If we perform measurement trials more than 6000 times for the separable state, it is expected to obtain a nontrivial 90%-confidence interval

Finally we analyze the behaviors of δ_1 and δ_2. In Figs. 4(a) and 4(b) vertical axes are for expected deviations $\mathbb{E}[\delta_1(x^n, \epsilon)]$ and $\mathbb{E}[\delta_2(x^n, \epsilon)]$. Solid and dashed (black) lines are $\mathbb{E}[\delta_1(x^n, \epsilon)]$ and $\mathbb{E}[\delta_2(x^n, \epsilon)]$ for the separable state, respectively. Solid and dashed (red) lines are $\mathbb{E}[\delta_1(x^n, \epsilon)]$ and $\mathbb{E}[\delta_2(x^n, \epsilon)]$ for the GHZ state, respectively.

In Fig. 4(a) the horizontal axis is the number of atoms N. Plots in the panel express the scaling of the expected δ_1 and δ_2 with respect to N with a fixed number of measurement trials, $n = 3000$. $\mathbb{E}[\delta_1]$ for the separable state (solid black)

is almost constant, although the other three plots decrease as N becomes large. This is caused by the difference between scalings of $V_{\max}$ and V_n. Roughly speaking, V_n and $\mathbb{V}[\mathbf{\Pi} | \rho_\phi]$ have the same scaling with respect to N. From Eqs. (D5), (D6), (D7), (D9), (D14), (D15), (D16), and (D18), we have the following scalings of δ_1 and δ_2 with respect to N for fixed ns:

$$\delta_1 = \begin{cases} O(1) & \text{(the separable state)} \\ O(1/N) & \text{(the GHZ state)} \end{cases}, \qquad (\mathrm{D20})$$

$$\delta_2 = \begin{cases} O(1) & \text{(the separable state, small } n) \\ O(1/\sqrt{N}) & \text{(the separable state, sufficiently large } n). \\ O(1/N) & \text{(the GHZ state)} \end{cases}$$

$$(\mathrm{D21})$$

The scaling of the dashed black line is between $O(1/\sqrt{N})$ and $O(1)$, which means that $n = 3000$ is not small and is not sufficiently large.

In Fig. 4(b) the horizontal axis is the number of measurement trials n. Plots in the panel express the scaling of the expected δ_1 and δ_2 with respect to n with a fixed number of atoms, $N = 100$. Plots for the separable and GHZ states have a same behavior, i.e., $\mathbb{E}[\delta_1] < \mathbb{E}[\delta_2]$ for small n and $\mathbb{E}[\delta_1] > \mathbb{E}[\delta_2]$ for large n. For small n, a correction term $\frac{8(b-a)}{3(n-1)} \ln \frac{4}{\epsilon}$ in δ_2 is not negligible, and $\delta_1 < \delta_2$ holds. For large n, the correction term becomes negligible, and $\delta_1 > \delta_2$ becomes true because $V_n \leqslant V_{\max}$ holds.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[2] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).

[3] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, arXiv:1405.7703 [quant-ph].

[4] J. G. Rarity, P. R. Tapster, E. Jakeman, T. Larchuk, R. A. Campos, M. C. Teich, and B. E. A. Saleh, Phys. Rev. Lett. **65**, 1348 (1990).

[5] V. Meyer, M. A. Rowe, D. Kielpinski, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Phys. Rev. Lett. **86**, 5870 (2001).

[6] A. Widera, O. Mandel, M. Greiner, S. Kreim, T. W. Hänsch, and I. Bloch, Phys. Rev. Lett. **92**, 160406 (2004).

[7] There are two types of the CRBs. One is for finite data, and the other is for infinite data. The sufficient and necessary condition for the attainability of the CRBs for finite data is that the

probability distribution of measurement outcomes is an affine function of the parameter ϕ [18]. In quantum metrology the probability distribution is not affine of ϕ, and the CRBs are not attainable for finite data. It becomes attainable at infinite data.

[8] Note that n and N are different. N is the number of particles in a probe system used for each measurement trial, and n is the number of measurement trials.

[9] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley Series in Probability and Statistics (Wiley, New York, 2002) (originally published in 1973).

[10] R. R. Bahadur, Sankhyā **22**, 229 (1960).

[11] J. M. Taylor, P. Cappellaro, L. Childress, L. Jiang, D. Budker, P. R. Hemmer, A. Yacoby, R. Walsworth, and M. D. Lukin, Nat. Phys. **4**, 810 (2008).

[12] M. Napolitano, M. Koschorreck, B. Dubost, N. Behbood, R. J. Sewell, and M. W. Mitchell, Nature (London) **471**, 486 (2010).

[13] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde, Nat. Photon. **5**, 43 (2011).

[14] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

[15] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, New York, 1982).

[16] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. (Springer, New York, 2005).

[17] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. Lett. **111**, 160406 (2013).

[18] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs (American Mathematical Society, Providence, 2000).

[19] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).

[20] A. Maurer and M. Pontil, in COLT 09: Proceedings of the 22nd Annual Conference on Learning Theory (2009), http://colt2009.cs.mcgill.ca/papers/012.pdf.

[21] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett. **79**, 3865 (1997).

[22] Y. Matsuzaki, S. C. Benjamin, and J. Fitzsimons, Phys. Rev. A **84**, 012103 (2011).

[23] A. W. Chin, S. F. Huelga, and M. B. Plenio, Phys. Rev. Lett. **109**, 233601 (2012).

[24] R. Chaves, J. B. Brask, M. Markiewicz, J. Kołodyński, and A. Acin, Phys. Rev. Lett. **111**, 120401 (2013).

[25] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus, Phys. Rev. Lett. **112**, 080801 (2014).

[26] H. Lee, P. Kok, and J. P. Dowling, J. Mod. Opt. **49**, 2325 (2002).

[27] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, Nat. Phys. **7**, 406 (2011).

[28] R. Demkowicz-Dobrzánski, J. Kołodyński, and M. Gută, Nat. Commun. **3**, 1063 (2012).

[29] G. Arrad, Y. Vinkler, D. Aharonov, and A. Retzker, Phys. Rev. Lett. **112**, 150801 (2014).

[30] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. **112**, 150802 (2014).

[31] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, Science **304**, 1476 (2004).

[32] C. D. Aiello, M. Hirose, and P. Cappellaro, Nat. Commun. **4**, 1419 (2012).

[33] Some coefficients in Eq. (B2) are different from the corresponding inequality in [20], because we have a proof of Eq. (B2) but we could not prove the original inequality.