

Quantum enhanced metrology of Hamiltonian parameters beyond the Cramèr–Rao bound

Luigi Seveso and Matteo G. A. Paris*

*Dipartimento di Fisica ‘Aldo Pontremoli’
Università di Milano, I-20133 Milano, Italy
matteo.paris@fisica.unimi.it

Received 18 February 2020

Accepted 10 March 2020

Published 20 May 2020

This is a tutorial aimed at illustrating some recent developments in quantum parameter estimation beyond the Cramèr–Rao bound, as well as their applications in quantum metrology. Our starting point is the observation that there are situations in classical and quantum metrology where the unknown parameter of interest, besides determining the state of the probe, is also influencing the operation of the measuring devices, e.g. the range of possible outcomes. In those cases, nonregular statistical models may appear, for which the Cramèr–Rao theorem does not hold. In turn, the achievable precision may exceed the Cramèr–Rao bound, opening new avenues for enhanced metrology. We focus on quantum estimation of Hamiltonian parameters and show that an achievable bound to precision (beyond the Cramèr–Rao) may be obtained in a closed form for the class of so-called controlled energy measurements. Examples of applications of the new bound to various estimation problems in quantum metrology are worked out in some details.

Keywords: Quantum estimation theory; Fisher information; Cramer-Rao bound.

1. Introduction

In the last decade, quantum signals and detectors carved out a place for themselves in mainstream technology. Characterization of those devices at the quantum level is thus a crucial ingredient for the development of quantum technologies. Quantum metrology, on the other hand, is the art of estimating the value of one or more parameters of interest, e.g. those characterizing the operation of a device, by exploiting the quantum features of *both* the probing system and the measuring apparatus. This second, broadly employed, understanding of the concept has attracted the interest of many researchers, causing a rapid development of the field.^{1,2}

* Corresponding author.

Quantum estimation theory (QET) is the mathematical framework where to address optimization of a quantum measurement.²⁻⁴ It applies to situations where one is interested in inferring the value of a parameter by performing a set of measurements on identical repeated preparations of the system, and then processing data in order to estimate the value of the unknown parameter. In turn, the goal of QET is to optimize the overall inference strategy, i.e. the two following steps: (1) the choice of the most convenient measurement apparatus and (2) the choice of the most convenient estimator, i.e. the data processing able to extract as much information as possible about the parameter of interest. The figure of merit used to assess the precision the estimation is the mean square error and an inference strategy is deemed optimal if the mean square error achieves a minimum. Step (2) is classical in nature, and amounts to choose a suitable data processing. On the other hand, the first one is where the quantum nature of physical devices come into play.

Usually, the choice of the optimal measurement is made by optimizing the figure of merit assuming that the information on the unknown parameter comes from the statistical manifold of possible quantum states of the system only. In other words, one assumes that the measurement apparatus aimed at estimating the parameter does not depend on its value. Such an assumption is necessary to employ standard tools of QET, i.e. the concept of quantum Fisher information and the so-called quantum Cramèr–Rao theorem.

As a matter of fact, there are relevant estimation problems where the above assumption does not hold. In those cases, an alternative approach is needed to obtain the ultimate precision bounds, as imposed by quantum mechanics. Relevant examples are provided by statistical models for Hamiltonian parameters, and by models where the sample space of possible results do depend itself on the parameter of interest. In order to address those scenarios, novel bounds have been proposed, some of them being tight and achievable. In particular, it has been proved that the achievable precision may exceed the Cramèr-Rao bound, thus opening new avenues for quantum enhanced metrology.^{5,6}

In this tutorial, we review some recent developments in quantum parameter estimation beyond the Cramèr-Rao bound, as well as their applications in quantum metrology. We focus on quantum estimation of Hamiltonian parameters, illustrate the novel bound (beyond the Cramèr–Rao one) for the so-called class of controlled energy measurements, and work out in details few examples of applications, especially those of interest for quantum magnetometry. In order to place the reader in a position to appreciate the recent developments, we will introduce in detail the basic notions of quantum parameter estimation, paying the necessary attention to the mathematical framework where those notions had been developed. In turn, the paper is structured as follows: In Sec. 2, we provide a brief summary of concepts and notations used in probability theory, whereas Sec. 3 is devoted to classical parameter estimation and Sec. 4 to quantum measurement theory. Quantum parameter estimation is briefly reviewed in Sec. 5, whereas nonregular measurements and parameter

estimation beyond the quantum Cram r–Rao theorem are discussed in Sec. 6. Non-regular estimation of general Hamiltonian parameters is the subject of Sec. 7. In particular, we analyze metrological scheme based on controlled energy measurements and present a tight achievable bound for the precision they may achieve. In Secs. 8 and 9, we discuss metrological applications of the above findings, and work out in details few examples of interest in quantum magnetometry. Section 10 closes the paper with some concluding remarks.

2. Elements of Probability Theory

The outcome of a random experiment is an *event*. At this stage, an event has no numerical counterpart: it is an abstract subset of a *sample space* Ω . In general, not every possible subset of Ω constitutes an event. A few desirable requirements are the following: an experiment may have no outcome, so the empty set should be an event; if A is a possible event, then its complement A^c , or logical negation, should also be an event; if A and B are events, then their union $A \cup B$, or logical conjunction, should also be an event. Such requirements naturally lead to the introduction of an σ -algebra structure on the set of events.

Definition 1 (σ -algebra). A σ -algebra \mathcal{A} on a sample space Ω is a family of subsets of Ω having the following properties: **(P1)**: The empty set \emptyset is an element of \mathcal{A} ; **(P2)**: If A is an element of \mathcal{A} , then also its relative complement $A^c \in \mathcal{A}$; **(P3)**: If $\{A_i\}_{i=1}^\infty$ is a countable collection of elements of \mathcal{A} , then also $\cup_{i=1}^\infty A_i \in \mathcal{A}$.

The tuple (Ω, \mathcal{A}) is called a *measurable space* and the elements of \mathcal{A} the *measurable sets*. Making use of properties **(P2)**–**(P3)**, one may prove that, if $\{A_i\}_{i=1}^\infty$ is a countable collection of elements in \mathcal{A} , then also their countable intersection $\cap_{i=1}^\infty A_i \in \mathcal{A}$. It follows that if \mathcal{A}_1 and \mathcal{A}_2 are two different σ -algebras on the same sample space Ω , then their intersection $\mathcal{A}_1 \cap \mathcal{A}_2$ is also a σ -algebra. From this, one may go on to prove that, given any family of sets τ , there is a unique smallest σ -algebra containing τ . A case of major interest is when τ is a topology on Ω , i.e. the tuple (Ω, τ) is a *topological space*.

Definition 2 (Topological space). A topological space (Ω, τ) is a set Ω provided with a topology τ , i.e. a family of subsets of Ω having the following properties: **(P1)**: Both Ω and the empty set \emptyset are elements of τ ; **(P2)**: If $\{T_i\}_{i=1}^\infty$ is a countable collection of elements of τ , then also $\cup_{i=1}^\infty T_i \in \tau$; **(P3)**: If $\{T_i\}_{i=1}^n$ is a finite collection of elements of τ , then also $\cap_{i=1}^n T_i \in \tau$.

The elements of a topology τ on Ω are called the *open sets* of Ω . If Ω is endowed with a topological structure to start with, a σ -algebra structure can be introduced by taking countable unions, countable intersections and relative complements of its open sets. The resulting σ -algebra is called the *Borel σ -algebra* $\mathcal{B}(\Omega)$: it is the smallest σ -algebra containing the open sets of Ω . Once a σ -algebra structure \mathcal{A} has

been introduced on Ω , the probability of different events is specified by a *probability measure* μ . The triple $(\Omega, \mathcal{A}, \mu)$ is called a *probability space*.

Definition 3 (Probability space). A probability space $(\Omega, \mathcal{A}, \mu)$ is a set Ω together with a σ -algebra structure \mathcal{A} and a probability measure μ , i.e. a function $\mu : \mathcal{A} \rightarrow [0, 1]$ having the following properties: **(P1)**: $\mu(\Omega) = 1$; **(P2)**: If $\{A_i\}_{i=1}^\infty$ is a countable collection of mutually disjoint elements of \mathcal{A} , then

$$\mu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i). \tag{1}$$

With the help of property **(P2)**, one may also prove that a probability measure satisfies the following intuitive properties: $\mu(\emptyset) = 0$; if $B \subset A$, then $\mu(B) \leq \mu(A)$; for any two events $A, B \in \mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$. Notice that if (Ω, \mathcal{A}) is a measurable space and $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ is a function from the measurable sets to the extended (nonnegative) real line, satisfying property **(P2)**, the triple $(\Omega, \mathcal{A}, \mu)$ is called a *measure space* and μ a *measure*. A measure μ is said to be *finite* if $\mu(\Omega)$ is a finite real number (it is said *σ -finite* if Ω is countable union of measurable sets having finite measure). A probability space is thus equivalent to a measure space with finite measure, normalized according to property **(P1)**. While the random outcomes of an experiment are only required to have an σ -algebra structure, a *random variable* is needed in order to associate values to elements of Ω .

Definition 4 (Random variable). Given a probability space $(\Omega, \mathcal{A}, \mu)$ and a measurable space $(\mathcal{X}, \mathcal{B})$, a random variable is a function $\mathbb{X} : \Omega \rightarrow \mathcal{X}$ having the following property: if $B \in \mathcal{B}$, then the preimage of B under \mathbb{X} , i.e. $\mathbb{X}^{-1}(B) = \{\omega \in \Omega : \mathbb{X}(\omega) \in B\}$, is an element of \mathcal{A} .

Notice that the measurable space $(\mathcal{X}, \mathcal{B})$ in Definition 4 can be naturally made into a probability space, by introducing the probability measure ν defined via the relation $\nu(B) = \mu(\mathbb{X}^{-1}(B))$, where B is any measurable set in \mathcal{B} . In practice, one often blurs the distinction between the two probability spaces $(\Omega, \mathcal{A}, \mu)$ and $(\mathcal{X}, \mathcal{B}, \nu)$, and says that the outcome of a random experiment is a real value $x \in \mathcal{X}$, rather than an event $A \in \mathcal{A}$. We will also make use of such abuse of terminology when the distinction can be safely ignored. We add that since, by definition, a measurable function between two measurable spaces is a function such that the preimage of any measurable set is measurable, then a random variable can equivalently be defined as a measurable function between probability spaces. For most random variables of interest, the image set \mathcal{X} is a subset of the real line \mathbb{R} . If the subset is finite or countably infinite, the random variable is said to be *discrete*; otherwise, it is a *continuous*. In the following, by a random variable, it will always be meant a *real* random variable, either discrete or continuous. We will also assume that the σ -algebra \mathcal{B} is fixed by defining first a topological structure on \mathcal{X} (i.e. the subspace topology induced by the real line standard topology) and then a σ -algebra structure, i.e. the Borel algebra of \mathcal{X} .

Let us now sketch how to define a notion of integration of a random variable with respect to a probability measure. This is done initially only for *simple* random variables.

Definition 5 (Simple random variable). A random variable $\mathbb{X} : \Omega \rightarrow \mathcal{X}$ is simple if \mathcal{X} is a finite set.

As a consequence, a simple random variable \mathbb{X} can be written as $\mathbb{X} = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$, where $\{x_i\}_{i=1}^n$ are real numbers, $\{A_i\}_{i=1}^n$ are elements of \mathcal{A} and $\mathbf{1}_{A_i}$ is the characteristic function of A_i , i.e.

$$\mathbf{1}_{A_i}(\omega) := \begin{cases} 1 & \text{if } \omega \in A_i, \\ 0 & \text{if } \omega \notin A_i. \end{cases} \quad (2)$$

This representation is, in general, nonunique. If \mathbb{X} is a simple random variable, its *expectation* is defined as

$$E(\mathbb{X}) := \sum_{i=1}^n x_i \mu(A_i), \quad (3)$$

which can also be denoted by $\int_{\Omega} \mathbb{X} d\mu$. It can be proven that $E(\mathbb{X})$ does not depend on the representation. The next step is to define the expectation of nonnegative random variables. A random variable \mathbb{X} is nonnegative if it takes only nonnegative values. Two random variables satisfy $\mathbb{X} \geq \mathbb{Y}$ if their difference $\mathbb{X} - \mathbb{Y}$ is nonnegative. One defines

$$E(\mathbb{X}) := \sup(E(\mathbb{Y}), \mathbb{Y} \text{ a simple random variable with } 0 \leq \mathbb{Y} \leq \mathbb{X}). \quad (4)$$

Let us remark that, by definition, $E(\mathbb{X}) \geq 0$ and that $E(\mathbb{X})$ always exists, but might be equal to $+\infty$, even if \mathbb{X} is everywhere finite. The final step is to consider an arbitrary random variable \mathbb{X} . Let $\mathbb{X}^{(+)} = \max(\mathbb{X}, 0)$ and $\mathbb{X}^{(-)} = -\min(\mathbb{X}, 0)$. Thus, $\mathbb{X} = \mathbb{X}^{(+)} - \mathbb{X}^{(-)}$, where $\mathbb{X}^{(+)}$ and $\mathbb{X}^{(-)}$ are positive random variables. Then, one defines

$$E(\mathbb{X}) := E(\mathbb{X}^{(+)}) - E(\mathbb{X}^{(-)}). \quad (5)$$

A random variable \mathbb{X} is *integrable* if both $E(\mathbb{X}^{(+)})$ and $E(\mathbb{X}^{(-)})$ are finite; then, its expectation is given by Eq. (5). It is easy to check that the set of integrable random variables on a probability space $(\Omega, \mathcal{A}, \mu)$ is a vector space, denoted by \mathcal{L}^1 , with expectation acting as a linear map on it. Notice that, if two random variables satisfy $\mathbb{X} = \mathbb{Y}$ almost surely, i.e. $\mu(\{\omega \in \Omega : \mathbb{X}(\omega) = \mathbb{Y}(\omega)\}) = 1$, then $E(\mathbb{X}) = E(\mathbb{Y})$. Therefore, equality almost surely is an equivalence relation, denoted by \sim , and equivalent random variables have the same expectation. To remove this redundancy, one introduces the quotient space $L^1 := \mathcal{L}^1 / \sim$, whose elements are equivalence classes of almost surely equal random variables. However, by abuse of terminology, one usually still refers to elements of L^1 as random variables. In a similar way, for $1 \leq p < \infty$, one defines \mathcal{L}^p as the vector space of random variables such that

$|\mathbb{X}|^p \in \mathcal{L}^1$, where $|\mathbb{X}| := \mathbb{X}^{(+)} + \mathbb{X}^{(-)}$. By taking equivalence classes with respect to \sim , one then obtains the spaces L^p of p -integrable random variables. In the following, we will only need the spaces L^1 and L^2 .

If two random variables are square-integrable, they satisfy the following inequality.

Proposition 6 (Cauchy–Schwarz inequality). *If $\mathbb{X}, \mathbb{Y} \in L^2$, then $\mathbb{X} \cdot \mathbb{Y} \in L^1$ and*

$$|E(\mathbb{X} \cdot \mathbb{Y})| \leq \sqrt{E(\mathbb{X}^2)E(\mathbb{Y}^2)}. \tag{6}$$

Given square-integrable random variables $\{\mathbb{X}_i\}_{i=1}^n$ with $\mathbb{X}_i \in L^2$, one defines their covariance matrix as follows.

Definition 7 (Covariance matrix). Let $\{\mathbb{X}_i\}_{i=1}^n$ be a collection of square-integrable random variables in L^2 . Their covariance matrix is the matrix with entries.

$$\text{Cov}(\mathbb{X}_i, \mathbb{X}_j) := E[(\mathbb{X}_i - E(\mathbb{X}_i))(\mathbb{X}_j - E(\mathbb{X}_j))]. \tag{7}$$

In particular, the diagonal elements of a covariance matrix are the variances $\text{Var}(\mathbb{X}_i) := E[(\mathbb{X}_i - E(\mathbb{X}_i))^2]$. As a concluding remark, since the product of two measurable functions is a measurable function and the characteristic function $\mathbf{1}_A$ of a set A is measurable if and only if A is measurable, the integral of a random variable on any measurable set $A \in \mathcal{A}$ is well-defined: one has to take the expectation of the product $\mathbf{1}_A \cdot \mathbb{X}$, i.e. $\int_A \mathbb{X}d\mu = \int_{\Omega} \mathbf{1}_A \cdot \mathbb{X}d\mu$.

We now introduce the concept of *probability density* of a random variable. As discussed before, a random variable \mathbb{X} on a probability space $(\Omega, \mathcal{A}, \mu)$ gives rise to a probability space $(\mathcal{X}, \mathcal{B}, \nu)$, where $\mathcal{X} \subseteq \mathbb{R}$, \mathcal{B} is the Borel algebra generated by the natural topology of \mathcal{X} and ν is a probability measure. Notice that there are already two natural notions of a measure on \mathcal{X} : the *Lebesgue measure* (if \mathcal{X} is an uncountable subset of \mathbb{R}) and the *counting measure* (if \mathcal{X} is a countable subset). The measure ν can always be expressed in terms of either the Lebesgue measure or the counting measure, provided it satisfies a technical assumption, which is contained in the following definition.

Definition 8 (Absolutely continuous measures). If ν and ν' are any two measures with the same σ -algebra \mathcal{B} of subsets of \mathcal{X} , then ν is said to be absolutely continuous with respect to ν' , denoted $\nu \ll \nu'$, if $\nu(B) = 0$ for any $B \in \mathcal{B}$ such that $\nu'(B) = 0$.

We henceforth assume that, if \mathbb{X} is a continuous random variable, ν is absolutely continuous with respect to the Lebesgue measure, i.e. it agrees with the Lebesgue measure on any set with Lebesgue measure zero. If instead \mathbb{X} is discrete, every probability measure ν is already absolutely continuous with respect to the counting measure (since the counting measure vanishes only on the empty set and $\nu(\emptyset) = 0$ always). The following theorem applies to any two absolutely continuous measures.

Theorem 9 (Radon–Nikodym). Let ν and ν' be two σ -finite measures on the same measurable space $(\mathcal{X}, \mathcal{B})$ such that $\nu \ll \nu'$. Then: **(T1)**: There exists a measurable function $h : \mathcal{X} \rightarrow \mathbb{R}_+$ such that, for all $B \in \mathcal{B}$,

$$\nu(B) = \int_B h d\nu'. \tag{8}$$

(T2): Such a function h is almost unique: any two functions satisfying Eq. (8) can differ only on sets of measure zero with respect to ν' . **(T3)**: h is integrable with respect to ν' if and only if ν is a finite measure.

The function h is called the Radon–Nikodym derivative of ν with respect to ν' , denoted $h = d\nu/d\nu'$. It allows to convert between the two measures by means of the symbolic identity $d\nu = h d\nu'$. Given the above let us introduce the following definition.

Definition 10. Let \mathbb{X} be a random variable with probability space $(\mathcal{X}, \mathcal{B}, \nu)$. **(D1)**: If \mathbb{X} is continuous, its probability density function (p.d.f.) is the Radon–Nikodym derivative of ν with respect to the Lebesgue measure, i.e. $p = d\nu/dx$. **(D2)**: If \mathbb{X} is discrete, its probability mass function (p.m.f.) is the Radon–Nikodym derivative of ν with respect to the counting measure, i.e. $p = d\nu/d\#$.

Knowledge of the p.d.f. p (respectively, the p.m.f.) fully characterizes a random experiment whose outcomes are described by a random variable \mathbb{X} , since the probability of any event B can be obtained by integrating p on B with respect to the Lebesgue measure (respectively, the counting measure) by means of Eq. (8).

3. Classical Parameter Estimation

Let us consider a random experiment, whose outcomes are described by a random variable \mathbb{X} , with probability space $(\mathcal{X}, \mathcal{B}, \nu)$ and probability density p . The task is to reconstruct p , which is referred to as the *true* probability density, starting from N independent sample points or observations of \mathbb{X} (in the following, a sample point is denoted by a lowercase letter, e.g. $x \in \mathcal{X}$, whereas a sample of N observations by a boldface letter, e.g. $x \in \mathcal{X}^{\times N}$). There are many ways to approach the problem of learning p but, if the functional form of p is already known, or can be guessed with reasonable accuracy, a *parametric approach* is quite natural. The true probability density p is assumed to belong to a parametric family of probability densities $\{p_\theta\}_{\theta \in \Theta}$, where $\Theta \subset \mathbb{R}^m$ is the *parameter space*. It is also assumed that there exists a suitable choice $\theta^* \in \Theta$ such that $p_{\theta^*} = p$. In this way, all lack of knowledge about p is reduced to lack of knowledge about the *true* parameter θ^* — a considerable simplification of the problem.

Definition 11 (Classical statistical model). A classical statistical model S is a family of probability densities on \mathcal{X} parametrized by m real parameters $\theta \in \Theta \subset \mathbb{R}^m$.

$$S = \{p_\theta : \theta = (\theta^1, \theta^2, \dots, \theta^m) \in \Theta\}, \tag{9}$$

where the parametrization map $\theta \rightarrow p_\theta$ is injective, the support \mathcal{X} is parameter-independent and p_θ can be differentiated as many times as needed with respect to the parameters, i.e. all possible derivatives $\partial_1^{k_1} \cdots \partial_m^{k_m} p_\theta$ (where ∂_i is short for ∂_{θ^i}) exist.

Notice that if \mathcal{X} is countable, then p_θ is a p.m.f. normalized such that

$$\sum_{x \in \mathcal{X}} p_{x,\theta} = 1, \quad \forall \theta \in \Theta. \tag{10}$$

If \mathcal{X} is uncountable, then p_θ is a p.d.f. normalized such that

$$\int_{\mathcal{X}} p_\theta(x) dx = 1, \quad \forall \theta \in \Theta. \tag{11}$$

In the following, we will employ the notation for continuous variables; for discrete variables, one should replace the Lebesgue measure dx by the counting measure $d\#$.

Given a statistical model $S = \{p_\theta\}_{\theta \in \Theta}$, the map $\varphi : S \rightarrow \mathbb{R}^m$ defined by $\varphi(p_\theta) = \theta$ can be considered as providing a coordinate system for S . If ψ is a smooth reparametrization which maps $\Theta \rightarrow \Theta'$, nothing prevents using $\psi(\theta) = \theta'$ as the new parameters, so that the model is rewritten as $S = \{p_{\psi^{-1}(\theta')}\}_{\theta' \in \Theta'}$. This defines the structure of a differentiable manifold on S , with different parametrizations representing different coordinate systems. Moreover, a Riemannian metric can be defined on the statistical manifold S as follows.

Definition 12 (Fisher information). Let S be a statistical model. Given a point θ , the (classical) Fisher information matrix $\mathcal{F}_C(\theta)$ at that point is the matrix having (i, j) th element

$$[\mathcal{F}_C(\theta)]_{ij} = \int_{\mathcal{X}} dx p_\theta(x) \partial_i \log p_\theta(x) \partial_j \log p_\theta(x). \tag{12}$$

When $m = 1$ and only one parameter $\theta = \theta^1$ is present, $\mathcal{F}_C(\theta)$ is referred to as the Fisher information (FI). For $m > 1$, $\mathcal{F}_C(\theta)$ is indeed a symmetric $m \times m$ real matrix. It is always positive semi-definite and, in particular, positive-definite if and only if for every $\theta \in \Theta$ the elements of the set $\{\partial_1 p_\theta, \dots, \partial_m p_\theta\}$ are linearly independent. Moreover, $\mathcal{F}_C(\theta)$ has the correct transformation properties of a $(0, 2)$ tensor under reparametrizations.⁷ It follows that $\mathcal{F}_C(\theta)$ provides a Riemannian metric on S . There is a precise sense in which the Fisher geometry, i.e. the geometry implied by the Fisher information metric, is the only possible geometry on a statistical manifold. To explain this, we introduce the notion of a *statistic*.

Definition 13 (Statistic). Given a random variable \mathbb{X} and a function $T : \mathcal{X} \rightarrow \mathcal{Y}$ which maps $x \rightarrow y = T(x)$, a statistic based on T is the random variable $\mathbb{Y} = T(\mathbb{X})$.

If \mathbb{X} is associated with a statistical model $S = \{p_\theta\}_{\theta \in \Theta}$, then a statistic T gives rise to a model $S_T = \{q_\theta\}_{\theta \in \Theta}$ associated with $\mathbb{Y} = T(\mathbb{X})$. A statistic is said to be *sufficient* if the two models are related as follows: $p_\theta(x) = h(x)q_\theta(y(x))$, $\forall x \in \mathcal{X}$, i.e. all dependence on the parameter θ is contained in q_θ . Intuitively, a sufficient statistic leads to no loss of information about θ . Notice that a one-to-one function is always a

sufficient statistic, but there exist sufficient statistics which are not one-to-one functions. We now have the following theorem.

Theorem 14 (Pre Cramèr–Rao). *The Fisher information matrix $\mathcal{F}_C^{(T)}$ of the statistical model S_T induced by a statistic T satisfies the monotonicity property $\mathcal{F}_C^{(T)} \leq \mathcal{F}_C$ (where \mathcal{F}_C is the Fisher information matrix of the original model S). The previous inequality must be interpreted in the sense that the difference $\mathcal{F}_C^{(T)} - \mathcal{F}_C$ is a positive semi-definite matrix. Equality holds if and only if T is a sufficient statistic.*

A Riemannian metric satisfying the monotonicity property is said to be a monotone metric. Monotone metrics are the natural metrics on classical statistical models: they reflect the fact that the points of the manifold are probability distributions and distances between points can only contract under any information processing. In this regard, the following theorem^{8–10} singles out the Fisher information metric as the only natural metric on statistical manifolds.

Theorem 15 (Chentsov). *The Fisher information metric \mathcal{F}_C is the essentially unique monotone Riemannian metric on a classical statistical model, in the sense that any other such metric is a scalar multiple of \mathcal{F}_C .*

Chentsov’s theorem establishes a first link between the statistical properties of parametric models and the geometry defined by the Fisher metric. A further link comes from the (classical) Cramèr–Rao theorem, which we now introduce.

Let us now return to the problem of estimating the true parameter θ^* from a sample $\mathbf{x} \in \mathcal{X}^{\times N}$. To this end, we introduce the following definition.

Definition 16 (Estimator). An estimator $\hat{\theta}^{(N)} : \mathcal{X}^{\times N} \rightarrow \Theta$ is a random variable from the sample space $\mathcal{X}^{\times N}$ to the parameter space Θ . In particular, **(D1)**: An unbiased estimator is an estimator satisfying $E_\theta(\hat{\theta}^{(N)}) = \theta$, $\forall \theta \in \Theta$, where $E_\theta(\cdot)$ denotes expectation with respect to p_θ , i.e.

$$E_\theta(\hat{\theta}^{(N)}) = \int_{\mathcal{X}^{\times N}} dx_1, \dots, dx_N p_\theta(x_1), \dots, p_\theta(x_N) \hat{\theta}^{(N)}(x). \quad (13)$$

(D2): A locally unbiased estimator is an estimator which is unbiased at $\theta = \theta^*$, i.e.

$$E_{\theta^*}(\hat{\theta}^{(N)}) = \theta^*, \quad (14)$$

and, moreover, satisfies

$$\partial_i E_\theta[(\hat{\theta}^{(N)})^j] |_{\theta=\theta^*} = \delta_i^j. \quad (15)$$

(D3): An asymptotically unbiased estimator is an estimator such that

$$\lim_{N \rightarrow \infty} E_\theta(\hat{\theta}^{(N)}) = \theta. \quad (16)$$

A typical (classical) estimation protocol consists in sampling data $x \in \mathcal{X}^{\times N}$ and the processing them using an estimator $\hat{\theta}^{(N)}$, finally providing an estimate $\hat{\theta}^{(N)}(x)$ of the true value. If the estimator is unbiased, the estimate will fluctuate around the

true value θ^* over many independent repetitions of the protocol. To quantify the performance of an estimator, it is usual to take as a figure of merit its mean square error:

$$[\text{MSE}(\hat{\theta}^{(N)})]_{ij} := E_{\theta}([\hat{\theta}^{(N)}]^i - \theta)[\hat{\theta}^{(N)}]^j - \theta). \quad (17)$$

Estimators with a smaller MSE are said to perform better than estimators with a larger one. Notice that for unbiased estimators, the MSE matrix coincides with the covariance matrix $[\text{Cov}(\hat{\theta}^{(N)})]_{ij}$. The following theorem provides a lower bound to the covariance matrix of unbiased estimators.^{11,12}

Theorem 17 (Cramér–Rao). *If S is a classical statistical model and $\hat{\theta}^{(N)}$ an unbiased estimator, its covariance matrix is bounded from below as follows.*

$$\text{Cov}(\hat{\theta}^{(N)}) \geq \frac{1}{N} [\mathcal{F}_C(\theta)]^{-1}, \quad (18)$$

where \mathcal{F}_C is the Fisher information matrix of S .

The proof of Theorem 17 amounts to an application of the Cauchy–Schwarz inequality of Proposition 6. Under the weaker assumption that $\hat{\theta}^{(N)}$ is only locally unbiased, inequality (18) still holds, but only at $\theta = \theta^*$. Notice that the Cramér–Rao theorem only provides a lower-bound: it does not guarantee that an estimator achieving the bound actually exists. If such an estimator exists, it is said to be *efficient*. An efficient estimator is the best unbiased estimator, since it minimizes the MSE among all unbiased estimators. Unfortunately, efficient estimators exist only under special circumstances (when the statistical model is of the exponential type and the parameters are its natural parameters, see e.g. Ref. 13). Finding the best unbiased estimator becomes then a nontrivial task. The situation improves in the asymptotic limit of a large number of samples. Let us remark that unbiasedness is a strong condition: for some models there exists no such estimator. A far more reasonable condition is that of consistency. A *consistent* estimator is such that, in the limit $N \rightarrow \infty$, its probability density becomes concentrated around θ , i.e. $\forall \epsilon > 0$ and $\forall \theta \in \Theta$, $\lim_{N \rightarrow \infty} \Pr_{\theta}(|\hat{\theta}^{(N)} - \theta| > \epsilon) = 0$, where $\Pr_{\theta}(\cdot)$ denotes the probability of an event computed with respect to p_{θ} . Under mild conditions (e.g. that $\text{Cov}(\hat{\theta}^{(N)})$ is uniformly bounded with respect to the number of samples N), one can prove that a consistent estimator is asymptotically unbiased, i.e. $\lim_{N \rightarrow \infty} E_{\theta}(\hat{\theta}^{(N)}) = \theta$, and satisfies $\lim_{N \rightarrow \infty} \partial_i E_{\theta}[(\hat{\theta}^{(N)})^j] = \delta_i^j$. With the help of the last two properties, one can prove the following asymptotic version of the Cramér–Rao theorem:

$$\lim_{N \rightarrow \infty} N \cdot \text{Cov}(\hat{\theta}^{(N)}) \geq [\mathcal{F}_C(\theta)]^{-1}. \quad (19)$$

A consistent estimator achieving equality is said to be *asymptotically efficient*. Remarkably, asymptotically efficient estimators always exist, e.g. the maximum-likelihood estimator and Bayes estimators are asymptotically efficient.¹³ In conclusion, at the classical level and in the asymptotic regime $N \gg 1$, the optimal protocol

consists in collecting a sample and processing it via an asymptotically efficient estimator; the asymptotic optimal rate at which distinct values of the parameters can be distinguished is given by the inverse Fisher information.

4. Quantum Measurement Theory

The outcomes of a quantum experiment are probabilistic. This means that there must exist a suitable probability measure $\nu_\rho^{(\mathcal{M})}$ such that, if $(\mathcal{X}, \mathcal{B})$ is the measurable space of outcomes (where $\mathcal{X} \subseteq \mathbb{R}$ is the sample space and \mathcal{B} the σ -algebra induced by the natural topology of \mathcal{X}), then the probability of any event $B \in \mathcal{B}$ is $\nu_\rho^{(\mathcal{M})}(B)$. The main difference compared with the classical case is that $\nu_\rho^{(\mathcal{M})}$ is not arbitrary, but is a specific function of both the state of the system ρ and the measurement \mathcal{M} . The mapping $(\rho, \mathcal{M}) \rightarrow \nu_\rho^{(\mathcal{M})}$ is given by Born’s rule. We will deal exclusively with *finite-dimensional* quantum systems, with Hilbert space $\mathcal{H} = \mathbb{C}^d$. A state is a *density matrix* $\rho \in \text{Her}_d^+(\mathbb{C})$, i.e. an Hermitian positive semi-definite matrix, usually normalized such that $\text{Tr}(\rho) = 1$. The set $\mathcal{S}(\mathcal{H})$ of all possible density operators on \mathcal{H} is a convex set. Its extremal elements are the pure states $|\psi\rangle\langle\psi|$, with $|\psi\rangle \in \mathcal{H}$ such that $\langle\psi|\psi\rangle = 1$. The *Hamiltonian* matrix $H \in \text{Her}_d(\mathbb{C})$ completely determines the dynamics of the system (assuming it is isolated from any external environment). That is, if $U_t := \exp(-itH)$ is the matrix exponential of H and ρ_0 is the state at time $t = 0$, then the state of the system at any subsequent time t is $\rho_t := U_t\rho_0U_t^\dagger$. A measurement on a quantum system can be described at three different levels of details. We begin with the first level, which is the more coarse-grained of the three.

(L1) POVM description: At this level, a measurement \mathcal{M} is a mapping that associates to any event $B \in \mathcal{B}$ a positive semi-definite operator $\mathcal{M}(B) \in \text{Her}_d^+(\mathbb{C})$. A few natural requirements are that $\mathcal{M}(\emptyset) = \mathbb{0}_d$; $\mathcal{M}(\mathcal{X}) = \mathbb{1}_d$; if $\{B_i\}_{i=1}^n$ are mutually disjoint measurable sets such that $\bigcup_{i=1}^n B_i := B \in \mathcal{B}$, then $\mathcal{M}(B) = \sum_{i=1}^n \mathcal{M}(B_i)$. These properties imply that \mathcal{M} is a positive-operator valued (probability) measure (POVM) on $(\mathcal{X}, \mathcal{B})$. In particular, they imply that if $\{B_i\}_{i=1}^n$ are mutually disjoint and $\bigcup_{i=1}^n B_i = \mathcal{X}$, then $\sum_{i=1}^n \mathcal{M}(B_i) = \mathbb{1}_d$. Apart for this normalization condition and for being non-negative, the operators $\mathcal{M}(B)$ are completely arbitrary.

The link between a measurement \mathcal{M} and the probability measure $\nu_\rho^{(\mathcal{M})}$ is provided by Born’s rule, i.e.

$$\nu_\rho^{(\mathcal{M})}(B) = \text{Tr}(\rho\mathcal{M}(B)). \tag{20}$$

It can be proven that Born’s rule is actually the unique possibility under a few reasonable assumptions.¹⁴ Equation (20) completely determines the statistics of any quantum experiment.

If \mathcal{X} is a countable sample space, one defines the *probability operators* $\{\Pi_x\}_{x \in \mathcal{X}}$ of a given measurement as follows: $\Pi_x := \mathcal{M}(x)$. The probability operators are

sufficient to compute the probability of any other event. A special case is when each Π_x is a projector P_x , i.e. $P_x^2 = P_x$. One can then associate to the measurement an Hermitian operator $X = \sum_{x \in \mathcal{X}} x P_x$, also called an *observable*. Vice versa, every Hermitian operator gives rise to a projective measurement via its eigendecomposition. An example is the Hamiltonian: a projective measurement over its eigenstates $\{|\xi_j\rangle\}_{j=0}^{d-1}$ is called an energy measurement.

(L2) Instrument description: A POVM description assigns probabilities to measurement outcomes, but does not specify how the state of the system is modified as a result of the measurement. However, quantum measurements can have dynamical effects: if the measurement is nondestructive, the state of the system is updated depending on the outcome. This requires introducing an *instrument*. Formally, an instrument \mathcal{I} is a mapping $\mathcal{B} \rightarrow \mathcal{T}(\mathcal{H})$, where $\mathcal{T}(\mathcal{H})$ denotes the set of bona fide quantum operations on the system (i.e. completely-positive, trace preserving maps). If $B \in \mathcal{B}$ is the observed event, then the state of the system after the measurement is, by definition, $\mathcal{I}_B(\rho)$. Assuming \mathcal{X} is countable, it is enough to consider the set $\{\mathcal{I}_x\}_{x \in \mathcal{X}}$. It can be proven¹⁵ that the most general form for \mathcal{I}_x is as follows:

$$\mathcal{I}_x(\rho) = \frac{\sum_{j=1}^n M_x^{(j)} \rho M_x^{(j)\dagger}}{\Pr(x)}, \quad (21)$$

where the operators $M_x^{(j)}$ are called *measurement operators* and $\Pr(x) := \text{Tr}(\rho \Pi_x)$. Since the post-measurement state $\mathcal{I}_x(\rho)$ must be normalized, one has the identification

$$\Pi_x = \sum_{j=1}^n M_x^{(j)\dagger} M_x^{(j)}. \quad (22)$$

In particular, if $n = 1$, $\forall x \in \mathcal{X}$, the measurement is said to be *fine-grained*. Notice that, in general, many different instruments correspond to the same positive-operator valued measure. This is true even for fine-grained measurements, since the condition $\Pi_x = M_x^\dagger M_x$ is solved by $M_x = U_x \sqrt{\Pi_x}$, where $\sqrt{\Pi_x}$ is the principal square-root of Π_x but U_x is an arbitrary unitary operator. If the measurement is fine-grained and $U_x = \mathbb{I}_d$, $\forall x \in \mathcal{X}$, the measurement is said to be *bare* and the corresponding instrument is known as the *Lüders instrument*.

(L3) Measurement model description: This is the most detailed level of description of a measurement and is obtained by explicitly modeling the interaction between the system and the measuring apparatus. It is assumed that the system is coupled to an ancillary system with Hilbert space \mathcal{H}_A ; the ancilla is prepared in an initial state $\eta \in \mathcal{S}(\mathcal{H}_A)$; the two systems evolve together for an interaction time t_{int} via a quantum channel $\mathcal{E}^{(t_{\text{int}})} \in \mathcal{T}(\mathcal{H} \otimes \mathcal{H}_A)$; finally, an observable $X = \sum_{x \in \mathcal{X}} x P_x$ on \mathcal{H}_A is measured, producing an outcome $x \in \mathcal{X}$. A

measurement model is therefore a quadruple $(\mathcal{H}_A, \eta, \mathcal{E}^{(t_{\text{int}})}, X)$. It gives rise to a positive-operator valued measure via the relation.

$$\text{Tr}(\rho \Pi_x) = \text{Tr}[\mathcal{E}^{(t_{\text{int}})}(\rho \otimes \eta) \mathbb{I}_d \otimes P_x]. \quad (23)$$

Moreover, it defines an instrument via

$$\mathcal{I}_x(\rho) = \frac{\text{Tr}_A[\mathcal{E}^{(t_{\text{int}})}(\rho \otimes \eta) \mathbb{I}_d \otimes P_x]}{\text{Pr}(x)}, \quad (24)$$

where $\text{Tr}_A(\cdot)$ denotes the partial trace over the ancilla’s degrees of freedom. Clearly, many measurement models can lead to the same instrument. In fact, Ozawa’s theorem¹⁶ states that one can recover all possible instruments just by considering measurement models $(\mathcal{H}_A, \eta, \mathcal{E}^{(t_{\text{int}})}, X)$ where η is pure, $\mathcal{E}^{(t_{\text{int}})}$ is a unitary channel and each P_x is rank-1. More precisely, let H_A be the free Hamiltonian of the ancillary system and H_I the interaction Hamiltonian between the system and the apparatus. Let $\eta = |\phi\rangle\langle\phi|$ be the initial preparation of the ancilla. Then, the unitary channel $\mathcal{U}^{(t)}$ generated by the total Hamiltonian $H_T = H + H_A + H_I$ acts as follows:

$$\rho \otimes |\phi\rangle\langle\phi| \rightarrow \mathcal{U}^{(t)}(\rho \otimes |\phi\rangle\langle\phi|) := U_t \rho \otimes |\phi\rangle\langle\phi| U_t^\dagger, \quad U_t := e^{-itH_T}. \quad (25)$$

From conditions (23) and (24), one may prove that the measurement operators M_x and probability operators Π_x take the following form, respectively,

$$M_x = \langle x | U_{t_{\text{int}}} | \phi \rangle, \quad \Pi_x = \langle \phi | U_{t_{\text{int}}}^\dagger \mathbb{I}_d \otimes P_x U_{t_{\text{int}}} | \phi \rangle. \quad (26)$$

5. Quantum Parameter Estimation

By analogy with the classical case, a *quantum statistical model* is defined as follows.^{17,18}

Definition 18 (Quantum statistical model). Given a quantum system described in the Hilbert space \mathcal{H} , a quantum statistical model S is a family of states, i.e. density operators, in $\mathcal{S}(\mathcal{H})$ labeled by m real parameters $\theta \in \Theta \subset \mathbb{R}^m$.

$$S = \{\rho_\theta : \theta = (\theta^1, \theta^2, \dots, \theta^m) \in \Theta\}, \quad (27)$$

where the parametrization map $\theta \rightarrow \rho_\theta$ is injective, the rank $\text{rk}(\rho_\theta)$ is parameter-independent and ρ_θ can be differentiated as many times as needed with respect to the parameters.

A quantum statistical model typically arises in this way: the system is prepared at time $t = 0$ in an initial state ρ_0 and then goes through a quantum channel $\mathcal{E}_{\theta^*} \in \mathcal{T}(\mathcal{H})$, which depends on the true value θ^* of one or more parameters. The associated model is defined as $\rho_\theta := \mathcal{E}_\theta(\rho_0)$, with $\theta \in \Theta$ and Θ containing, by assumption, the true value θ^* . The mapping $\rho_0 \rightarrow \mathcal{E}_\theta(\rho_0)$ is called the *dynamical encoding*. A typical example is

the unitary channel generated by the system's Hamiltonian, i.e.

$$\rho_\theta = U_t \rho_0 U_t^\dagger, \quad U_t = e^{-itH_\theta}. \quad (28)$$

The parameter θ is usually referred to as a *Hamiltonian parameter*. One then further distinguishes between Hamiltonian *shift* or *phase* parameters and *general* parameters. In the first case, the parameter is just an overall multiplicative constant, i.e. $H_\theta = \theta G$, i.e. it appears linearly in the Hamiltonian. In the second case, the parameter may appear in any way, e.g. nonlinearly, and the eigenvectors $|\xi_{j,\theta}\rangle$ of H_θ generally depend in general on θ . Dynamical encoding is not, however, the only possibility. For certain models, the encoding is static. A typical example is that of a *thermal model*, describing the equilibrium state of a quantum system in contact with a thermal bath,

$$\rho_\beta = \frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})}, \quad (29)$$

where the parameter, conventionally denoted by β , is the inverse temperature of the bath and H is the Hamiltonian of the system. In both cases, given a quantum statistical model $S = \{\rho_\theta\}_{\theta \in \Theta}$, performing a measurement with probability operators $\{\Pi_x\}_{x \in \mathcal{X}}$ gives rise to a classical statistical model, via the relation $p_\theta(x) = \text{Pr}_\theta(x) = \text{Tr}(\rho_\theta \Pi_x)$ (where the sample space \mathcal{X} is henceforth assumed to be countable). Notice that the choice of the measurement to perform is an additional degree of freedom the experimentalist is called to optimize upon, which is not present in the classical case. Furthermore, if the encoding is dynamical, one also has to optimize over the initial state of the probe ρ_0 . As a consequence, the search for optimal quantum estimation protocols is considerably more complicated.

A quantum statistical model can be naturally given the structure of a differentiable manifold. Whereas in the classical case there is a fundamentally unique metric, in the quantum case noncommutativity breaks uniqueness and, in fact, leads to an infinite number of possible metrics. Notice that monotonicity now translates into the requirement that, for any completely-positive, trace-preserving map $\mathcal{E} \in \mathcal{T}(\mathcal{H})$, the difference between the metric on the original statistical model $\{\rho_\theta\}_{\theta \in \Theta}$ and on the derived model $\{\mathcal{E}(\rho_\theta)\}_{\theta \in \Theta}$ is positive semi-definite. In the quantum case, all possible monotone Riemannian metrics have been classified by Petz.¹⁹ Each such metric is in one-to-one correspondence with an operator monotone function, which in turn is one-to-one related to an operator mean. We give the following definition.

Definition 19 (Operator mean). An operator mean $\mathfrak{m} : \text{Her}_d^+ \times \text{Her}_d^+ \rightarrow \text{Her}_d^+$ is a function such that, for any positive semi-definite operators A, B, C, D : **(P1)**: $\mathfrak{m}(A, A) = A$; **(P2)**: $\mathfrak{m}(\alpha A, \alpha A) = \alpha A, \forall \alpha \in \mathbb{R}$; **(P3)**: $A \geq C, B \geq D \Rightarrow \mathfrak{m}(A, B) \geq \mathfrak{m}(C, D)$; **(P4)**: $\mathfrak{m}(UAU^\dagger, UBU^\dagger) = U\mathfrak{m}(A, B)U^\dagger, \forall U$ unitary; **(P5)**: $\mathfrak{m}(A, B) = \mathfrak{m}(B, A)$.

Any function aspiring to be a mean for positive semi-definite matrices should intuitively satisfy conditions **(P1)**–**(P5)**. The following proposition fully characterizes the family of operator means.

Proposition 20. *Every operator mean can be written in the form*

$$\mathfrak{m}^{(f)}(A, B) = \sqrt{A}f\left(\frac{1}{\sqrt{A}}B\frac{1}{\sqrt{A}}\right)\sqrt{A}, \quad (30)$$

where f is an operator monotone function (i.e. a function such that, $\forall A, B \in \text{Her}_d^+$, $A \geq B \Rightarrow f(A) \geq f(B)$) with the constraints $f(1) = 1$ and $f(1/x) = f(x)/x$. Vice versa, any such function gives rise to an operator mean.

Each quantum monotone metric is now put in one-to-one correspondence with a suitable operator mean via Petz’s classification theorem.

Theorem 21 (Petz¹⁹). *If $S = \{\rho_\theta\}_{\theta \in \Theta}$ is a quantum statistical model such that, $\forall \theta \in \Theta$, ρ_θ is full-rank, the generic monotone Riemannian metric on S is of the form:*

$$[\mathcal{F}_Q^{(f)}(\theta)]_{ij} = \text{Tr}(\partial_i \rho_\theta \mathcal{J}^{-1} \partial_j \rho_\theta), \quad (31)$$

where \mathcal{J} is the superoperator $\mathcal{J} = Rf(LR^{-1})$, f is an operator-monotone function satisfying $f(1) = 1$ and $f(1/x) = f(x)/x$, and L (resp. R) is the left (respectively, right) multiplication superoperator, which by definition acts on $\eta \in \mathcal{S}(\mathcal{H})$ as follows:

$$L(\eta) = \rho_\theta \eta, \quad R(\eta) = \eta \rho_\theta. \quad (32)$$

One may rewrite (31) more expressively by introducing the logarithmic derivative operators $L_{i,\theta}^{(f)}$ which satisfy the following relations:

$$\partial_i \rho_\theta = \mathcal{J} L_{i,\theta}^{(f)}, \quad i \in \{1, \dots, m\}. \quad (33)$$

The metric $\mathcal{F}_Q^{(f)}$ can therefore be rewritten as

$$[\mathcal{F}_Q^{(f)}(\theta)]_{ij} = \text{Tr}[\partial_i \rho_\theta L_{j,\theta}^{(f)}] = \text{Tr}[\mathcal{J}(L_{i,\theta}^{(f)})L_{j,\theta}^{(f)}]. \quad (34)$$

For each choice of an operator monotone function f , one obtains a corresponding monotone metric.

(M1) Let us consider the operator monotone function $f_{\text{ari}}(x) = (1+x)/2$. The corresponding operator mean is the *arithmetic mean* since, if A, B are commuting matrices, then $\mathfrak{m}^{(f_{\text{ari}})} = (A+B)/2$. The logarithmic derivative operator $L_{i,\theta}^{(f_{\text{ari}})}$ satisfies, from Eq. (33),

$$\partial_i \rho_\theta = \frac{R+L}{2} L_{i,\theta}^{(f_{\text{ari}})} = \frac{1}{2} \{\rho_\theta, L_{i,\theta}^{(f_{\text{ari}})}\}, \quad (35)$$

so that $L_{i,\theta}^{(f_{\text{ari}})}$ is also called the *symmetric* logarithmic derivative (SLD) of ρ_θ . The corresponding quantum metric is

$$[\mathcal{F}_Q^{(f_{\text{ari}})}(\theta)]_{ij} = \Re \text{tr}(\rho_\theta L_{i,\theta}^{(f_{\text{ari}})} L_{j,\theta}^{(f_{\text{ari}})}), \quad (36)$$

which is usually referred to as the *quantum Fisher information* (QFI) metric and denoted simply by $\mathcal{F}_Q(\theta)$. It can be obtained by “quantizing” the *Bures distance* d_B^2 ,²⁰ in the sense that

$$d_B^2(\rho_\theta, \rho_{\theta+d\theta}) = \frac{1}{4} [\mathcal{F}_Q(\theta)]_{ij} d\theta^i d\theta^j, \quad (37)$$

where $d_B^2(\rho, \sigma) = 2[1 - \sqrt{F(\rho, \sigma)}]$ and $F(\rho, \sigma) = (\text{Tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}])^2$ is the *fidelity*.

- (M2)** The operator monotone function $f_{\text{har}} = 2x/(1+x)$ corresponds to the harmonic mean, since for commuting matrix A, B one has $\mathfrak{m}^{(f_{\text{har}})}(A, B) = 2AB/(A+B)$. From Eq. (33), one finds:

$$\partial_i \rho_\theta = \frac{2LR}{L+R} L_{i,\theta}^{(f_{\text{har}})} \Rightarrow L_{i,\theta}^{(f_{\text{har}})} = \frac{1}{2} \{\rho_\theta^{-1}, \partial_i \rho_\theta\}. \quad (38)$$

The corresponding metric is

$$[\mathcal{F}_Q^{(f_{\text{har}})}]_{ij} = \Re \text{tr}(\partial_i \rho_\theta \partial_j \rho_\theta \rho_\theta^{-1}). \quad (39)$$

- (M3)** The logarithmic mean corresponds to $f_{\text{log}} = (x-1)/\log x$ since, for commuting A and B , $\mathfrak{m}^{(f_{\text{log}})}(A, B) = (B-A)/(\log B - \log A)$. From Eq. (33), one obtains the condition:

$$\partial_i \rho_\theta = \frac{L-R}{\log L - \log R} L_{i,\theta}^{(f_{\text{log}})} \Rightarrow [\log \rho_\theta, \partial_i \rho_\theta] = [\rho_\theta, L_{i,\theta}^{(f_{\text{log}})}]. \quad (40)$$

One can solve for $L_{i,\theta}^{(f_{\text{log}})}$ as follows. First of all, let us recall the identity

$$\log \rho_\theta = \int_0^\infty \frac{dt}{1+t} - \int_0^\infty \frac{dt}{\rho_\theta + t}. \quad (41)$$

The commutator $[\log \rho_\theta, \partial_i \rho_\theta]$ can now be rewritten as follows:

$$\begin{aligned} [\log \rho_\theta, \partial_i \rho_\theta] &= \int_0^\infty dt \left[\partial_i \rho_\theta, \frac{1}{\rho_\theta + t} \right] \\ &= \int_0^\infty dt \left[\partial_i \frac{1}{\rho_\theta + t}, \rho_\theta \right] \\ &= \int_0^\infty dt \left[\rho_\theta, \frac{1}{\rho_\theta + t} \partial_i \rho_\theta \frac{1}{\rho_\theta + t} \right], \end{aligned} \quad (42)$$

where we made use of the fact that, for any invertible matrix M , $\partial_i M^{-1} = -M^{-1}\partial_i M M^{-1}$. From Eq. (42), $L_{i,\theta}^{(f_{\log})}$ can be read-off directly, i.e.

$$L_{i,\theta}^{(f_{\log})} = \int_0^\infty dt \frac{1}{\rho_\theta + t} \partial_i \rho_\theta \frac{1}{\rho_\theta + t}. \quad (43)$$

The corresponding metric is the *Bogoliubov–Kubo–Mori metric*:

$$[\mathcal{F}_Q^{(f_{\log})}]_{ij} = \int_0^\infty dt \text{Tr} \left(\partial_i \rho_\theta \frac{1}{\rho_\theta + t} \partial_j \rho_\theta \frac{1}{\rho_\theta + t} \right). \quad (44)$$

It can be obtained by “quantizing” the *quantum relative entropy* $S(\rho||\sigma)$,²⁰ in the sense that

$$S(\rho_\theta || \rho_{\theta+d\theta}) = \frac{1}{2} [\mathcal{F}_Q(\theta)^{(f_{\log})}]_{ij} d\theta^i d\theta^j, \quad (45)$$

where $S(\rho||\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)]$.

It is also possible to derive a closed-form expression for $\mathcal{F}_Q^{(f)}$, with f an arbitrary operator monotone function. Notice that the superoperators L and R commute. Moreover, if $\rho_\theta = \sum_{k=1}^d p_k |k\rangle\langle k|$ (where $\{|k\rangle\}_{k=1}^d$ are the normalized eigenvectors of ρ_θ), then

$$L|k\rangle\langle l| = p_k |k\rangle\langle l|, \quad R|k\rangle\langle l| = p_l |k\rangle\langle l|. \quad (46)$$

It follows that $\{|k\rangle\langle l|\}_{k,l=1}^d$ is a complete system of eigenvectors for both R and L . They are also the eigenvectors of the superoperator $\mathcal{J} = Rf(LR^{-1})$, with eigenvalues:

$$\mathcal{J}|k\rangle\langle l| = p_l f\left(\frac{p_k}{p_l}\right) |k\rangle\langle l|. \quad (47)$$

Let us expand the symmetric derivative operators as

$$L_{i,\theta}^{(f)} = \sum_{k,l=1}^d \ell_{kl}^{(i)} |k\rangle\langle l|. \quad (48)$$

Notice that since ρ_θ is full-rank, the coefficients $\ell_{kl}^{(i)}$ completely determine $L_{i,\theta}^{(f)}$. Next, one substitutes Eq. (48) into Eq. (42) and compares terms, which leads to the conditions:

$$\ell_{kl}^{(i)} = \begin{cases} \frac{\partial_i p_k}{p_k} & (k = l), \\ \frac{p_l - p_k}{p_l f(p_k/p_l)} \langle k|\partial_i l\rangle & (k \neq l). \end{cases} \quad (49)$$

From Eq. (34) and the previous relation, one finds

$$[\mathcal{F}_Q^{(f)}(\theta)]_{ij} = \sum_{k=1}^d \frac{\partial_i p_k \partial_j p_k}{p_k} + \sum_{l \neq k} \frac{(p_l - p_k)^2}{p_l f(p_k/p_l)} \langle k | \partial_i l \rangle \langle \partial_j l | k \rangle, \quad (50)$$

which is our final result. If the statistical model is not full-rank, one can still recover all possible monotone metrics by extending the metrics of Eq. (31) via a suitable fiber bundle construction (see e.g. Ref. 21). In particular, for a pure model $S = \{|\psi_\theta\rangle\}_{\theta \in \Theta}$, the extension of the metric $\mathcal{F}_Q^{(f)}$ on S exists if and only if $f(0) \neq 0$, in which case it is always proportional to the Fubini–Study metric (which is in fact the unique unitarily invariant metric on pure states²⁰). For instance, the quantum Fisher information metric evaluates to

$$[\mathcal{F}_Q(\theta)]_{ij} = 4\Re[\langle \partial_i \psi_\theta | \partial_j \psi_\theta \rangle + \langle \psi_\theta | \partial_i \psi_\theta \rangle \langle \psi_\theta | \partial_j \psi_\theta \rangle]. \quad (51)$$

See also Ref. 22 for a closed-form expression of $\mathcal{F}_Q(\theta)$ when $1 < \text{rk}(\rho_\theta) < d$.

In spite of the infinite number of possible metrics, Braunstein and Caves²³ have shown that the quantum Fisher information metric $\mathcal{F}_Q(\theta)$ is the only relevant one from an estimation viewpoint. This is true, at least, in the case of uniparametric models (i.e. when there is only one parameter $\theta = \theta^1$ to be estimated), to which from now on we restrict our attention (see however Remark 24). Let us recall that a typical quantum estimation protocol is specified by a triple $(\rho_0, \mathcal{M}, \hat{\theta}^{(N)})$ and can be broken down into the following steps:

- (S1) **Initialization:** The statistical model ρ_θ is prepared by suitably encoding the parameter into an initial state ρ_0 .
- (S2) **Measurement:** A measurement \mathcal{M} is performed, yielding an outcome $x \in \mathcal{X}$. When N independent measurements are taken onto identically prepared systems, one obtains a sample $x \in \mathcal{X}^{\times N}$.
- (S3) **Data processing:** The sample x is processed through the estimator $\hat{\theta}^{(N)}$.

The problem is to optimize over each step in order to minimize a given objective function, which is generally taken to be the mean-square-error $\text{MSE}(\hat{\theta}^{(N)})$. Notice that, among the three steps, only (S1) and (S2) are properly quantum. Moreover, in the asymptotic limit of a large number of sample points, optimization over (S3) is trivially carried out by employing an asymptotically efficient estimator. In contrast, optimization over the measurement step (S2) is a nontrivial task. However, as long as $N \gg 1$, minimization of $\text{MSE}(\hat{\theta}^{(N)})$ is equivalent to maximization of the Fisher information $\mathcal{F}_C(\theta)$ corresponding to the classical statistical model $p_\theta(x) = \text{Tr}(\rho_\theta \Pi_x)$ (with $\{\Pi_x\}_{x \in \mathcal{X}}$ the probability operators of a generic measurement \mathcal{M}). Therefore, the strategy usually followed is first to identify the family \mathcal{F} of measurements which are available to the experimentalist, and then to maximize the Fisher information over all measurements $\mathcal{M} \in \mathcal{F}$.

We now introduce the family of *regular* measurements.

Definition 22 (Regular measurement). A measurement \mathcal{M} is called regular if its probability operators are parameter-independent, i.e.

$$\partial_\theta \Pi_x = 0, \quad \forall x \in \mathcal{X}; \tag{52}$$

otherwise, the measurement is nonregular.

Braunstein and Caves have maximized the Fisher information over the family \mathcal{F}_R of regular measurements.

Theorem 23 (Braunstein-Caves²³). For uniparametric model, the maximum Fisher information, optimized over the family \mathcal{F}_R , is the quantum Fisher information:

$$\mathcal{F}_Q(\theta) = \max_{\mathcal{M} \in \mathcal{F}_R} \mathcal{F}_C(\theta). \tag{53}$$

Proof. For a generic measurement, the Fisher information can be written as

$$\mathcal{F}_C(\theta) = \sum_{x \in \mathcal{X}^*} \frac{[\partial_\theta \text{Tr}(\rho_\theta \Pi_x)]^2}{\text{Tr}(\rho_\theta \Pi_x)}, \tag{54}$$

where $\mathcal{X}^* := \{x \in \mathcal{X} : \text{Tr}(\rho_\theta \Pi_x) \neq 0\}$. Notice that, in Definition 12, summation is only over those outcomes belonging to the support of p_θ . In the quantum case the role of p_θ is taken by $\text{Pr}_\theta(x) = \text{Tr}(\rho_\theta \Pi_x)$, so one should exclude outcomes $x \in \mathcal{X} \setminus \mathcal{X}^*$ for which $\text{Pr}_\theta(x) = 0$. This clarification becomes irrelevant if ρ_θ is full-rank, since then $\mathcal{X} = \mathcal{X}^*$. Equation (54) can be manipulated as follows:

$$\mathcal{F}_C(\theta) = \sum_{x \in \mathcal{X}^*} \frac{\Re^2 \text{Tr}(\rho_\theta L_\theta \Pi_x)}{\text{Tr}(\rho_\theta \Pi_x)} \tag{55}$$

$$\leq \sum_{x \in \mathcal{X}^*} \frac{|\text{Tr}(\rho_\theta L_\theta \Pi_x)|^2}{\text{Tr}(\rho_\theta \Pi_x)} \tag{56}$$

$$= \sum_{x \in \mathcal{X}^*} \frac{|\text{Tr}(\sqrt{\Pi_x} \sqrt{\rho_\theta} \sqrt{\rho_\theta} L_\theta \sqrt{\Pi_x})|^2}{\text{Tr}(\rho_\theta \Pi_x)} \tag{57}$$

$$\leq \sum_{x \in \mathcal{X}^*} \frac{\text{Tr}(\rho_\theta \Pi_x) \text{Tr}(L_\theta \rho_\theta L_\theta \Pi_x)}{\text{Tr}(\rho_\theta \Pi_x)} \tag{58}$$

$$= \sum_{x \in \mathcal{X}^*} \text{Tr}(L_\theta \rho_\theta L_\theta \Pi_x) \tag{59}$$

$$\leq \sum_{x \in \mathcal{X}} \text{Tr}(L_\theta \rho_\theta L_\theta \Pi_x) \tag{60}$$

$$= \text{Tr}(\rho_\theta L_\theta^2) = \mathcal{F}_Q(\theta). \tag{61}$$

In the first line, we have employed the defining relation of the symmetric logarithmic derivative $\partial_\theta \rho_\theta = \{\rho_\theta, L_\theta\}/2$; in the second line, the inequality $\Re^2 z \leq |z|^2, \forall z \in \mathbb{C}$; in the fourth line, the Cauchy–Schwarz inequality; in the sixth, we have extended

summation over all outcomes \mathcal{X} , noting that $\text{Tr}(L_\theta \rho_\theta L_\theta \Pi_x) \geq 0, \forall x \in \mathcal{X}$;^a finally, in the last line, we have made use of the completeness relation $\sum_{x \in \mathcal{X}} \Pi_x = \mathbb{I}_d$. We have thus proved that, for any regular measurement $\mathcal{M} \in \mathcal{F}_R, \mathcal{F}_C(\theta) \leq \mathcal{F}_Q(\theta)$.

We will now show that there always exists a measurement saturating the previous inequality, which will establish the theorem. The above manipulations involved three separate inequalities, that to be simultaneously saturated require: **(R1)**: $\Im \text{Tr}(\rho_\theta L_\theta \Pi_x) = 0, \forall x \in \mathcal{X}^*$; **(R2)**: There exist complex numbers $\{\alpha_x\}_{x \in \mathcal{X}^*}$ such that $\sqrt{\rho_\theta} L_\theta \sqrt{\Pi_x} = \alpha_x \sqrt{\rho_\theta} \sqrt{\Pi_x}$; **(R3)**: $\sum_{x \in \mathcal{X} \setminus \mathcal{X}^*} \text{Tr}(L_\theta \rho_\theta L_\theta \Pi_x) = 0$. It is easy to check that requirements **(R1)**–**(R3)** are satisfied by performing a projective measurement of the symmetric logarithmic derivative L_θ . More precisely, let us remark that the defining relation $\partial_\theta \rho_\theta = \{\rho_\theta, L_\theta\}/2$ determines L_θ only on the support of ρ_θ : outside the support $\text{supp}(\rho_\theta)$, L_θ may be defined in an arbitrary way, compatible with Hermiticity. The SLD L_θ may thus be written as follows:

$$L_\theta = \sum_{x \in \mathcal{X}} \lambda_{x,\theta} |\lambda_{x,\theta}\rangle \langle \lambda_{x,\theta}|, \tag{62}$$

where $\{|\lambda_{x,\theta}\rangle\}_{x \in \mathcal{X} \setminus \mathcal{X}^*}$ are chosen arbitrarily so as to give rise to an orthonormal basis. The eigenvectors and eigenvalues of L_θ are, in general, parameter-dependent. Then, if θ^* is the actual value of the parameter to be estimated, the optimal measurement is described by

$$\Pi_x^{(\text{opt})} = |\lambda_{x,\theta}\rangle \langle \lambda_{x,\theta}|_{\theta=\theta^*}, \quad \forall x \in \mathcal{X}, \tag{63}$$

i.e. the corresponding Fisher information satisfies $\mathcal{F}_C(\theta^*) = \mathcal{F}_Q(\theta^*)$. Notice that, for each $\theta^* \in \Theta$, there is a *different* optimal measurement: it is not required to engineer the measurement so that it satisfies Eq. (63) for any possible value of θ^* . Such a measurement would instead have probability operators $|\lambda_{x,\theta}\rangle \langle \lambda_{x,\theta}|$ and would be nonregular. However, implementing the optimal measurement *does* require to know the value of θ^* for the problem at hand, which is a priori unknown. The obstacle is overcome by employing an *adaptive procedure*, which involves constructing a sequence of estimates $\{\theta_n^*\}$ such that $\theta_n^* \rightarrow \theta^*$ and modifying the implemented measurement at each step so as to match condition (63). See e.g. Ref. 24 for more details. □

Remark 24. One may generalize Theorem 23 to the multiparameter case. The quantum Fisher information $\mathcal{F}_Q(\theta)$ can be proven to be the least monotone metric such that $\mathcal{F}_Q(\theta) - \mathcal{F}_C(\theta)$ is positive semi-definite for any regular measurement. However, equality is not in general attainable, unless the commutativity condition $\text{Tr}(\rho_\theta [L_{i,\theta}, L_{j,\theta}]) = 0$ is satisfied $\forall i, j \in \{1, \dots, m\}$.^{25,26} A widely employed solution²⁷ is to regularize the problem, by changing the objective function to $\text{Tr}[C \cdot \text{Cov}(\theta)]$ (where C is a positive-definite diagonal matrix assigning different weights to different

^aIn fact, $L_\theta \rho_\theta L_\theta$ and Π_x are positive-semidefinite matrices and the trace of the product of two positive semi-definite matrices is always nonnegative.

parameters). However, for this problem, the QFI metric is no longer necessarily the one providing the tightest bound.²⁸

With some caveats, the quantum Fisher information therefore sets the ultimate asymptotic sensitivity bound in uniparametric problems.

Theorem 25 (Quantum Cram er–Rao). *For regular models and any uniparametric estimation protocol $(\rho_0, \mathcal{M}, \hat{\theta}^{(N)})$, where $\mathcal{M} \in \mathcal{F}_R$ and the estimator $\hat{\theta}^{(N)}$ is unbiased, the following inequality holds:*

$$\text{Var}(\hat{\theta}^{(N)}) \geq \frac{1}{N \cdot \mathcal{F}_Q(\theta)}. \tag{64}$$

As in the classical case, the bound (64) is saturable only for a few special statistical models (see Ref. 29 for a precise statement). In contrast, in the asymptotic limit $N \gg 1$, one has that, for any regular measurement and any consistent estimator,

$$\lim_{N \rightarrow \infty} N \cdot \text{Var}(\hat{\theta}^{(N)}) \geq \frac{1}{\mathcal{F}_Q(\theta)}. \tag{65}$$

Equality can be achieved by resorting to the optimal measurement of Eq. (63) and to an asymptotically efficient estimator. The last logical step is to maximize the QFI over the choice of the initial state ρ_0 . To this end, the following *extended convexity* property is going to be useful.

Proposition 26. *Given a quantum statistical model $S = \{\rho_\theta\}_{\theta \in \Theta}$, where each ρ_θ is written as a convex superposition of the form $\rho_\theta = \sum_i \lambda_{i,\theta} \rho_{i,\theta}$, the quantum Fisher information satisfies the inequality:*

$$\mathcal{F}_Q[\rho_\theta] \leq \sum_i \lambda_{i,\theta} \mathcal{F}_Q[\rho_{i,\theta}] + \mathcal{F}_C[\{\lambda_{i,\theta}\}]. \tag{66}$$

The terms in square brackets specify the statistical models on which the (quantum) Fisher information is computed. From Proposition 26, assuming that the system is prepared in the parameter-independent state $\rho_0 = \sum_i \lambda_i \rho_i$ and that the parameter is encoded via a channel \mathcal{E}_θ , one has

$$\mathcal{F}_Q[\rho_\theta] \leq \sum_i \lambda_i \mathcal{F}_Q[\mathcal{E}_\theta(\rho_i)]; \tag{67}$$

notice that the classical term $\mathcal{F}_C[\{\lambda_i\}]$ vanishes since $\partial_\theta \lambda_i = 0$. It follows that the QFI achieves its maximum on the set of pure states. It is not possible, in general, to further determine the optimal preparation, with the significant exception of unitary models.

Let us assume that $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and the encoding is provided by the unitary channel associated to $U_t = \exp(-itH_\theta)$. Then, substituting into Eq. (51), one obtains

$$\mathcal{F}_Q(\theta) = 4[|\langle\psi_\theta|\mathfrak{g}_\theta^2[U_t]|\psi_\theta\rangle - (\langle\psi_\theta|\mathfrak{g}_\theta[U_t]|\psi_\theta\rangle)^2], \tag{68}$$

where $|\psi_\theta\rangle = U_t|\psi_0\rangle$ and $\mathfrak{g}_\theta[U_t] := i\partial_\theta U_t U_t^\dagger$ is the local generator of U_t . Equation (68) may be rewritten as

$$\mathcal{F}_Q(\theta) = 4\text{Var}_{U_t|\psi_0}[\mathfrak{g}_\theta[U_t]] = 4\text{Var}_{|\psi_0\rangle}[U_t^\dagger \mathfrak{g}_\theta[U_t] U_t], \quad (69)$$

where $\text{Var}_{|\psi\rangle}[O]$ is by definition the variance of the operator O over a state $|\psi\rangle$. Let us recall that, by Popoviciu's inequality,³⁰ for any random variable \mathbb{Y} ,

$$\text{Var}(\mathbb{Y}) \leq \frac{(Y - y)^2}{4}, \quad (70)$$

where Y (respectively, y) is the maximum (respectively, minimum) value of \mathbb{Y} and equality holds when \mathbb{Y} is equally distributed over the two values Y and y . Let us also introduce the following standard notation for the eigenvalues of a matrix $M \in \text{Her}_d(\mathbb{C})$: if we denote the d real eigenvalues of M by $\text{spec}(M) = \{\lambda_1(M), \dots, \lambda_d(M)\}$ (ordered nondecreasingly, i.e. $\lambda_1(M) \geq \dots \geq \lambda_d(M)$), it then follows that

$$\begin{aligned} \mathcal{F}_Q(\theta) &\leq [\lambda_1(U_t^\dagger \mathfrak{g}_\theta[U_t] U_t) - \lambda_d(U_t^\dagger \mathfrak{g}_\theta[U_t] U_t)]^2 = [\lambda_1(\mathfrak{g}_\theta[U_t]) - \lambda_d(\mathfrak{g}_\theta[U_t])]^2 \\ &= [\sigma(\mathfrak{g}_\theta[U_t])]^2, \end{aligned}$$

where the *spectral gap* of a matrix $M \in \text{Her}_d(\mathbb{C})$ is defined as $\sigma[M] := \lambda_1(M) - \lambda_d(M)$. The equal sign is achieved by preapring a (any) balanced superposition of the extremal eigenvectors of the generator. Overall, we may summarize the result by the following proposition.

Proposition 27. *Given the unitary model $\{\rho_\theta\}_{\theta \in \Theta}$, with $\rho_\theta = U_t \rho_0 U_t^\dagger$ and $U_t = \exp(-itH_\theta)$, one has*

$$\max_{\rho_0} \mathcal{F}_Q[U_t \rho_0 U_t^\dagger] = [\sigma(\mathfrak{g}_\theta[U_t])]^2. \quad (71)$$

The maximum is reached upon setting $\rho_0 = |\psi_0^{(\text{opt})}\rangle\langle\psi_0^{(\text{opt})}|$, where $|\psi_0^{(\text{opt})}\rangle$ is a balanced superposition of the extremal eigenvectors of the generator $\mathfrak{g}_\theta[U_t]$:

$$|\psi_0^{(\text{opt})}\rangle = \frac{1}{\sqrt{2}} (|\lambda_1(\mathfrak{g}_\theta[U_t])\rangle + e^{i\phi} |\lambda_d(\mathfrak{g}_\theta[U_t])\rangle), \quad \phi \in \mathbb{R}. \quad (72)$$

6. Nonregular Measurements and Parameter Estimation Beyond the Quantum Cramér–Rao Theorem

Let us now extend the theory of quantum parameter estimation, by enlarging the class of measurements under consideration to nonregular measurements, i.e. measurements carrying an intrinsic dependence on the unknown value of the parameter. Such measurements will be shown to lead to an improvement of the achievable precision, beyond the bound encoded by the quantum Cramér–Rao theorem.^{5,6}

A measurement \mathcal{M}_θ is said to be *nonregular* if its probability operators $\{\Pi_{x,\theta}\}_{x \in \mathcal{X}}$ are parameter-dependent. Since nonregular measurements, by definition, do not

belong to the family \mathcal{F}_R over which the Fisher information was optimized in Theorem 23, they might outperform the optimal Braunstein–Caves measurement. Explicitly, their Fisher information $\mathcal{F}_C(\theta)$ reads

$$\begin{aligned} \mathcal{F}_C(\theta) = & \sum_{x \in \mathcal{X}^*} \frac{\Re^2 \text{Tr}(\rho_\theta L_\theta \Pi_{x,\theta})}{\text{Tr}(\rho_\theta \Pi_{x,\theta})} + \sum_{x \in \mathcal{X}^*} \frac{[\text{Tr}(\rho_\theta \partial_\theta \Pi_{x,\theta})]^2}{\text{Tr}(\rho_\theta \Pi_{x,\theta})} \\ & + 2 \sum_{x \in \mathcal{X}^*} \frac{\Re \text{Tr}(\rho_\theta L_\theta \Pi_{x,\theta}) \text{Tr}(\rho_\theta \partial_\theta \Pi_{x,\theta})}{\text{Tr}(\rho_\theta \Pi_{x,\theta})}. \end{aligned} \quad (73)$$

The first term on the RHS is the same that appears on the first line of Eq. (55) and that is bounded from above by the QFI, but there are also two additional contributions. In general, they will have an important effect on the achievable sensitivity (though they are not always positive, so a precision enhancement is not guaranteed).

It is not immediately clear how to implement nonregular measurements. Seemingly, one would need to know beforehand the true value of the parameter. The same could be said of the statistical model ρ_θ but, in the latter case, the true value of the parameter is encoded into the initial state, e.g. by making use of the time-evolution of the system as a resource. In the same way, a nonregular measurement requires the parameter to be suitably encoded into its probability operators. We now describe two scenarios where this is possible.

6.1. Measurement models with parameter-dependent interactions

Let us model a nonregular measurement as in Sec. 4, by specifying the interaction between the system and the apparatus. The total Hamiltonian is $H_T = H_\theta + H_A + H_{I,\theta}$, where we assume that the free Hamiltonian H_A of the apparatus does not depend on the parameter, but the coupling term $H_{I,\theta}$ does. We also assume that the duration of the measurement t_{int} is short and the interaction is strong, such that the free evolution of the two systems may be neglected, i.e. the time-evolution operator during the measurement process may be written as $U_t \sim \exp(-itH_{I,\theta})$. If the apparatus is prepared in a reference state $|\phi\rangle$ and a projective measurement $\{P_x\}_{x \in \mathcal{X}}$ is made on the ancilla after a time t_{int} , the resulting probability operators read

$$\Pi_{x,\theta} = \langle \phi | e^{itH_{I,\theta}} \mathbb{I}_d \otimes P_x e^{-itH_{I,\theta}} | \phi \rangle \quad (74)$$

and are, in general, parameter-dependent. A simple example of this scenario is provided by the estimation of the frequency of a bosonic mode, see a schematic diagram in Fig. 1.

The parameter to be estimated is the frequency ω of a bosonic mode in a cavity. The system’s Hamiltonian is $H_\omega = \omega(a^\dagger a + 1/2)$, the initial state is chosen as $|\psi_0\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and the statistical model at time t is $|\psi_\omega\rangle := U_t|\psi_0\rangle = \alpha_0 e^{-i\omega t/2}|0\rangle + \alpha_1 e^{-3i\omega t/2}|1\rangle$, where $U_t := \exp(-itH_\omega)$. The QFI may be written as

$$\mathcal{F}_Q(\omega) = 4t^2 |\alpha_0|^2 |\alpha_1|^2,$$

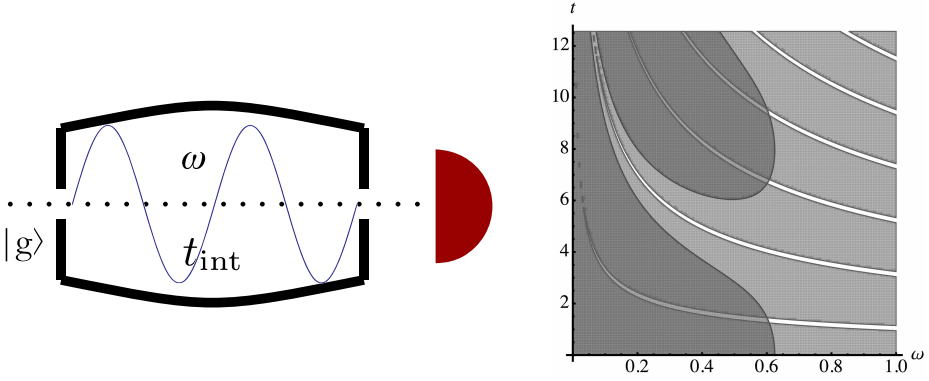


Fig. 1. Ancilla-assisted estimation of the frequency of a bosonic mode in a cavity. A nonregular measurement can be engineered by coupling the bosonic mode to a two-level atom, which is initially in its ground state $|g\rangle$, and by measuring whether the atom has been excited or not after an interaction time t . In the right panel, we show the regions in the $\omega - t$ plane where the ratio $\gamma = \mathcal{F}_C(\omega)/\mathcal{F}_Q(\omega)$ is larger than one. The dark-gray region is for $\Omega = 0.5\sqrt{\omega}$ and the light-gray one for $\Omega = 1.5\sqrt{\omega}$.

which is the maximum information extractable via regular measurements. A non-regular measurement can be engineered by coupling the bosonic mode to a two-level atom, which is initially in its ground state $|g\rangle$, and by measuring whether the atom has been excited or not after an interaction time t_{int} . The interaction Hamiltonian is of the Jaynes–Cummings type $H_I = \Omega(a^\dagger\sigma_- + a\sigma_+)$, where $\Omega := d\sqrt{\omega/2\epsilon_0 V}$, $d := \epsilon \cdot \langle e|\mathbf{d}|g\rangle$, ϵ is the photon polarization, ϵ_0 is the dielectric constant, V the volume of the cavity, \mathbf{d} the dipole operator, $|g\rangle$ the atom’s ground state, $|e\rangle$ the excited state, $\sigma_+ := |e\rangle\langle g|$ and $\sigma_- := |g\rangle\langle e|$. Notice that $\Omega = \kappa\sqrt{\omega}$ with $\kappa = d/V\sqrt{2\epsilon_0}$, such that the interaction Hamiltonian is parameter-dependent. Explicitly, the evolution operator U_t during the measurement process is

$$U_t = U_{gg}|g\rangle\langle g| + U_{ge}|g\rangle\langle e| + U_{eg}|e\rangle\langle g| + U_{ee}|e\rangle\langle e|, \quad (75)$$

where letting $N := a^\dagger a$ denote the number operator for the radiation field, we have defined

$$\begin{aligned} U_{gg} &:= \cos(\Omega t\sqrt{N}), & U_{ge} &:= -i \frac{\sin(\Omega t\sqrt{N})}{\sqrt{N}} a^\dagger, \\ U_{eg} &:= -i \frac{\sin(\Omega t\sqrt{1+N})}{\sqrt{1+N}} a, & U_{ee} &:= \cos(\Omega t\sqrt{1+N}). \end{aligned} \quad (76)$$

By convention, the outcome 0 is obtained if the atom is measured in the ground state and the outcome 1 if measured in the excited state. From Eq. (26), the measurement operators and the corresponding probability operators are given by

$$M_{0,\omega} = \langle g|U_t|g\rangle = \cos(\Omega t\sqrt{N}), \quad M_{1,\omega} = \langle e|U_t|g\rangle = -i \frac{\sin(\Omega t\sqrt{1+N})}{\sqrt{1+N}} a, \quad (77)$$

$$\Pi_{0,\omega} = \cos^2(\Omega t\sqrt{N}), \quad \Pi_{1,\omega} = \sin^2(\Omega t\sqrt{N}). \quad (78)$$

They depend on the parameter ω via the coupling constant Ω . The Fisher information is then given by

$$\mathcal{F}_C(\omega) = \left(\frac{\Omega t}{\omega}\right)^2 \frac{|\alpha_1|^2 \cos^2(\Omega t)}{1 - |\alpha_1|^2 \sin^2(\Omega t)}, \quad (79)$$

which is not necessarily bounded from above by the QFI. For instance, if the system is initially prepared in the excited state, then the QFI vanishes (there is no regular measurement that can estimate the parameter with finite precision), but $\mathcal{F}_C(\omega) = (\Omega t/\omega)^2$. More generally, for small values of α_0 we have a diverging ratio $\mathcal{F}_C(\omega)/\mathcal{F}_Q(\omega) \simeq \Omega^2/(4\omega^2|\alpha_0|^2)$, and we have $\mathcal{F}_C(\omega)/\mathcal{F}_Q(\omega) > 1$ for values of α_0 satisfying the condition

$$|\alpha_0|^2 < \frac{\sqrt{1 + \frac{\Omega^2}{\omega^2} \tan^2(\Omega t)} - 1}{2 \tan^2(\Omega t)}. \quad (80)$$

In the right panel of Fig. 1 we show the regions in the $\omega - t$ plane where the ratio $\gamma = \mathcal{F}_C(\omega)/\mathcal{F}_Q(\omega)$ is larger than one. The dark region is for $\Omega = 0.5\sqrt{\omega}$ and the light one for $\Omega = 1.5\sqrt{\omega}$.

6.2. Energy measurements of nonlinear Hamiltonians

If the Hamiltonian H_θ depends on the parameter θ in a nonlinear way (i.e. it is not of the form $H_\theta = \theta G$), its eigenstates $\{|\xi_{j,\theta}\rangle\}_{j=0}^{d-1}$ are in general parameter-dependent. An energy measurement corresponds to the projective probability operators $\Pi_{\xi_{j,\theta}} = |\xi_{j,\theta}\rangle\langle\xi_{j,\theta}|$, thus the measurement is nonregular. As an example, let us consider the estimation of the strength g of a uniform gravitational field. The probing system is a mechanical oscillator, with Hamiltonian $H_g = -\partial_x^2/2m + kx^2/2 + mgx$, where m is the mass of the oscillator, k its elastic constant and x denotes the vertical displacement of the oscillator from equilibrium, see Fig. 2.

The energy eigenstates have the following wavefunctions:

$$\psi_j = \left(\frac{m\omega}{\pi}\right)^{1/4} \frac{1}{\sqrt{2^j j!}} H_j(\varkappa + \varkappa_g) e^{-(\varkappa + \varkappa_g)^2/2}, \quad (81)$$

where $j \in \mathbb{N}_0$, H_j is the j th Hermite polynomial, $\omega := \sqrt{k/m}$, \varkappa is the dimensionless coordinate $\varkappa := x/\ell$, ℓ is the characteristic length of the oscillator $\ell := 1/\sqrt{m\omega}$ and $\varkappa_g := mg/k\ell$. The corresponding eigenvalues are $\xi_{j,g} = \omega(j + 1/2) - mg^2/2\omega^2$. At time $t = 0$, the oscillator is cooled to its ground state ψ_0 ; it is henceforth mechanically displaced from its equilibrium point by a distance δx , so that the initial state is

$$\psi(x, 0) = \left(\frac{m\omega}{\pi}\right)^{1/4} e^{-(\varkappa + \varkappa_\delta)^2/2}, \quad \varkappa_\delta := \delta x/\ell. \quad (82)$$

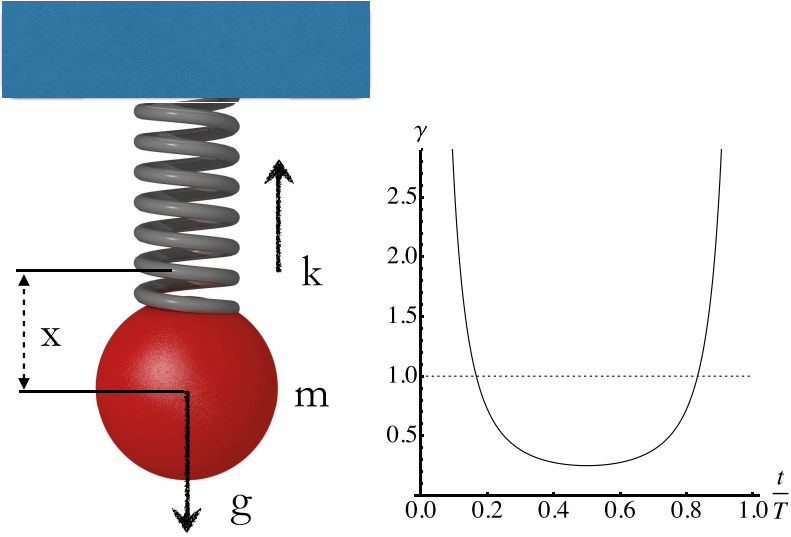


Fig. 2. Estimation of the strength g of a uniform gravitational field using a (quantum) mechanical oscillator as a probing system. The mass of the oscillator is m , k its elastic constant, and x denotes the vertical displacement of the oscillator from equilibrium. In the right panel, we show the ratio $\gamma = \mathcal{F}_C(g)/\mathcal{F}_Q(g)$ as a function of t/T , $T = 2\pi/\omega$ being the period of the oscillator.

At the generic time t , the wavefunction of the oscillator reads

$$\psi(x, t) = \left(\frac{m\omega}{\pi}\right)^{1/4} e^{-i\omega t(1-\kappa_g^2)/2} e^{-(x+\kappa_g)^2/2} \exp[\Phi_g], \quad (83)$$

where

$$\Phi_g = -e^{-i\omega t} \left(\frac{(\kappa_\delta - \kappa_g)^2}{2} \cos \omega t + (\kappa_\delta - \kappa_g)(\kappa + \kappa_g) \right). \quad (84)$$

The computation of the QFI for the statistical model of Eq. (84) can be carried out straightforwardly (see Ref. 5 for details); the final result is

$$\mathcal{F}_Q(g) = \frac{8m}{\omega^3} \sin^2\left(\frac{\omega t}{2}\right). \quad (85)$$

It should be compared with the Fisher information $\mathcal{F}_C(g)$ corresponding to an energy measurement, which is independent on time and given by $\mathcal{F}_C(g) = 2m/\omega^3$. Notice that $\mathcal{F}_C(g)$ exceeds the QFI for certain values of the interrogation time t , i.e. whenever $|\sin(\omega t/2)| < 1/2$, e.g. for $\omega t < \pi/3$ (see the left panel of Fig. 2, where we show the ratio $\gamma = \mathcal{F}_C(g)/\mathcal{F}_Q(g)$ as a function of t/T , $T = 2\pi/\omega$ being the period of the oscillator).

7. Nonregular Estimation of General Hamiltonian Parameters

In this section, we further study nonregular estimation protocols based on energy measurements of nonlinear Hamiltonians. The plan is to introduce a family of

measurements that are nonregular and have a clear-cut physical interpretation; to maximize the Fisher information over such a family; to identify the best-performing measurement and, finally, to compare it with the optimal Braunstein–Caves measurement.

7.1. Controlled energy measurements

Let us consider a projective measurement of H_θ , with θ a general Hamiltonian parameter. It is assumed that H_θ has eigenvalues $\xi_{j,\theta} = \lambda_{d-j}(H_\theta)$. With no significant loss of generality, the spectrum is taken to be nondegenerate. The probability of each measurement outcome is

$$\Pr_\theta(\xi_{j,\theta}) = \text{Tr}(\rho_\theta P_{\xi_{j,\theta}}) = \langle \xi_{j,\theta} | \rho_0 | \xi_{j,\theta} \rangle, \quad (86)$$

where $\rho_\theta = U_t \rho_0 U_t^\dagger$, $U_t = \exp(-itH_\theta)$ and $P_{\xi_{j,\theta}} = |\xi_{j,\theta}\rangle\langle\xi_{j,\theta}|$. The corresponding sample space $\mathcal{X}_\theta = \{\xi_{j,\theta}\}_{j=0}^{d-1}$ is, in general, parameter-dependent. This is a significant complication, since there is no established theory for statistical models with parameter-dependent sample spaces. In fact, if the sample space is allowed to depend on θ , the proof of the classical Cramèr–Rao theorem, Theorem 23, breaks down. In some cases, it is even possible to construct unbiased estimators having vanishing variance.³¹ To exclude such pathological situations, we assume in the following that either the eigenstates of H_θ are parameter-dependent, but not its eigenvalues; or that the outcomes of an energy measurement are processed via a suitable statistic $Y : \mathcal{X}_\theta \rightarrow \mathcal{Y}$, where \mathcal{Y} is a conventional parameter-independent sample space. Within these assumptions, estimators having vanishing variance no longer occur and the Fisher information is again providing the relevant bounds. Let us now consider a specific family of nonregular measurements, which were referred to as *controlled energy measurements*. They are obtained by first applying a unitary control $V \in U(d)$ and then performing a projective energy measurement (see Fig. 3).

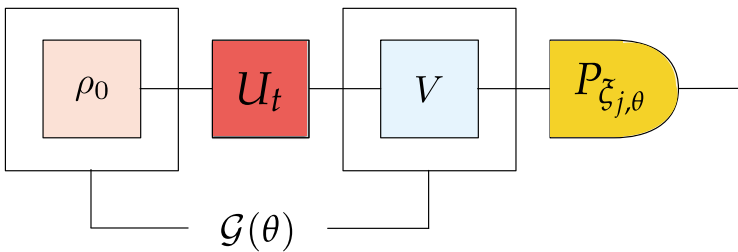


Fig. 3. Schematic diagram of an estimation strategy based on a controlled energy measurement. The optimal performance is quantified by $\mathcal{G}(\theta)$, see Eq. (87), which is optimized over both the preparation stage and the unitary control. The different stages of the schemes correspond to preparation (ρ_0), encoding (U_t), control (V), and energy measurement ($P_{\xi_{j,\theta}}$).

Definition 28 (Controlled energy measurement). A controlled energy measurement $\mathcal{M}_\theta^{(V)}$ has sample space $\mathcal{Y} = \{\zeta_j := Y(\xi_{j,\theta})\}_{j=0}^{d-1}$ and probability operators $\{\Pi_{\zeta_j}\}_{j=0}^{d-1}$, where $\Pi_{\zeta_j} := V^\dagger P_{\xi_{j,\theta}} V$, $V \in U(d)$ is a unitary parameter-independent control and $P_{\xi_{j,\theta}}$ is the projector over the j th energy eigenstate of H_θ .

The Fisher information of a controlled energy measurement $\mathcal{M}_\theta^{(V)}$ is denoted by $\mathcal{F}_C^{(V)}(\theta)$. Let us remark that an energy measurement corresponds to the choice $V = \mathbb{I}_d$. Its probability measure (see Eq. (86)) is t -independent, which implies that also the Fisher information does not depend on t . In contrast, the QFI generically grows quadratically with t .³² Therefore, for sufficiently long times, an energy measurement can never outperform the optimal Braunstein–Caves measurement. If, on the other hand, a unitary control is applied before performing the measurement, then the Fisher information $\mathcal{F}_C^{(V)}(\theta)$ may grow again like t^2 and, in fact, it may even outperform the optimal Braunstein–Caves measurement at any t , as it will be discussed in the following. If an experimentalist is allowed to implement arbitrary controlled energy measurements, the maximum Fisher information she can extract is

$$\mathcal{G}(\theta) := \max_{\rho_\theta} \max_{V \in U(d)} \mathcal{F}_C^{(V)}[\rho_\theta]. \quad (87)$$

Compared with regular measurements, an enhancement is achievable if and only if $\mathcal{G}(\theta) > [\sigma(\mathfrak{g}_\theta[U_t])]^2$. However, computing $\mathcal{G}(\theta)$ directly from its definition is a non-trivial task. In the following section, a closed-form formula for $\mathcal{G}(\theta)$ is derived under the assumption that the Hamiltonian H_θ satisfies a rather general condition.

7.2. A tight achievable bound for the precision of controlled energy measurements

For a generic controlled energy measurement $\mathcal{M}_\theta^{(V)}$, the probability of the outcome ζ_j is

$$\Pr_\theta(\zeta_j) = \text{Tr}(\rho_\theta V^\dagger P_{\xi_{j,\theta}} V). \quad (88)$$

Let us denote by $\{|j\rangle\}_{j=0}^{d-1}$ the computational basis on the Hilbert space \mathcal{H} of the system. The two orthonormal basis $\{|j\rangle\}_{j=0}^{d-1}$ and $\{|\xi_{j,\theta}\rangle\}_{j=0}^{d-1}$ are connected by a unitary transformation, denoted by $S \in U(d)$, such that $|j\rangle = S|\xi_{j,\theta}\rangle$. Explicitly, the matrix elements of S are $\langle j|S|k\rangle = \langle \xi_{j,\theta}|k\rangle$. Notice that, for a general Hamiltonian parameter, the matrix S is θ -dependent and that S reduces H_θ to diagonal form, i.e. $SH_\theta S^\dagger = \text{diag}(\xi_{0,\theta}, \dots, \xi_{d-1,\theta})$. One may thus rewrite Eq. (88) as follows:

$$\Pr_\theta(\zeta_j) = \text{Tr}[(SVU_t)\rho_0(SVU_t)^\dagger P_j] = \text{Tr}(\tilde{U}^{(V)}\rho_0\tilde{U}^{(V)\dagger}P_j), \quad (89)$$

where $P_j := |j\rangle\langle j|$ and all dependence on θ has been collected into the unitary matrix $\tilde{U}^{(V)} := SVU_t$. Formally, a controlled energy measurement on the model ρ_θ is

equivalent to a projective measurement in the computational basis on the model $\rho_\theta^{(V)} := \tilde{U}^{(V)} \rho_0 \tilde{U}^{(V)\dagger}$. The Fisher information corresponding to $\mathcal{M}_\theta^{(V)}$ can thus be written as

$$\mathcal{F}_C^{(V)}(\theta) = \sum_{j \in \mathcal{J}^*} \frac{[\partial_\theta \text{Tr}(\rho_\theta^{(V)} P_j)]^2}{\text{Tr}(\rho_\theta^{(V)} P_j)}, \quad (90)$$

where \mathcal{J}^* is the subset of $\mathcal{J} := \{0, \dots, d-1\}$ such that $j \in \mathcal{J}^*$ if and only if $\text{Pr}_\theta(\zeta_j) \neq 0$. The task is to maximize the RHS of Eq. (90) over the unitary group $U(d)$ of available controls V and over the initial preparation ρ_0 .

Theorem 29. *The maximum Fisher information $\mathcal{G}(\theta)$ that can be extracted via controlled energy measurements satisfies the inequality*

$$\mathcal{G}(\theta) \leq [\sigma(\mathfrak{g}_\theta[U_t]) + \sigma(\mathfrak{g}_\theta[S])]^2, \quad (91)$$

where $U_t = \exp(-itH_\theta)$ is the unitary encoding, S is the similarity transformation diagonalizing H_θ , $\mathfrak{g}_\theta[U_t]$ (respectively, $\mathfrak{g}_\theta[S]$) is the generator of U_t (respectively, S), i.e.

$$\mathfrak{g}_\theta[U_t] = i\partial_\theta U_t U_t^\dagger, \quad \mathfrak{g}_\theta[S] = i\partial_\theta S S^\dagger, \quad (92)$$

and $\sigma(M)$ denotes the spectral gap of a matrix $M \in \text{Her}_d(\mathbb{C})$.

Proof. The Fisher information for $\mathcal{M}_\theta^{(V)}$ is given by Eq. (90). Introducing the symmetric logarithmic derivative $L_\theta^{(V)}$ of $\rho_\theta^{(V)}$,

$$\mathcal{F}_C^{(V)}(\theta) = \sum_{j \in \mathcal{J}^*} \frac{\Re^2 \text{Tr}(\rho_\theta^{(V)} L_\theta^{(V)} P_j)}{\text{Tr}(\rho_\theta^{(V)} P_j)}. \quad (93)$$

Using the inequality $\Re z \leq |z|$, $\forall z \in \mathbb{C}$, and then the Cauchy–Schwarz inequality, the numerator can be bounded as follows:

$$\begin{aligned} \Re^2 \text{Tr}(\rho_\theta^{(V)} L_\theta^{(V)} P_j) Z &\leq |\text{Tr}(\rho_\theta^{(V)} L_\theta^{(V)} P_j)|^2 \\ &\leq \text{Tr}(L_\theta^{(V)} \rho_\theta^{(V)} L_\theta^{(V)} P_j) \text{Tr}(\rho_\theta^{(V)} P_j). \end{aligned} \quad (94)$$

Therefore,

$$\mathcal{F}_C^{(V)}(\theta) \leq \sum_{j \in \mathcal{J}^*} \text{Tr}(L_\theta^{(V)} \rho_\theta^{(V)} L_\theta^{(V)} P_j) \leq \sum_{j \in \mathcal{J}} \text{Tr}(L_\theta^{(V)} \rho_\theta^{(V)} L_\theta^{(V)} P_j) = \text{Tr}[\rho_\theta^{(V)} (L_\theta^{(V)})^2]. \quad (95)$$

Taking the maximum over the initial preparation

$$\max_{\rho_0} \mathcal{F}_C^{(V)}(\theta) \leq \max_{\rho_0} \text{Tr}[\rho_\theta^{(V)} (L_\theta^{(V)})^2]. \quad (96)$$

By convexity, the maximum of the expression on the RHS is achieved when the system is prepared in a pure state. Let us set $\rho_0 = |\psi_0\rangle\langle\psi_0|$. One can then rewrite it as

$$\text{Tr}[\rho_\theta^{(V)}(L_\theta^{(V)})^2]_{|\rho_0=|\psi_0\rangle\langle\psi_0|} = 4\text{Var}_{|\psi_0\rangle}(\tilde{U}^{(V)\dagger}\mathfrak{g}_\theta[\tilde{U}^{(V)}]\tilde{U}^{(V)}), \quad (97)$$

where

$$\mathfrak{g}_\theta[\tilde{U}^{(V)}] = \mathfrak{g}_\theta[S] + (SV)\mathfrak{g}_\theta[U_t](SV)^\dagger \quad (98)$$

is the local generator of $\tilde{U}^{(V)}$. By Popoviciu's inequality, we have,

$$\max_{\rho_0} \mathcal{F}_C^{(V)}(\theta) \leq [\sigma(\mathfrak{g}_\theta[S] + (SV)\mathfrak{g}_\theta[U_t](SV)^\dagger)]^2, \quad (99)$$

and after maximizing over the unitary control V ,

$$\mathcal{G}(\theta) \leq \max_{V \in U(d)} [\sigma(\mathfrak{g}_\theta[S] + (SV)\mathfrak{g}_\theta[U_t](SV)^\dagger)]^2. \quad (100)$$

The above maximization may be carried out explicitly. To this aim we employ the following lemma: the maximum spectral gap of the sum of any two Hermitian matrices with given spectra is equal to the sum of their spectral gaps, i.e.

$$\max_{U_1, U_2 \in U(d)} \sigma(U_1 M_1 U_1^\dagger + U_2 M_2 U_2^\dagger) = \sigma(M_1) + \sigma(M_2), \quad M_1, M_2 \in \text{Her}_d(\mathbb{C}). \quad (101)$$

See Ref. 6 for a proof. From Eq. (101), Eq. (91) follows immediately. \square

Let us now discuss tightness of inequality (91). The proof of Theorem 29 can be broken down into three main steps:

(S1) In Eq. (95), the Fisher information $\mathcal{F}_C^{(V)}(\theta)$ was bounded from above. This step actually made use of three different inequalities: the inequality $\Re z \leq |z|$ (on the first line of Eq. (94)), the Cauchy–Schwarz inequality (on the second line of Eq. (94)) and the inequality on the second line of Eq. (95), which follows from:

$$\sum_{j \in \mathcal{J} \setminus \mathcal{J}^*} \text{Tr}(L_\theta^{(V)} \rho_\theta^{(V)} L_\theta^{(V)} P_j) \geq 0. \quad (102)$$

(S2) Next, the quantity on the RHS of Eq. (96) was maximized over the initial preparation ρ_0 , which led to Eq. (99).

(S3) Finally, maximization over the unitary control V was performed.

Steps **(S2)** and **(S3)** are proper maximizations, that can be made tight by implementing the optimal control $V^{(\text{opt})}$ and the optimal initial preparation $|\psi_0^{(\text{opt})}\rangle$. It is easy to check that the optimal control has the form

$$V^{(\text{opt})} = S^\dagger R_1^\dagger R_2, \quad (103)$$

where R_1 (respectively, R_2) is the similarity transformation that diagonalizes $\mathfrak{g}_\theta[S]$ (respectively, $\mathfrak{g}_\theta[U_t]$), with eigenvalues ordered decreasingly, i.e.

$$\begin{aligned} R_1 \mathfrak{g}_\theta[S] R_1^\dagger &= \text{diag}(\lambda_1(\mathfrak{g}_\theta[S]), \dots, \lambda_d(\mathfrak{g}_\theta[S])), \\ R_2 \mathfrak{g}_\theta[U_t] R_2^\dagger &= \text{diag}(\lambda_1(\mathfrak{g}_\theta[U_t]), \dots, \lambda_d(\mathfrak{g}_\theta[U_t])). \end{aligned} \quad (104)$$

Moreover, from Popoviciu’s inequality, the optimal initial preparation is

$$|\psi_0^{(\text{opt})}\rangle = \frac{1}{\sqrt{2}} \tilde{U}^{(V^{(\text{opt})})\dagger} [|\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})})\rangle\rangle + e^{i\phi}|\lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})})\rangle\rangle], \quad \phi \in \mathbb{R}, \quad (105)$$

where $\tilde{U}^{(V^{(\text{opt})})} = SV^{(\text{opt})}U_t$. The previous expression for $|\psi_0^{(\text{opt})}\rangle$ can be slightly simplified by noticing that the extremal eigenvalues of the generator of $\tilde{U}^{(V^{(\text{opt})})}$ coincide with the extremal eigenvalues of the generator of S . This can be proven as follows. From Eqs. (98) and (103), the generator of $\tilde{U}^{(V^{(\text{opt})})}$ can be written as

$$\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})}] = \mathfrak{g}_\theta[S] + R_1^\dagger R_2 \mathfrak{g}_\theta[U_t] R_2^\dagger R_1 = R_1^\dagger D R_1, \quad (106)$$

where D is the diagonal matrix

$$D = \text{diag}[\lambda_1(\mathfrak{g}_\theta[S]) + \lambda_1(\mathfrak{g}_\theta[U_t]), \dots, \lambda_d(\mathfrak{g}_\theta[S]) + \lambda_d(\mathfrak{g}_\theta[U_t])]. \quad (107)$$

Therefore, the extremal eigenvectors of $\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})}]$ are given by

$$|\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})})\rangle\rangle = R_1^\dagger |0\rangle, \quad |\lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V^{(\text{opt})})})\rangle\rangle = R_1^\dagger |d-1\rangle. \quad (108)$$

But, by the very definition of R_1 , $R_1^\dagger |0\rangle = |\lambda_1(\mathfrak{g}_\theta[S])\rangle\rangle$ and $R_1^\dagger |d-1\rangle = |\lambda_d(\mathfrak{g}_\theta[S])\rangle\rangle$, which establishes our claim. One may thus write

$$|\psi_0^{(\text{opt})}\rangle = \frac{1}{\sqrt{2}} \tilde{U}^{(V^{(\text{opt})})\dagger} [|\lambda_1(\mathfrak{g}_\theta[S])\rangle\rangle + e^{i\phi}|\lambda_d(\mathfrak{g}_\theta[S])\rangle\rangle], \quad \phi \in \mathbb{R}. \quad (109)$$

Proving tightness of inequality (91) is therefore equivalent to proving that of step (S1), under the constraints that the control and the initial preparation are chosen according to Eqs. (103) and (109), respectively. Let us first consider the majorization based on the Cauchy–Schwarz inequality, which is saturated if and only if, $\forall j \in \mathcal{J}^*$, there exist complex numbers $\{\alpha_j\}$ such that

$$\sqrt{\rho_\theta^{(V^{(\text{opt})})}} P_j = \alpha_j \sqrt{\rho_\theta^{(V^{(\text{opt})})}} L_\theta^{(V^{(\text{opt})})} P_j. \quad (110)$$

When the model is pure, condition (110) is automatically satisfied since it reduces to

$$\langle \psi_\theta^{(V^{(\text{opt})})} | j \rangle | \psi_\theta^{(V^{(\text{opt})})} \rangle \langle j | = \alpha_j \langle \psi_\theta^{(V^{(\text{opt})})} | L_\theta^{(V^{(\text{opt})})} | j \rangle | \psi_\theta^{(V^{(\text{opt})})} \rangle \langle j | \quad (111)$$

(where we have set $|\psi_\theta^{(V^{(\text{opt})})}\rangle := \tilde{U}^{(V^{(\text{opt})})} |\psi_0^{(\text{opt})}\rangle$), which implies

$$\alpha_j = \frac{\langle \psi_\theta^{(V^{(\text{opt})})} | L_\theta^{(V^{(\text{opt})})} | j \rangle}{\langle \psi_\theta^{(V^{(\text{opt})})} | j \rangle}. \quad (112)$$

The remaining two inequalities used in step (S1) cannot be saturated without making further assumptions about the Hamiltonian H_θ . For the inequality $\Re z \leq |z|$ to be tight, one should have, $\forall j \in \mathcal{J}^*$,

$$\Im [\langle j | L_\theta^{(V^{(\text{opt})})} | \psi_\theta^{(V^{(\text{opt})})} \rangle \langle \psi_\theta^{(V^{(\text{opt})})} | j \rangle] = 0. \quad (113)$$

Upon writing explicitly the SLD $L_\theta^{(V(\text{opt}))}$ and using the optimal preparation given in Eq. (105), one may prove that the inequality is tight provided that

$$|\langle j|\lambda_1(\mathfrak{g}_\theta[S])\rangle| = |\langle j|\lambda_d(\mathfrak{g}_\theta[S])\rangle|, \quad \forall j \in \mathcal{J}^*, \quad (114)$$

i.e. the extremal eigenvectors of the generator of S , written in the computational basis, are such that corresponding entries have the same complex moduli.

It remains to discuss tightness of inequality (102). Let $j \in \mathcal{J} \setminus \mathcal{J}^*$. This is equivalent to $\langle j|\psi_\theta^{(V(\text{opt}))}\rangle = 0$ where, as before,

$$|\psi_\theta^{(V(\text{opt}))}\rangle = \frac{1}{\sqrt{2}} [|\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle + e^{i\phi}|\lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle]. \quad (115)$$

For $\langle j|\psi_\theta^{(V(\text{opt}))}\rangle = 0$ to hold, there are two possibilities: either both

$$\langle j|\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle = 0 \quad \text{and} \quad \langle j|\lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle = 0; \quad (116)$$

or they are different from zero, have the same moduli and the correct phase difference to cancel each other out. This last possibility can be excluded since the phase ϕ is arbitrary and can always be set such that no cancelation occurs. So, the only possibility is for Eq. (116) to hold. Now, to prove tightness, one should show that

$$\langle j|L_\theta^{(V(\text{opt}))}|\psi_\theta^{(V(\text{opt}))}\rangle = 0, \quad \forall j \in \mathcal{J} \setminus \mathcal{J}^*. \quad (117)$$

Using Eqs. (114) and (115), one arrives at the equivalent condition

$$\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\langle j|\lambda_1(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle + \lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])e^{i\phi}\langle j|\lambda_d(\mathfrak{g}_\theta[\tilde{U}^{(V(\text{opt}))}])\rangle = 0, \quad (118)$$

which is trivially satisfied because of Eq. (116). Thus, no additional assumption is needed for equality to hold in Eq. (102). We summarize our results via the following proposition.

Proposition 30. *If H_θ is such that the extremal eigenvectors of the generator of its diagonalizing matrix S satisfy the condition*

$$|\langle j|\lambda_1(\mathfrak{g}_\theta[S])\rangle| = |\langle j|\lambda_d(\mathfrak{g}_\theta[S])\rangle|, \quad \forall j \in \mathcal{J}^*, \quad (119)$$

then the maximum Fisher information extractable via controlled energy measurements is

$$\mathcal{G}(\theta) = [\sigma(\mathfrak{g}_\theta[U_i]) + \sigma(\mathfrak{g}_\theta[S])]^2. \quad (120)$$

The optimal preparation is given by Eq. (105) and the optimal control by Eq. (103).

The condition imposed by Eq. (119) on H_θ may seem quite restricting. However, it turns out to be satisfied for many Hamiltonians of practical use in quantum metrology, see the examples discussed in Sec. 9. Equation (120) thus often provides a way to directly compute $\mathcal{G}(\theta)$, without the need of any optimization procedure.

8. Metrological Applications

In this section, we discuss how to implement controlled energy measurements in a realistic metrological scenario. In principle, a controlled energy measurement requires to apply a unitary control V , and then to measure the energy projectively. The question is how to perform a projective measurement of the Hamiltonian when the Hamiltonian is not fully known. The problem has first been investigated in Refs. 33 and 34. In the following, we associate to each controlled energy measurement $\mathcal{M}_\theta^{(V)}$ a family of measurements, called *realistic controlled energy measurements*, denoted by $\mathcal{M}_{n,m}^{(V)}$ (with $n, m \in \mathbb{N}$), that are experimentally feasible and allow to approximate $\mathcal{M}_\theta^{(V)}$ to any desired level of accuracy (in the sense that, as $n, m \rightarrow \infty$ the probability measure of $\mathcal{M}_{n,m}^{(V)}$ converges to that of $\mathcal{M}_\theta^{(V)}$). Our exposition can be divided into two parts. First, we describe a simplified version, denoted by $\mathcal{M}_n^{(V)}$, which is based on the *phase estimation* algorithm.^{35–37} It is assumed that the experimentalist can implement the controlled time-evolution operator

$$C_{U_t} := |0\rangle\langle 0| \otimes \mathbb{I}_d + |1\rangle\langle 1| \otimes U_t. \quad (121)$$

This is an unrealistic assumption, since C_{U_t} still depends on the true value of the parameter via U_t . Next, we remove such assumption, which will lead to the introduction of realistic controlled energy measurements. In order to implement $\mathcal{M}_n^{(V)}$, one introduces n control qubits, each one having Hilbert space $\mathcal{H}_c = \mathbb{C}^2$. The total Hilbert space is thus $\mathcal{H}_c^{\otimes n} \otimes \mathcal{H}$, with $\mathcal{H} = \mathbb{C}^d$ the Hilbert space of the original system. All the control qubits are initially prepared in their ground state $|0\rangle$, such that at time $t = 0$ the state of the total system is $|0 \cdots 0\rangle\langle 0 \cdots 0| \otimes \rho_0$. Then, a Hadamard gate is applied to each control qubit, i.e. $|0\rangle \rightarrow H|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and the parameter is encoded into the model $\rho_\theta = U_t \rho_0 U_t^\dagger$. Next, the unitary control V is applied. At time t , the state of the system is thus given by

$$\frac{1}{2^n} \sum_{x,y \in \{0,1\}^{xn}} |x_1, \dots, x_n\rangle\langle y_1, \dots, y_n| \otimes V \rho_\theta V^\dagger, \quad (122)$$

where x stands for the generic binary n -string x_1, \dots, x_n and y for the binary string y_1, \dots, y_n .

Next, given an arbitrary unitary U on \mathcal{H} , we define the superoperator \mathcal{C}_U as follows:

$$\mathcal{C}_U[\rho] := C_U \rho C_U^\dagger. \quad (123)$$

For $l = 1, \dots, n$, the n superoperators $\mathcal{C}_{U_\tau^{2^{l-1}}}$ couple the l th control qubit to the main system (here τ represents a free parameter, which corresponds to the typical time of the measurement process). In particular, when $\mathcal{C}_{U_\tau^{2^{l-1}}}$ is applied to $\rho_l := |x_l\rangle\langle y_l| \otimes$

$V\rho_\theta V^\dagger$, one obtains

$$C_{U_\tau^{2^{l-1}}}[\rho_l] = |x_l\rangle\langle y_l| \otimes U_\tau^{x_l 2^{l-1}} V\rho_\theta V^\dagger (U_\tau^\dagger)^{y_l 2^{l-1}}. \quad (124)$$

Denoting by $X = x_1 + 2 \cdot x_2 + \dots + 2^{n-1} \cdot x_n$ the decimal representation of the binary string x , one obtains

$$\frac{1}{2^n} \sum_{X=0}^{2^n-1} \sum_{Y=0}^{2^n-1} |x\rangle\langle y| \otimes U_\tau^X V\rho_\theta V^\dagger (U_\tau^\dagger)^Y. \quad (125)$$

Let us now expand $V\rho_\theta V^\dagger$ on the energy eigenbasis, i.e.

$$V\rho_\theta V^\dagger = \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} c_{jk} |\xi_{j,\theta}\rangle\langle \xi_{k,\theta}|. \quad (126)$$

Equation (125) then becomes

$$\frac{1}{2^n} \sum_{j,k=0}^{d-1} \sum_{X,Y=0}^{2^n-1} c_{jk} e^{-i\tau(X\xi_{j,\theta} - Y\xi_{k,\theta})} |x\rangle\langle y| \otimes |\xi_{j,\theta}\rangle\langle \xi_{k,\theta}|. \quad (127)$$

The subsequent step of the protocol involves the use of inverse quantum Fourier transform QFT^{-1} on the set of n control qubits. The action of QFT^{-1} on the computational basis (of $\mathcal{H}_c^{\otimes n}$) is given by

$$QFT^{-1}|x\rangle = \frac{1}{2^{n/2}} \sum_{Q=0}^{2^n-1} e^{-\frac{2\pi i x Q}{2^n}} |q\rangle, \quad (128)$$

and thus the total state of the system after QFT^{-1} may be written as

$$\frac{1}{2^{2n}} \sum_{j,k=0}^{d-1} \sum_{X,Y=0}^{2^n-1} \sum_{Q,P=0}^{2^n-1} \tilde{c}_{jk} |q\rangle\langle p| \otimes |\xi_{j,\theta}\rangle\langle \xi_{k,\theta}|, \quad (129)$$

where

$$\tilde{c}_{jk} = c_{jk} e^{-iX(\tau\xi_{j,\theta} + \frac{2\pi Q}{2^n})} e^{iY(\tau\xi_{k,\theta} + \frac{2\pi P}{2^n})}. \quad (130)$$

The final step consists in a read-out, i.e. one performs a measurement (in the computational basis) on the n control qubits. The probability $\Pr_\theta(q)$ of obtaining the (binary) string q as outcome is given by

$$\Pr_\theta(q) = \frac{1}{2^{2n}} \sum_{j=0}^{d-1} \sum_{X,Y=0}^{2^n-1} \Pr_\theta(\xi_{j,\theta}) e^{-i(X-Y)\alpha_{j,Q}}, \quad (131)$$

where

$$\alpha_{j,Q} := \tau\xi_{j,\theta} + \frac{2\pi Q}{2^n}, \quad \Pr_\theta(\xi_{j,\theta}) = \langle \xi_{j,\theta} | V\rho_\theta V^\dagger | \xi_{j,\theta} \rangle. \quad (132)$$

After straightforward manipulation, Eq. (131) can also be written as

$$\Pr_\theta(q) = \sum_{j=0}^{d-1} \Pr_\theta(\xi_{j,\theta}) \left(\frac{1}{2^n} \frac{\sin(2^n \alpha_{j,Q}/2)}{\sin(\alpha_{j,Q}/2)} \right)^2. \quad (133)$$

In the limit $n \rightarrow \infty$, $\Pr_\theta(q)$ converges to the probability $\Pr_\theta(\xi_{j,\theta})$, which corresponds to a controlled energy measurement $\mathcal{M}_\theta^{(V)}$.

In order to obtain a controlled energy measurement $\mathcal{M}_{n,m}^{(V)}$ in a realistic scenario, one exploits $\mathcal{M}_n^{(V)}$ and implements the controlled time-evolution operator C_{U_t} using a quantum subroutine referred to as *universal controllization* (see Fig. 4).³³ In order to briefly illustrate the protocol let us address the case $l = 1$ and consider the problem to approximate the action of C_{U_τ} on the state $\rho_1 = |x_1\rangle\langle y_1| \otimes V\rho_\theta V^\dagger$. The transformation C_{U_τ} is obtained (i.e. replaced) by m applications of the superoperator $\Gamma_{U_{\tau/m}}$, constructed as follows (see Fig. 5). At first, an ancilla system with the same dimensionality as the main system, is introduced. The total Hilbert space is $\mathcal{H}_c^{\otimes n} \otimes \mathcal{H} \otimes \mathcal{H}_a$, with $\mathcal{H}_a = \mathbb{C}^d$. The ancillary system is then prepared in a maximally mixed state: The state of the first control qubit, the main system and the ancilla (before application of C_{U_τ}) is thus given by $\rho'_1 = |x_1\rangle\langle y_1| \otimes V\rho_\theta V^\dagger \otimes \mathbb{I}_d/d$. Let us now consider the quantum operation,

$$W_{U_\tau} := C_{\text{SWAP}}(\mathbb{I}_2 \otimes U_\tau \otimes \mathbb{I}_d)C_{\text{SWAP}}, \quad (134)$$

where C_{SWAP} is the controlled-SWAP gate acting on $\mathcal{H}_c \otimes \mathcal{H} \otimes \mathcal{H}_a$ as

$$\begin{aligned} C_{\text{SWAP}}(|0\rangle \otimes |\psi\rangle \otimes |\phi\rangle) &= |0\rangle \otimes |\phi\rangle \otimes |\psi\rangle, \\ C_{\text{SWAP}}(|1\rangle \otimes |\psi\rangle \otimes |\phi\rangle) &= |0\rangle \otimes |\psi\rangle \otimes |\phi\rangle. \end{aligned} \quad (135)$$

At this point, it is crucial to remark that for the realization (implementation) of the transformation W_{U_τ} we do not need to know the form of the Hamiltonian, since only the *uncontrolled* version of the time-evolution operator U_τ is required. Let us divide τ into m subintervals of duration τ/m . During each subinterval, $W_{U_{\tau/m}}$ is applied; then the ancilla

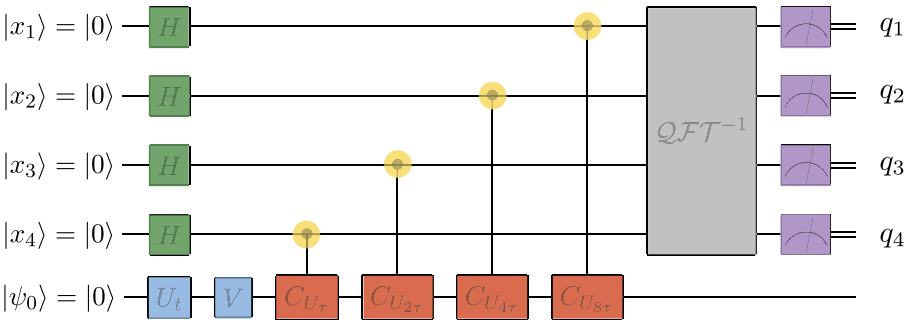


Fig. 4. Circuit diagram of $\mathcal{M}_n^{(V)}$ with $n = 4$ control qubits. In a realistic implementation of controlled energy measurement, one replaces each box C_{U_τ} by its effective implementation, i.e. by m repeated applications of the transformation $\Gamma_{U_{\tau/m}}$, defined in Eq. (136).

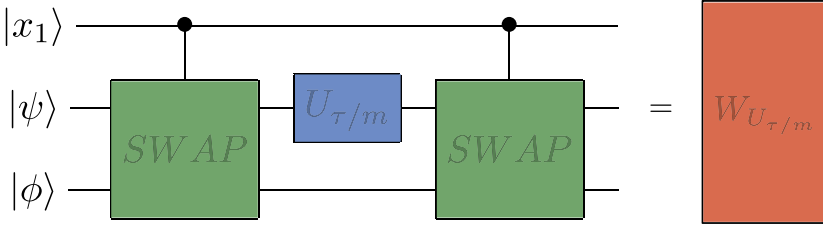


Fig. 5. Circuit diagram of $W_{U_{\tau/m}}$.

is traced out and finally it is reset to its initial state. As for example: after the first interval, one obtains $\Gamma_{U_{\tau/m}}[\rho_1] \otimes \mathbb{I}_d/d$, where

$$\Gamma_{U_{\tau/m}}[\rho_1] := \text{Tr}_{\mathcal{H}_a}(W_{U_{\tau/m}}\rho_1'W_{U_{\tau/m}}^\dagger). \tag{136}$$

A simple computation reveals that

$$\Gamma_{U_{\tau/m}}[\rho_1] = \frac{1}{d} \text{Tr}(U_{\tau/m}^{y_1-x_1})\mathcal{C}_{U_{\tau/m}}[\rho_1]. \tag{137}$$

For future convenience, we write

$$\frac{1}{d} \text{Tr}(U_{\tau/m}) = a_{\tau/m}e^{i\phi_{\tau/m}}, \tag{138}$$

where $a_{\tau/m} \in \mathbb{R}^+$ and $\phi_{\tau/m} \in \mathbb{R}$. Note that, since $x_1 - y_1 \in \{-1, 0, 1\}$, one can write

$$\Gamma_{U_{\tau/m}}^m[\rho_1] = a_{\tau/m}^{|x_1-y_1|}e^{i(y_1-x_1)m\phi_{\tau/m}}\mathcal{C}_{U_{\tau/m}}[\rho_1]. \tag{139}$$

Universal controllization thus replaces $\mathcal{C}_{U_{\tau/m}}$ with $\Gamma_{U_{\tau/m}}^m$. In the limit $m \rightarrow \infty$, it can be proven that the error

$$\epsilon_m := [\text{Tr}(U_{\tau/m})/d]^m - 1 \tag{140}$$

tends to zero. A realistic controlled energy measurement is obtained by substituting each application of $\mathcal{C}_{U_{\tau/m}}$ by $2^{l-1}m$ applications of $\Gamma_{U_{\tau/m}}$. For instance, instead of Eq. (125), one would have

$$\frac{1}{2^n} \sum_{X,Y=0}^{2^n-1} \pi_{X,Y}e^{i(Y-X)m\phi_{\tau/m}}|x\rangle\langle y| \otimes U_{\tau/m}^X V_{\rho_\theta} V_{\rho_\theta}^\dagger (U_{\tau/m}^\dagger)^Y, \tag{141}$$

where

$$\pi_{X,Y} := \prod_{l=1}^n a_{\tau/m}^{|x_l-y_l|}2^{l-1}m. \tag{142}$$

After applying the inverse quantum Fourier transform and measuring in the computational basis, the probability of obtaining the outcome $q \in \{0, 1\}^{\times n}$ is

$$\text{Pr}_\theta(q) = \frac{1}{2^{2n}} \sum_{j=0}^{d-1} \text{Pr}_\theta(\xi_{j,\theta}) \sum_{X,Y=0}^{2^n-1} \pi_{X,Y}e^{i(Y-X)\beta_{j,\theta}}, \tag{143}$$

with

$$\beta_{j,Q} := \alpha_{j,Q} + m\phi_{\tau/m}. \tag{144}$$

Equation (143) can be further expanded by rewriting it as follows:

$$\begin{aligned} \Pr_{\theta}(q) &= \frac{1}{2^{2n}} \sum_{j=0}^{d-1} \Pr_{\theta}(\xi_{j,\theta}) \prod_{l=1}^n \sum_{u,v=0}^1 a_{\tau/m}^{|u-v|2^{l-1}m} e^{i(v-u)2^{l-1}\beta_{j,Q}} \\ &= \frac{1}{2^n} \sum_{j=0}^{d-1} \Pr_{\theta}(\xi_{j,\theta}) \prod_{l=1}^n [1 + a_{\tau/m}^{2^{l-1}m} \cos(2^{l-1}\beta_{j,Q})]. \end{aligned} \tag{145}$$

If $m \rightarrow \infty$, then $\phi_{\tau/m} \rightarrow 0$ and $a_{\tau/m} \rightarrow 1$, so that Eq. (145) converges to Eq. (133). In conclusion, a realistic controlled energy measurement allows to approximate to any desired precision a controlled energy measurement $\mathcal{M}_{\theta}^{(V)}$, without requiring any *a priori* knowledge about the parameter θ .

9. Examples

In this section, we work out a collection of examples. For each example, we compute the QFI $\mathcal{F}_Q(\theta)$ and compare it with $\mathcal{G}(\theta)$. We will find that, in general, $\mathcal{G}(\theta)$ majorizes $\mathcal{F}_Q(\theta)$, thus controlled energy measurements lead to a precision enhancement. Moreover, we study numerically the performance of realistic controlled energy measurements $\mathcal{M}_{n,m}^{(V^{(\text{opt})})}$. From the previous section, as $n, m \rightarrow \infty$, $\mathcal{M}_{n,m}^{(V^{(\text{opt})})}$ converges to $\mathcal{M}_{\theta}^{(V^{(\text{opt})})}$, and thus its Fisher information also converges to \mathcal{G}_{θ} . We will show that, already for relatively small values of n and m , realistic controlled energy measurements perform very close to the the ultimate bound \mathcal{G}_{θ} .

9.1. Estimation of the direction of a magnetic field

Let us consider a situation where the parameter of interest is direction of a magnetic field. More precisely, we want to estimate that the polar angular direction θ of an external magnetic field, whose magnitude B is known. The probing system is a two-level atom, with Hilbert space $\mathcal{H} = \mathbb{C}^2$ and Hamiltonian is $H_{\theta} = \omega(\cos \theta \sigma_z + \sin \theta \sigma_x)$. The energy splitting ω is proportional to the magnitude B of the field and it is thus known. At time $t = 0$, the atom is initialized in its ground state: $|\psi_0\rangle = |0\rangle$. At the generic time t , the state of the probe is $|\psi_{\theta}\rangle = U_t|\psi_0\rangle$, with $U_t := \exp(-iH_{\theta}t)$, see the left panel of Fig. 6 for a schematic diagram.

If an experimentalist is constrained to perform regular measurements, the best performance she can achieve is quantified by the QFI

$$\mathcal{F}_Q(\theta) = 4 \sin^2(\omega t) - \sin^2(2\omega t) \sin^2\theta. \tag{146}$$

Optimizing also over the initial preparation,

$$\max_{|\psi_0\rangle} \mathcal{F}_Q(\theta) = 4 \sin^2(\omega t). \tag{147}$$

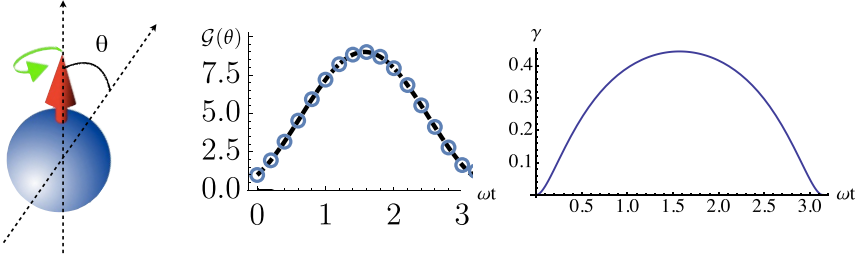


Fig. 6. Left: schematic diagrams of the qubit-based estimation of the direction of a magnetic field. Center: the line corresponds to $\mathcal{G}(\theta)$, computed by Eq. (151), whereas the circular marks correspond to values of $\mathcal{G}(\theta)$ computed by the numerical optimization involved in its definition, see (87). Right: comparison between the optimal Braunstein–Caves measurement and the optimal controlled energy measurement, the plot shows the ratio between the QFI, optimized over the initial preparation and $\mathcal{G}(\theta)$, computed by Eq. (151).

If instead the experimentalist is allowed to implement only controlled energy measurements, the maximum Fisher information that she can extract is given by $\mathcal{G}(\theta)$. To compute $\mathcal{G}(\theta)$, one first computes the matrix S , built from the eigenvectors of H_θ , and its generator $\mathfrak{g}_\theta[S]$

$$S = \begin{pmatrix} -sc(\theta) \sin \frac{\theta}{2} & sc(\theta) \cos \frac{\theta}{2} \\ ss(\theta) \cos \frac{\theta}{2} & ss(\theta) \sin \frac{\theta}{2} \end{pmatrix} \quad \mathfrak{g}_\theta[S] = \begin{pmatrix} 0 & -\frac{i}{2} ss(\theta) \\ \frac{i}{2} sc(\theta) & 0 \end{pmatrix}. \quad (148)$$

where $sc(\theta) = \text{sgn}[\cos \frac{\theta}{2}]$, $ss(\theta) = \text{sgn}[\sin \frac{\theta}{2}]$, and $\text{sgn}(x) := |x|/x$. The extremal eigenvectors of $\mathfrak{g}_\theta[S]$ are then given by

$$|\lambda_1(\mathfrak{g}_\theta[S])\rangle = \frac{1}{\sqrt{2}}(-i, 1)^t, \quad |\lambda_2(\mathfrak{g}_\theta[S])\rangle = \frac{1}{\sqrt{2}}(i, 1)^t. \quad (149)$$

Since condition (119) is satisfied, $\mathcal{G}(\theta)$ can be obtained via Proposition (30). The explicit expressions for U_t and its generator are

$$U_t = \begin{pmatrix} A & B \\ B & A^* \end{pmatrix}, \quad \mathfrak{g}_\theta[U_t] = \begin{pmatrix} -C & D \\ D^* & C \end{pmatrix}, \quad (150)$$

where

$$A = \cos \omega t - i \cos \theta \sin \omega t, \quad C = \frac{1}{2} \sin \theta \sin 2\omega t, \\ B = -i \sin \theta \sin \omega t, \quad D = (\cos \theta \cos \omega t - i \sin \omega t) \sin \omega t.$$

One thus obtains

$$\mathcal{G}(\theta) = (2|\sin(\omega t)| + 1)^2. \quad (151)$$

As an overall check, in the central panel of Fig. 6 we report $\mathcal{G}(\theta)$ computed by Eq. (151) together with its values computed by numerical optimization from its definition (87). In the right panel we instead show a comparison of $\mathcal{G}(\theta)$ with the QFI

in terms of the ratio $\gamma = \max_{|\psi_0\rangle} \mathcal{F}_Q(\theta)/\mathcal{G}(\theta)$, which is apparently below unit at all times.

In order to check whether the above protocol may be of practical interest one may also study numerically the performance of $\mathcal{M}_{n,m}^{(V^{(\text{opt})})}$ (with $V^{(\text{opt})}$ the optimal control of Eq. (103)). Recall that n is the number of ancillary qubits needed to implement the phase estimation algorithm, while m is the number of subintervals the timescale τ is subdivided into. During each subinterval, the action of the controlled time-evolution operator \mathcal{C}_{U_τ} is approximated by applying m times the superoperator $\Gamma_{U_{\tau/m}}$ of Eq. (136). As $n, m \rightarrow \infty$, the probability measure of $\mathcal{M}_{n,m}^{(V^{(\text{opt})})}$ converges to that of the optimal controlled energy measurement. Our results show that already for reasonably small values of the two parameters, say $n = 6, m = 3$, one is close to the ultimate bound $\mathcal{G}(\theta)$.

9.2. Estimation of a component of a magnetic field

The parameter to be estimated is the component of a magnetic field along the x direction. The probing system is again a two-level atom. The Hamiltonian is $H_\theta = -\omega\sigma_z + \theta\sigma_x$, with eigenvalues $\pm\Omega_\theta$ and $\Omega_\theta := \sqrt{\omega^2 + \theta^2}$. We report the matrices U_t and S , with their corresponding generators. For U_t and $\mathfrak{g}_\theta[U_t]$, one obtains

$$U_t = \begin{pmatrix} A & B \\ B & A^* \end{pmatrix}, \quad \mathfrak{g}_\theta[U_t] = \begin{pmatrix} -C & D \\ D^* & C \end{pmatrix}, \quad (152)$$

where

$$A = \cos(\Omega_\theta t) + \frac{i\omega \sin(\Omega_\theta t)}{\Omega_\theta}, \quad C = -\frac{\omega\theta[\sin(2\Omega_\theta t) - 2\Omega_\theta t]}{2\Omega_\theta^3},$$

$$B = -\frac{i\theta \sin(\Omega_\theta t)}{\Omega_\theta}, \quad D = \frac{\sin(2\Omega_\theta t)\omega^2 - i\Omega_\theta \cos(2\Omega_\theta t)\omega + \Omega_\theta(2t\theta^2 + i\omega)}{2\Omega_\theta^3}.$$

For the matrix S and its generator,

$$S = \frac{1}{\sqrt{2\Omega_\theta}} \begin{pmatrix} -\frac{\omega + \Omega_\theta}{\sqrt{\Omega_\theta + \omega}} & \frac{\theta}{\sqrt{\Omega_\theta + \omega}} \\ \frac{\theta}{\sqrt{\Omega_\theta + \omega}} & \frac{\theta}{\sqrt{\Omega_\theta - \omega}} \end{pmatrix}, \quad \mathfrak{g}_\theta[S] = \frac{i\omega}{2\theta^2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (153)$$

The maximum QFI is

$$\max_{|\psi_0\rangle} \mathcal{F}_Q(\theta) = \frac{2}{\Omega_\theta^4} [2\Omega_\theta^2 t^2 \theta^2 - \omega^2 \cos(2\Omega_\theta t) + \omega^2]. \quad (154)$$

Since the eigenvectors of $\mathfrak{g}_\theta[S]$ satisfy condition (119), $\mathcal{G}(\theta)$ can be computed directly and is given by

$$\mathcal{G}(\theta) = \left(\frac{\omega}{\Omega_\theta^2} + \frac{\sqrt{2[2\Omega_\theta^2 t^2 \theta^2 - \omega^2 \cos(2\Omega_\theta t) + \omega^2]}}{\Omega_\theta^2} \right)^2, \quad (155)$$

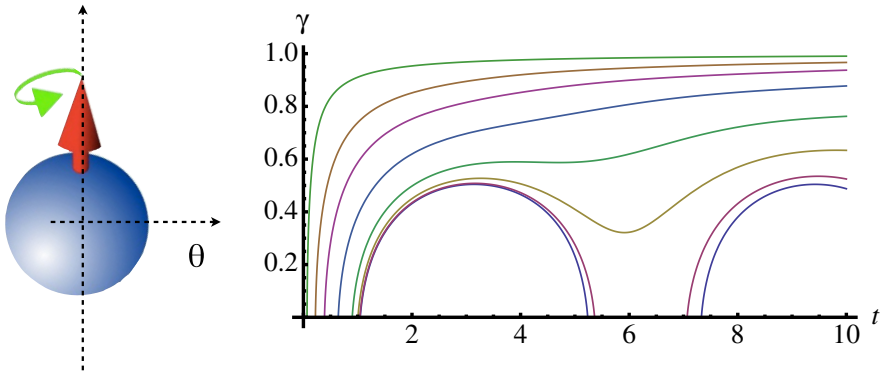


Fig. 7. Left: schematic diagrams of the qubit-based estimation the x -component of a magnetic field. Right: the ratio $\gamma = \max_{|\psi_0\rangle} \mathcal{F}_Q(\theta)/\mathcal{G}(\theta)$ as a function of time for $\omega = 1$ and different values of θ . From lower to upper curves, we have $\theta = 0, \pi/50, \pi/20, \pi/10, \pi/5, \pi/3, \pi/2$. Notice the loss of periodicity for higher values of θ .

which is larger than the QFI at any time. In particular, we may write

$$\mathcal{G}(\theta) = \left(\frac{\omega}{\Omega_\theta^2} + \sqrt{\max_{|\psi_0\rangle} \mathcal{F}_Q(\theta)} \right)^2. \quad (156)$$

The ratio $\gamma = \max_{|\psi_0\rangle} \mathcal{F}_Q(\theta)/\mathcal{G}(\theta)$ may be written as $\gamma = 1 - \omega/(\theta^2 t)$ for $\omega \ll 1$, whereas the difference between the difference between $\mathcal{F}_Q(\theta)$ and $\mathcal{G}(\theta)$ may be more pronounced in other regimes. For $\omega \gg 1$, the ratio γ oscillates at small times, and then it approaches unity for large times. In Fig. 7, we show the ratio γ as a function of time for $\omega = 1$ and different values of θ .

9.3. Estimation of a weak magnetic field by spin-1 probes

NV-center in diamond has been suggested as quantum probes to precisely estimate the magnitude of a weak magnetic field. The probing system is made of a nitrogen atom (N) inside a diamond crystal lattice, having a vacancy (V) in one of its neighboring sites. Two different classes of the defects are known and employed: the neutral state, usually referred to as NV_0 , and the negatively-charged state NV_- . The second class NV_- is the one exploited in metrological applications, since it provides a spin triplet state which can be accurately prepared, manipulated with long coherence time, and finally read out by purely optical means.³⁸ Upon assuming that the interactions with the surrounding nuclear spins may be neglected, the Hamiltonian H_{NV} governing the evolution of the triplet state is given by

$$H_{NV} = \mu \mathbf{B} \cdot \mathbf{S} + DS_z^2 + E(S_x^2 - S_y^2), \quad (157)$$

where the external magnetic field is denoted by \mathbf{B} and $\mathbf{S} = (S_x, S_y, S_z)$ is a vector whose elements are the three spin 1 matrices:

$$S_x = \sqrt{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_y = \sqrt{2}i \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_z = 2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In the above formulas μ is the Bohr magneton and the couplings D and E are given by $D \approx \pi \times 1.44$ GHz and $E \approx \pi \times 50$ kHz, respectively. Upon assuming that the magnetic field is weak, the two transverse components B_x and B_y may be neglected in comparison to the component B_z , which is aligned along the NV-center defect axis. By renaming B_z as θ , the Hamiltonian becomes

$$H_\theta = \mu\theta S_z + DS_z^2 + E(S_x^2 - S_y^2). \quad (158)$$

The maximum QFI is

$$\max_{|\psi_0\rangle} \mathcal{F}_Q(\theta) = \frac{8\mu^2[2\theta^2\mu^2t^2\chi^2 + E^2 - E^2 \cos(4\chi t)]}{\chi^4}, \quad (159)$$

where $\chi := \sqrt{\theta^2\mu^2 + 4E^2}$. Instead, $\mathcal{G}(\theta)$ is given by

$$\begin{aligned} \mathcal{G}(\theta) &= \left(\frac{2E\mu}{\chi^2} + 2\sqrt{2}\mu \frac{\sqrt{2\theta^2\mu^2t^2\chi^2 + E^2 - E^2 \cos(4\chi t)}}{\chi^2} \right)^2 \\ &= \left(\frac{2E\mu}{\chi^2} + \sqrt{\max_{|\psi_0\rangle} \mathcal{F}_Q(\theta)} \right)^2 \geq \max_{|\psi_0\rangle} \mathcal{F}_Q(\theta), \end{aligned} \quad (160)$$

with the maximised QFI approaching the value of $\mathcal{G}(\theta)$ only in the limit $\theta \gg 1$.

10. Conclusion

In this paper, we have addressed nonregular measurements as a novel resource for quantum metrology. In particular, we have analyzed the family of controlled energy measurements and applied them to Hamiltonian parameter estimation problems. A controlled energy measurement is obtained by applying a unitary control and then performing a projective energy measurement. It is nonregular whenever the Hamiltonian depends nonlinearly on the parameter θ .

We have then maximized the Fisher information over the set of controlled energy measurements and initial preparations. The maximum, denoted by $\mathcal{G}(\theta)$, can be computed by the closed-form expression given in Eq. (120), and it may be larger than the QFI of the corresponding regular statistical model. We have discussed how controlled energy measurements can be implemented in realistic scenarios, via an adaptation of the quantum phase estimation algorithm.

Finally, in order to clarify the details of our estimation techniques, we have worked out a collection of examples, showing that a precision enhancement, compared with regular measurements, is often possible. In particular, we have emphasized that, if the parameter is not a simple phase, the quantum Fisher information no longer necessarily embodies the ultimate precision limit. Our results show that precision of quantum metrological protocols is not necessarily bounded by the inverse of the quantum Fisher information, i.e. quantum enhanced estimation may be more precise than previously thought. We foresee further applications in the field of quantum sensing³⁹ and quantum probing.^{40–43}

Acknowledgments

The authors thanks Matteo Rossi, Francesco Albarelli, Marco Genoni, Claudia Benedetti, Stefano Olivares, Dario Tamascelli, C. M. Chandrasekar, Ilaria Pizio, Shivani Singh and Sholeh Razavian for interesting discussions. This work has been supported by SERB through project VJR/2017/000011. MGAP is a member of GNFM-INDAM.

Appendix A. Abbreviations and Symbols Used in this Paper

MSE	mean-square error
POVM	Positive operator-valued measure
FI	Fisher Information
SLD	Symmetric logarithmic derivative
QFI	Quantum Fisher Information
\mathbb{N}	Set of positive integers
\mathbb{N}_0	Set of nonnegative integers
\mathbb{R}	Set of real numbers
$\overline{\mathbb{R}}$	Extended set of real numbers
\mathbb{R}_+	Set of nonnegative real numbers
\mathbb{C}	Set of complex numbers
$ S $	Cardinality of a set
$\mathcal{P}(S)$	Power set of S
$\text{conv}(S)$	Convex hull of a set of points S
$\text{vert}(\Pi)$	Set of vertices of a convex polytope Π
M_{ij}	Element ij of M
M^t	Transpose of a matrix M
$\text{spec}(M)$	Spectrum of a matrix M
$\text{rk}(M)$	Rank of a matrix M
$\sigma(M)$	Spectral gap of a matrix M
$\text{col}(M)$	Set made up of the columns of a matrix M
$\text{diag}(\{\lambda_i\}_{i=1}^n)$	Diagonal matrix, with diagonal elements $\{\lambda_i\}_{i=1}^n$
$\text{im}_S(M)$	Image of a matrix M on a set S

$\mathcal{M}_{n,m}(\mathbb{K})$	Set of $n \times m$ matrices over a field \mathbb{K}
$\text{Her}_n(\mathbb{K})$	Set of $n \times n$ Hermitian matrices over a field \mathbb{K}
$\text{Her}_n^+(\mathbb{K})$	Set of $n \times n$ positive semi-definite Hermitian matrices over a field \mathbb{K}
\mathbb{I}_n	$n \times n$ identity matrix
\mathbb{O}_n	$n \times n$ zero matrix
$\mathbb{J}_{p \times q}$	$p \times q$ matrix made up of all ones
$\mathbb{X}, \mathbb{Y} \dots$	Classical random variables
$\text{E}(\mathbb{X})$	Expectation value of \mathbb{X}
$\text{Var}(\mathbb{X})$	Variance of \mathbb{X}
$\text{Cov}(\mathbb{X}, \mathbb{Y})$	Covariance of \mathbb{X} and \mathbb{Y}

References

1. J. F. Haase, A. Smirne, J. Kołodyński, R. Demkowicz-Dobrzański and S. F. Huelga, *Quantum Meas. Quantum Metrol.* **5** (2018) 13.
2. M. G. A. Paris, *Int. J. Quant. Inf.* **7** (2009) 125.
3. C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
4. A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd edn. (Edizioni della Normale, Pisa, 2011).
5. L. Seveso, M. A. C. Rossi and M. G. A. Paris, *Phys. Rev. A* **95** (2017) 012111.
6. L. Seveso and M. G. A. Paris, *Phys. Rev. A* **98** (2018) 032114.
7. S. Amari and H. Nagaoka, *Methods of Information Geometry*, Vol. 191 (American Mathematical Society, 2007).
8. L. L. Campbell, *Proc. Am. Math. Soc.* **98** (1986) 135.
9. N. Chentsov, *Ser. Stat.* **9** (1978) 267.
10. N. Ay, J. Jost, H. V. Le and L. Schwachhöfer, *Probab. Theory Relat. Fields* **162** (2015) 327.
11. C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, in *Breakthroughs in Statistics*, S. Kotz and N. L. Johnson (eds.) (Springer, Berlin, 1992), p. 235.
12. H. Cramèr, *Mathematical Methods of Statistics* (PMS-9) (Princeton University Press, Princeton, 2016).
13. G. Casella and R. L. Berger, *Statistical Inference*, Vol. 2 (Duxbury Pacific Grove, CA, 2002).
14. A. M. Gleason, *J. Math. Mech.* **6** (1957) 885.
15. K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).
16. M. Ozawa, *J. Math. Phys.* **25** (1984) 79.
17. H. Nagaoka, *IEICE Tech. Rep.* **IT 89-42** (1989) 9.
18. A. Fujiwara and H. Nagaoka, *Phys. Lett. A* **201** (1995) 119.
19. D. Petz, *Linear Algebra Appl.* **244** (1996) 81.
20. I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2017).
21. D. Petz and C. Sudàr, *J. Math. Phys.* **37** (1996) 2662.
22. L. Jing, J. Xiao-Xing, Z. Wei and W. Xiao-Guang, *Commun. Theor. Phys.* **61** (2014) 45.

23. S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72** (1994) 3439.
24. H. Nagaoka, An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators, in *Proc. Int. Symp. on Inform. Theory* **198** (1988) 577.
25. S. Ragy, M. Jarzyna and R. Demkowicz-Dobrzanski, *Phys. Rev. A* **94** (2016) 052108.
26. K. Matsumoto, *J. Phys. A: Math. Gen.* **35** (2002) 3111.
27. R. D. Gill and S. Massar, *Phys. Rev. A* **61** (2000) 042312.
28. H. Yuen and M. Lax, *IEEE Trans. Inf. Theory* **19** (1973) 740.
29. O. E. Barndorff-Nielsen, R. D. Gill and P. E. Jupp, *J. Royal Stat. Soc. Series B Stat. Methodol.* **65** (2003) 775.
30. T. Popoviciu, *Mathematica* **9** (1935) 129.
31. M. Akahira and K. Takeuchi, *Non-regular Statistical Estimation* (Springer, Berlin, 2012).
32. S. Pang and T. A. Brun, *Phys. Rev. A* **90** (2014) 022117.
33. S. Nakayama, A. Soeda and M. Muraio, *Phys. Rev. Lett.* **114** (2015) 190501.
34. Y. Matsuzaki, S. Nakayama, A. Soeda, S. Saito and M. Muraio, *Phys. Rev. A* **95** (2017) 062106.
35. K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin and F. Verstraete, *Nature* **471** (2011) 87.
36. A. Riera, C. Gogolin and J. Eisert, *Phys. Rev. Lett.* **108** (2012) 080402.
37. A. Y. Kitaev, *Electr. Coll. Comp. Compl.* **TR96** (1996) 3.
38. L. Rondin, J. Tetienne, T. Hingant, J. Roch, P. Maletinsky and V. Jacques, *Rep. Prog. Phys.* **77** (2014) 056503.
39. F. Piacentini, A. Avella, E. Rebufello, R. Lussana, F. Villa, A. Tosi, M. Gramegna, G. Brida, E. Cohen, L. Vaidman, I. P. Degiovanni and M. Genovese, *Nature Phys.* **13** (2017) 1191.
40. C. Benedetti, F. Buscemi, P. Bordone and M. G. A. Paris, *Phys. Rev. A* **89** (2014) 032114.
41. C. Benedetti and M. G. A. Paris, *Phys. Lett. A* **378** (2014) 2495.
42. M. A. C. Rossi and M. G. A. Paris, *Phys. Rev. A* **92** (2015) 010302.
43. D. Tamascelli, C. Benedetti, S. Olivares and M. G. A. Paris, *Phys. Rev. A* **94** (2016) 042129.