


Optimal Protocols for Quantum Metrology with Noisy Measurements

Sisi Zhou^{Ⓧ,*§}, Spyridon Michalakis^{Ⓧ,†} and Tuvia Gefen[‡]

Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA

 (Received 20 March 2023; revised 9 August 2023; accepted 8 September 2023; published 9 October 2023)

Measurement noise is a major source of noise in quantum metrology. Here, we explore preprocessing protocols that apply quantum controls to the quantum sensor state prior to the final noisy measurement (but after the unknown parameter has been imparted), aiming to maximize the estimation precision. We define the quantum preprocessing-optimized Fisher information, which determines the ultimate precision limit for quantum sensors under measurement noise, and conduct a thorough investigation into optimal preprocessing protocols. First, we formulate the preprocessing optimization problem as a biconvex optimization using the error observable formalism, based on which we prove that unitary controls are optimal for pure states and derive analytical solutions of the optimal controls in several practically relevant cases. Then we prove that for classically mixed states (whose eigenvalues encode the unknown parameter) under commuting-operator measurements, coarse-graining controls are optimal, while unitary controls are sub-optimal in certain cases. Finally, we demonstrate that in multiprobe systems where noisy measurements act independently on each probe, the noiseless precision limit can be asymptotically recovered using global controls for a wide range of quantum states and measurements. Applications to noisy Ramsey interferometry and thermometry are presented, as well as explicit circuit constructions of optimal controls.

DOI: [10.1103/PRXQuantum.4.040305](https://doi.org/10.1103/PRXQuantum.4.040305)

I. INTRODUCTION

Quantum metrology is one of the pillars of quantum science and technology [1–5]. This field deals with fundamental precision limits of parameter estimation imposed by quantum physics. Notably, it seeks to use nonclassical effects to enhance the estimation precision of unknown parameters in quantum systems, which has led to the development of improved sensing protocols in various experimental platforms [6–11]. To characterize the metrological limit of quantum sensors, the quantum Cramér-Rao bound (QCRB) [12,13], which is saturable for a large number of experiments, is conventionally used. It is defined using the quantum Fisher information (QFI) [14–16], which is one of the most useful and celebrated tools in quantum

metrology, with a considerable amount of research focused on developing better ways to calculate and bound it [17–22].

Although the QCRB and the QFI apply extensively in quantum sensing, they are defined assuming that arbitrary quantum measurements can be applied on quantum states to extract information about the unknown parameter. However, in actual experimental platforms, such as nitrogen-vacancy centers [23–29], superconducting qubits [30], trapped ions [31,32], and more, measurements are often noisy and time expensive, rendering the sensitivity of practical quantum devices far from the theoretical limits given by the QCRB. In particular, measurement noise remains a significant source of noise in quantum sensing experiments. Other sources of noise, such as system evolution and state preparation, have been studied extensively, with methods developed to mitigate their effect [17,33–45].

To tackle the effect of measurement noise on quantum metrology, interaction-based readouts were proposed [46–51] and demonstrated experimentally [52–54], where bespoke inter-particle interactions that enhance phase estimation precision in spin ensembles are applied before the noisy measurement step and after the probing step. The idea of employing unitary controls in a preprocessing manner, i.e., after the unknown parameter has been imparted but prior to the final measurement, was later formulated

*sisi.zhou26@gmail.com

†spiros@caltech.edu

‡tgefen@caltech.edu

§Present address: Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

as the imperfect (or noisy) QFI problem [50,55], where the preprocessing is optimized over all unitary operations. Classical postprocessing methods, such as measurement error mitigation [56–58], can then work in complement to the quantum preprocessing method for parameter estimation under noisy measurements.

Apart from a few specific cases, such as qubit sensors with lossy photon detection [55], setting the metrological limit under measurement noise by computing imperfect QFI has been difficult, limiting its practical application. In this work, we propose a more general measurement optimization scheme, where arbitrary quantum controls (i.e., general quantum channels that can be implemented utilizing unitary gates and ancillas) are applied before the noisy measurement. The goal is to identify the FI optimized over all quantum preprocessing channels for general quantum states and measurements, that we call the quantum preprocessing-optimized FI (QPFI) and quantifies the ultimate power of quantum sensors with measurement noise, and to obtain the corresponding optimal controls, that can be applied to achieve the optimal sensitivity in practical experiments.

We systematically study the QPFI, along with the corresponding optimal preprocessing controls in this work. In Sec. II, we first define the QPFI and review related concepts. We then introduce the concept of *error observables* in Sec. III, and use it to demonstrate that the QPFI problem can be cast as a biconvex optimization problem [59]. In turn, this allows us to find analytical conditions for optimality, and to identify optimal controls saturating the QPFI in the setting of commuting measurements applied to pure states (see Sec. IV). The case of classically mixed states (i.e., states for which the unknown parameter is encoded in the eigenvalues) is studied in Sec. V. Besides analytical solutions, we also manage to prove that unitary controls are optimal for pure states under general measurements, and that coarse-graining controls are optimal for classically mixed states under commuting-operator measurements, with a counterexample illustrating the nonoptimality of unitary controls. For general mixed states, we further prove useful bounds on the QPFI in Sec. VI. In terms of the asymptotic behavior of identical local measurements acting on multiprobe systems, in Sec. VII, we identify a sufficient condition for the convergence of the QPFI to the QFI using an optimal encoding protocol based on the Holevo-Schumacher-Westmoreland (HSW) theorem [60,61]. We show that the relevant condition is satisfied by a generic class of quantum states, including low-rank states, permutation-invariant states, and Gibbs states (with an unknown temperature), while previously only the pure-state case was proven [55].

Our results provide a theoretically accessible precision bound for quantum metrology under noisy measurements, along with a roadmap towards preprocessing optimization in sensing experiments.

II. DEFINITIONS

Given a quantum state ρ_θ as a function of an unknown parameter θ , the procedure to estimate θ goes as follows [see Fig. 1(a)]: (1) perform a quantum measurement $\{M_i\}$ on ρ_θ , which gives a measurement outcome i with probability $p_{i,\theta} = \text{Tr}(\rho_\theta M_i)$; (2) infer the value of θ using an estimator $\hat{\theta}$, which is a function of the measurement outcome i ; (3) repeat the above two steps multiple times and use the average of $\hat{\theta}$ over many trials as the final estimate of θ . Here, the quantum measurement $\{M_i\}$ is mathematically formulated as a positive operator-valued measure (POVM) [62] that satisfies $M_i \geq 0$ and $\sum_i M_i = \mathbb{1}$ (we use $A \geq 0$ to indicate an operator A that is positive semidefinite). We also assume in this work that ρ_θ and M_i lie in finite-dimensional Hilbert spaces, with measurement outcomes contained in a finite set.

In estimation theory, the Cramér–Rao bound (CRB) [63–65] provides a lower bound on the estimation error for any locally unbiased estimator $\hat{\theta}$ at a local point θ_0 where

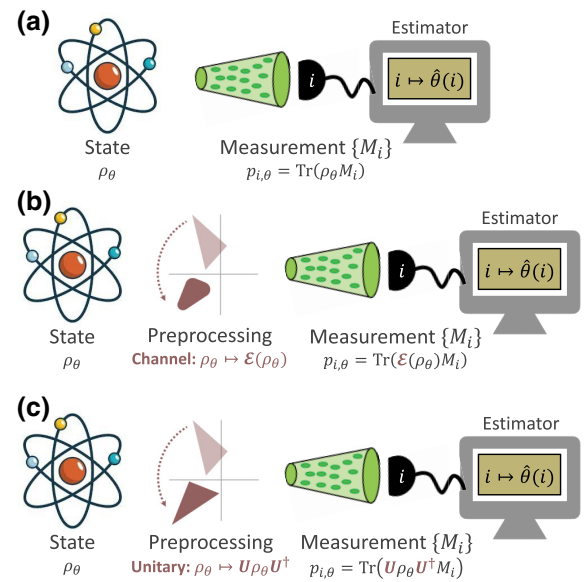


FIG. 1. (a) Standard parameter estimation procedure of a quantum state ρ_θ using a quantum measurement $\{M_i\}$. The estimation of θ is through an unbiased estimator $\hat{\theta}$ as a function of measurement outcomes i . The CRB states $\Delta \hat{\theta} \geq 1/\sqrt{N_{\text{expr}} F(\rho_\theta, \{M_i\})}$. (b) Preprocessing protocols where the measurement device is fixed, and the quantum control acting before the measurement is optimized over all quantum channels. The CRB states $\Delta \hat{\theta} \geq 1/\sqrt{N_{\text{expr}} F^{\text{IP}}(\rho_\theta, \{M_i\})}$. (c) Preprocessing protocols where the measurement device is fixed, and the quantum control acting before the measurement is optimized over all unitary channels. The CRB states $\Delta \hat{\theta} \geq 1/\sqrt{N_{\text{expr}} F^{\text{U}}(\rho_\theta, \{M_i\})}$. Different types of FIs discussed in this work satisfy $F(\rho_\theta, \{M_i\}) \leq F^{\text{U}}(\rho_\theta, \{M_i\}) \leq F^{\text{IP}}(\rho_\theta, \{M_i\}) \leq J(\rho_\theta)$ (and each inequality can be strict).

ρ_θ is differentiable, satisfying

$$\mathbb{E}[\hat{\theta}|\theta_0] = \theta_0, \quad \text{and} \quad \left. \frac{\partial}{\partial \theta} \mathbb{E}[\hat{\theta}|\theta] \right|_{\theta=\theta_0} = 1, \quad (1)$$

where we use $\mathbb{E}[\cdot|\theta]$ to denote the conditional expectation over the probability distribution $\{p_{i,\theta}\}$. The above condition indicates that locally unbiased estimators $\hat{\theta}$ provide an unbiased estimation of θ at the point θ_0 , which is also precise up to first order in its neighborhood. Note that in the following we will implicitly use $\mathbb{E}[\cdot]$ to represent $\mathbb{E}[\cdot|\theta]$ and consider locally unbiased estimators at a local point θ . The CRB states that the estimation error $\Delta\hat{\theta}$ (i.e., the standard deviation of the estimator $\hat{\theta}$) has the following lower bound:

$$\Delta\hat{\theta} := (\mathbb{E}[(\hat{\theta} - \theta)^2])^{\frac{1}{2}} \geq \frac{1}{\sqrt{N_{\text{expr}} F(\rho_\theta, \{M_i\})}}, \quad (2)$$

where N_{expr} is the number of experiments performed, and $F(\rho_\theta, \{M_i\})$ is the FI of the probability distribution $\{p_{i,\theta} = \text{Tr}(\rho_\theta M_i)\}$ [63–65], defined by

$$F(\rho_\theta, \{M_i\}) := \sum_{i: \text{Tr}(\rho_\theta M_i) \neq 0} \frac{(\text{Tr}(\partial_\theta \rho_\theta M_i))^2}{\text{Tr}(\rho_\theta M_i)}. \quad (3)$$

The CRB is often saturable asymptotically (i.e., when $N_{\text{expr}} \rightarrow \infty$) using the maximum-likelihood estimator [63–65] and therefore the FI, which is inversely proportional to the variance of the estimator, serves as a good measure of the degree of sensitivity of $\{p_{i,\theta}\}$ with respect to θ . One caveat is the CRB applies only to locally unbiased estimators and can be violated by biased estimators. Additionally, there exist singular cases where maximum-likelihood estimators are no longer necessarily asymptotically unbiased, e.g., when the support of $\{p_{i,\theta}\}$ varies in the neighborhood of θ , and the CRB may not apply to them [66]. However, for self-consistency, this paper will focus only on optimizing the FI, regardless of the limitations of the CRB.

The QFI of ρ_θ is the FI maximized over all possible quantum measurements on ρ_θ (see Appendix A for further details) and we will refer to the optimal measurements as QFI-attainable measurements. Formally, the QFI is defined by [12–14]

$$J(\rho_\theta) = \max_{\{M_i\}} F(\rho_\theta, \{M_i\}), \quad (4)$$

giving rise to the QCRB

$$\Delta\hat{\theta} \geq \frac{1}{\sqrt{N_{\text{expr}} J(\rho_\theta)}}, \quad (5)$$

which characterizes the ultimate lower bound on the estimation error. Going forward, we will also overload the

notation and write

$$J(\{p_{i,\theta}\}) := \sum_{i: p_{i,\theta} \neq 0} \frac{(\partial_\theta p_{i,\theta})^2}{p_{i,\theta}}, \quad (6)$$

to denote the FI of a classical probability distribution $\{p_{i,\theta}\}$, satisfying $p_{i,\theta} \geq 0$ and $\sum_i p_{i,\theta} = 1$. Note that, from now on, we will implicitly assume that the summation is taken over terms with nonzero denominators.

In practice, the optimal measurements achieving the QFI are not always implementable, restricting the range of applications of the QCRB. For example, the projective measurement onto the basis of the symmetric logarithmic operators, which is usually a correlated measurement among multiple probes, is known to be optimal [14], while quantum measurements in experiments are usually noisy and not exactly projective. Here, we consider a metrological protocol in which arbitrary quantum controls can be implemented, after the unknown parameter θ has been imparted to the quantum sensor state ρ_θ and before a fixed quantum measurement is performed [see Fig. 1(b)]. We call this additional step “preprocessing,” “premeasurement processing” in full. Note that the idea of implementing preprocessing quantum controls to improve sensitivity goes beyond the FI formalism and applies to other figures of merit of quantum sensors [67]. This model effectively describes quantum experiments where the measurement error is dominant, while the gate implementation error and the state preparation error is relatively small, a noise model that arises naturally in modern quantum devices such as nitrogen-vacancy centers [23–26] and superconducting qubits [30].

To quantify the sensitivity of estimating θ on ρ_θ with the measurement $\{M_i\}$ fixed, we define the FI optimized over all preprocessing quantum channels, or the *quantum preprocessing-optimized Fisher information (QPFI)*, to be

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = \sup_{\mathcal{E}} F(\mathcal{E}(\rho_\theta), \{M_i\}), \quad (7)$$

where \mathcal{E} is an arbitrary quantum channel (or a CPTP map [68]). See Appendix B for mathematical properties of the QPFI. In particular, when the quantum measurement is fixed, the CRB induced by the QPFI, i.e.,

$$\Delta\hat{\theta} \geq \frac{1}{\sqrt{N_{\text{expr}} F^{\mathbb{P}}(\rho_\theta, \{M_i\})}}, \quad (8)$$

provides a practical and tighter Cramér-Rao-type bound, compared to the QCRB, for parameter estimation under noisy measurements. We assume in the following discussions that all measurements are nontrivial (i.e., $\exists M_i \not\propto \mathbb{1}$, for all $\{M_i\}$) and $\partial_\theta \rho_\theta \neq 0$ so that the QPFI is always positive.

Unless stated otherwise, we will denote the systems that ρ_θ and $\{M_i\}$ act on by \mathcal{H}_S and $\mathcal{H}_{S'}$, respectively, and we

will refer to \mathcal{H}_S as the input system and $\mathcal{H}_{S'}$ as the output system. We do not assume $\mathcal{H}_S \cong \mathcal{H}_{S'}$ here. This broader context is of particular interest when the quantum state ρ_θ cannot be directly measured (e.g., readout of superconducting qubits via a resonator [30] and readout of nuclear spins via an electron spin in a nitrogen-vacancy center [69–72]); or when the quantum state is restricted to a subsystem of the entire system while quantum measurement can be performed globally.

Note that for generic noisy measurements, the supremum in Eq. (7) is usually attainable, i.e., there exists an optimal \mathcal{E} such that $F(\mathcal{E}(\rho_\theta), \{M_i\})$ is maximized (see Appendix C). However, there exist singular cases where $F(\mathcal{E}(\rho_\theta), \{M_i\})$ has no maximum, due to the singularity of the FI at the point $\text{Tr}(\mathcal{E}(\rho_\theta)M_i) = 0$ (see Sec. IV C for an example). In such cases, there still exist near-optimal quantum controls that attain $\sup_{\mathcal{E}} F(\mathcal{E}(\rho_\theta), \{M_i\}) - \eta$ for any small $\eta > 0$. In fact, we prove in Appendix C the following.

Theorem 1.—Let $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \mathbb{1}/d$, where $d = \dim(\mathcal{H}_{S'})$ and $0 < \epsilon < 1$. Then

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = \lim_{\epsilon \rightarrow 0^+} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}), \quad (9)$$

and the QPFI $F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\})$ is attainable for any $\epsilon \in (0, 1]$.

In the following, we will focus mostly on the case where the QPFI is attainable. We will discuss the behavior of the QPFI, exploring numerical optimization algorithms and analytical solutions to the optimal controls for certain practically relevant quantum states and measurements.

We will also examine the FI optimized over all unitary preprocessing channels, which we call the *quantum unitary-preprocessing-optimized Fisher information (QUPFI)* [50,55]

$$F^{\mathbb{U}}(\rho_\theta, \{M_i\}) = \sup_U F(U\rho_\theta U^\dagger, \{M_i\}), \quad (10)$$

where U is an arbitrary unitary gate. (Note that our QUPFI is the same as the imperfect QFI in Ref. [55].) Unlike the QPFI, we assume $\mathcal{H}_{S'} \cong \mathcal{H}_S$ (and do not distinguish between S' and S) when we talk about the QUPFI, so that it is well defined. We note here that Theorem 1 holds for the QUPFI, as well.

The optimal preprocessing controls that attain the QPFI and the QUPFI usually depend on θ , whose value should be roughly known before the experiment. Otherwise, one might use the two-step method by first using $\sqrt{N_{\text{expr}}}$ states to obtain a rough estimate $\tilde{\theta} \approx \theta$, and then performing the optimal controls based on $\tilde{\theta}$ on the remaining $N_{\text{expr}} - \sqrt{N_{\text{expr}}}$ states [73–75]. The two-step procedure introduces a negligible amount of error asymptotically.

Before we proceed, we prove a relation between the QPFI and the QUPFI that will be useful later.

Proposition 2.—Let \mathcal{H}_S and $\mathcal{H}_{S'}$ be the input and output systems of \mathcal{E} . Suppose \mathcal{H}_{A_1} and \mathcal{H}_{A_2} are ancillary systems such that $\mathcal{H}_{A_1} \otimes \mathcal{H}_S \cong \mathcal{H}_{A_2} \otimes \mathcal{H}_{S'}$. If $\dim(\mathcal{H}_{A_1}) \geq \dim(\mathcal{H}_{S'})^2$ [or equivalently, $\dim(\mathcal{H}_{A_2}) \geq \dim(\mathcal{H}_S) \dim(\mathcal{H}_{S'})$], then

$$\begin{aligned} & F^{\mathbb{P}}((\rho_\theta)_S, \{(M_i)_{S'}\}) \\ &= F^{\mathbb{U}}((\rho_\theta)_S \otimes |0_{A_1}\rangle \langle 0_{A_1}|, \{(M_i)_{S'} \otimes \mathbb{1}_{A_2}\}), \end{aligned} \quad (11)$$

where we use subscripts to denote the systems the operators are acting on.

Proof.—Any quantum channel $\mathcal{E}(\cdot) = \sum_{i=1}^{r_{\mathcal{E}}} K_i(\cdot)K_i^\dagger$ from \mathcal{H}_S to $\mathcal{H}_{S'}$ can be implemented by acting unitarily on \mathcal{H}_S and an ancillary system \mathcal{H}_{A_1} , and then tracing over an auxiliary system \mathcal{H}_{A_2} , if $\dim(\mathcal{H}_{A_2}) \geq r_{\mathcal{E}}$ (Stinespring’s dilation [68]). For any quantum channel with the input system \mathcal{H}_S and the output system $\mathcal{H}_{S'}$, there always exists a Kraus representation $\mathcal{E}(\cdot) = \sum_{i=1}^{r_{\mathcal{E}}} K_i(\cdot)K_i^\dagger$ such that $r_{\mathcal{E}} \leq \dim(\mathcal{H}_{S'}) \dim(\mathcal{H}_S)$ [68]. Therefore, if $\dim(\mathcal{H}_{A_2}) \geq \dim(\mathcal{H}_{S'}) \dim(\mathcal{H}_S)$, the unitary extension should exist.

Let $\mathcal{H}_S \otimes \mathcal{H}_{A_1} \cong \mathcal{H}_{S'} \otimes \mathcal{H}_{A_2}$ be the enlarged, isomorphic input and output Hilbert spaces, respectively. If $\dim(\mathcal{H}_{A_1}) \geq \dim(\mathcal{H}_{S'})^2$, then $\dim(\mathcal{H}_{A_2}) = \dim(\mathcal{H}_{A_1}) \dim(\mathcal{H}_S) / \dim(\mathcal{H}_{S'}) \geq \dim(\mathcal{H}_{S'}) \dim(\mathcal{H}_S)$. Thus, there is a unitary $U_{\mathcal{E}}$ mapping $\mathcal{H}_S \otimes \mathcal{H}_{A_1}$ to $\mathcal{H}_{S'} \otimes \mathcal{H}_{A_2}$ such that

$$\mathcal{E}(\sigma) = \text{Tr}_{A_2}(U_{\mathcal{E}}(\sigma \otimes |0\rangle \langle 0|)U_{\mathcal{E}}^\dagger). \quad (12)$$

From Eq. (3), it follows that

$$\begin{aligned} F(\mathcal{E}(\rho_\theta), \{M_i\}) &= F(\text{Tr}_{A_2}(U_{\mathcal{E}}(\rho_\theta \otimes |0\rangle \langle 0|)U_{\mathcal{E}}^\dagger), \{M_i\}) \\ &= F(U_{\mathcal{E}}(\rho_\theta \otimes |0\rangle \langle 0|)U_{\mathcal{E}}^\dagger, \{M_i \otimes \mathbb{1}\}), \end{aligned}$$

where we omit the subscripts for simplicity. Note that the Stinespring’s dilation technique is also useful in relating the QFI of a mixed state to the QFI of its purification in an extended Hilbert space [17,18]. Taking the supremum over \mathcal{E} in the above equality, we have

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq F^{\mathbb{U}}(\rho_\theta \otimes |0\rangle \langle 0|, \{M_i \otimes \mathbb{1}\}). \quad (13)$$

On the other hand, for any U from $\mathcal{H}_S \otimes \mathcal{H}_{A_1}$ to $\mathcal{H}_{S'} \otimes \mathcal{H}_{A_2}$, $\text{Tr}_{A_2}(U(\cdot) \otimes |0\rangle \langle 0|)U^\dagger$ is a quantum channel from \mathcal{H}_S to $\mathcal{H}_{S'}$, proving the other direction of Eq. (11). ■

III. ERROR OBSERVABLE FORMULATION

In this section, we will formalize the optimization of FI over quantum preprocessing controls as a biconvex optimization problem using the concepts of error observables. Using this new formulation, the preprocessing optimization problem becomes numerically tractable with standard algorithms for biconvex optimization [59]; and also analytically tractable for practically relevant quantum states (see Sec. IV).

Here, we consider the preprocessing optimization problem in Eq. (7). On the surface, it may appear from the definition of FI [Eq. (3)] that the target function $F(\mathcal{E}(\rho), \{M_i\})$ is mathematically formidable. To simplify the target function, we introduce the *error observable* X and the *squared error observable* X_2 , defined by

$$X = \sum_i x_i M_i, \quad \text{and} \quad X_2 = \sum_i x_i^2 M_i, \quad (14)$$

where x_i is interpreted as the difference between the estimator value $\hat{\theta}(i)$ and the true value θ , i.e. $x_i = \hat{\theta}(i) - \theta$. We assume there are r measurement outcomes and use \mathbf{x} to denote the vector (x_1, \dots, x_r) . The local unbiasedness conditions [Eq. (1)] for a single-shot measurement then become

$$\text{Tr}(\rho_\theta X) = 0, \quad \text{and} \quad \text{Tr}(\partial_\theta \rho_\theta X) = 1. \quad (15)$$

It can be verified mathematically (which is essentially a proof of the CRB) that the minimum of the variance of the estimator under the local unbiasedness conditions is the inverse of the FI; that is,

$$F(\rho_\theta, \{M_i\})^{-1} = \min_{\mathbf{x}} \text{Tr}(\rho_\theta X_2), \quad \text{such that Eq. (15)}. \quad (16)$$

The problem above is a convex optimization over variables \mathbf{x} , which can be solved using, e.g., the method of Lagrange multipliers [76]. The optimal solution to \mathbf{x} is

$$x_i = \frac{\frac{\text{Tr}(\partial_\theta \rho_\theta M_i)}{\text{Tr}(\rho_\theta M_i)}}{\sum_{j: \text{Tr}(\rho_\theta M_j) \neq 0} \frac{(\text{Tr}(\partial_\theta \rho_\theta M_j))^2}{\text{Tr}(\rho_\theta M_j)}}, \quad (17)$$

when $\text{Tr}(\rho_\theta M_i) \neq 0$, and $x_i = 0$ when $\text{Tr}(\rho_\theta M_i) = 0$. Note that the error observable formulation was previously used to derive the QCRB [77], where the QFI satisfies

$$J(\rho_\theta)^{-1} = \min_X \text{Tr}(\rho_\theta X^2), \quad \text{such that Eq. (15)}, \quad (18)$$

and X an arbitrary Hermitian matrix subject to the constraints in Eq. (15). This formulation has several useful applications [78–80]. In particular, an algorithm was proposed in Ref. [55] based on Eq. (18), to optimize the QFI of quantum channels.

Combining Eqs. (16) and (7), we have that

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta, \{M_i\})^{-1} &= \inf_{(\mathbf{x}, \mathcal{E})} \text{Tr}(\mathcal{E}(\rho_\theta) X_2), \\ &\text{such that } \text{Tr}(\mathcal{E}(\rho_\theta) X) = 0, \\ &\text{Tr}(\mathcal{E}(\partial_\theta \rho_\theta) X) = 1. \end{aligned} \quad (19)$$

Let \mathcal{H}_S and $\mathcal{H}_{S'}$ be the input and output systems of \mathcal{E} and let $\{|k\rangle_S\}_{k=1}^{\dim(\mathcal{H}_S)}$ and $\{|j\rangle_{S'}\}_{j=1}^{\dim(\mathcal{H}_{S'})}$ be two sets of orthonormal basis of \mathcal{H}_S and $\mathcal{H}_{S'}$, respectively. In the rest of

this section, we use matrix representations of operators in the above bases. It is convenient to represent a CPTP map \mathcal{E} using a linear operator acting on $\mathcal{H}_{S'} \otimes \mathcal{H}_S$. Let $\mathcal{E}(\cdot) = \sum_i K_i(\cdot)K_i^\dagger$ be the Kraus representation of \mathcal{E} . Then, the linear operator $\Omega = \sum_i |K_i\rangle\rangle\langle\langle K_i|$ is usually called the Choi matrix of \mathcal{E} [68], where $|\star\rangle\rangle := \sum_{jk} (\star)_{jk} |j\rangle_{S'} |k\rangle_S$ and $(\star)_{jk} = \langle j|_{S'} (\star) |k\rangle_S$. Ω corresponds to a CPTP map if and only if $\Omega \geq 0$ and $\text{Tr}_{S'}(\Omega) = \mathbb{1}_S$. \mathcal{E} acting on any density operator σ can be expressed using Ω through $\mathcal{E}(\sigma) = \text{Tr}_S((\mathbb{1} \otimes \sigma^T)\Omega)$ [we use $(\cdot)^T$ to denote matrix transpose]. Using the Choi matrix representation in Eq. (19), we have the following.

Theorem 3.—The optimal value of the following biconvex optimization problem gives the inverse of the QPFI.

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta, \{M_i\})^{-1} &= \inf_{(\mathbf{x}, \Omega)} \text{Tr}((X_2 \otimes \rho_\theta^T)\Omega), \\ &\text{such that } \Omega \geq 0, \quad \text{Tr}_{S'}(\Omega) = \mathbb{1}_S, \\ &\text{Tr}((X \otimes \rho_\theta^T)\Omega) = 0, \\ &\text{Tr}((X \otimes \partial_\theta \rho_\theta^T)\Omega) = 1. \end{aligned} \quad (20)$$

Equation (20) is a biconvex optimization problem of variables \mathbf{x} and Ω . Fixing Ω , Eq. (20) is a quadratic program with respect to \mathbf{x} , and fixing \mathbf{x} , Eq. (20) is a semidefinite program with respect to Ω ; each of which is efficiently solvable when the system dimensions are moderate and the domain of variables is compact.

Note that the domain of \mathbf{x} is unbounded in Eq. (20). In practice, one may impose a bounded domain on \mathbf{x} so that the minimum of Eq. (20) always exists. For cases where the QPFI is attainable, the optimal value of the bounded version will be equal to the one of Eq. (20) when the size of the bounded domain is sufficiently large. For singular cases where the QPFI is not attainable, the optimal value of the bounded version will approach the one of Eq. (20) with an arbitrarily small error as the size of the domain increases. We describe an algorithm called the global optimization algorithm [81] in Appendix D that can solve the bounded version of Eq. (20).

Finally, we note that Theorem 3 does not directly generalize to the case of QUPFI because the Choi matrices of unitary operators do not form a convex set. On the other hand, besides the set of quantum channels, our approach is also useful in optimizing the FI over other sets of quantum controls when the constraints on their Choi matrices can be represented using semidefinite constraints, e.g., the set of quantum channels that act only on a subsystem of the entire system.

IV. PURE STATES

In this section, we consider the special case where $\rho_\theta = \psi_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ is pure, which is most common in sensing experiments. We first consider the optimization of

the FI over the error vector \mathbf{x} and the unitary control U , and obtain two necessary conditions for the optimality of (\mathbf{x}, U) . We use these conditions to prove equality between the QPFI and the QUPFI for pure states, showing that unitary controls are optimal for such states (when $\mathcal{H}_S \cong \mathcal{H}_{S'}$). We also obtain an analytical expression of the QPFI for binary measurements (i.e., measurements with only two outcomes), and a semianalytical expression and analytical bounds for general commuting-operator measurements (i.e., measurements $\{M_i\}$ that satisfy $[M_i, M_j] = 0$ for all i, j). In particular, we prove that the optimal control is given by rotating the pure state and its derivative into a two-dimensional subspace spanned by two of the common eigenstates of the commuting-operator measurements.

A. Necessary conditions for optimal controls

Proposition 2 shows that the optimization for the QPFI can be reduced to an optimization for the QUPFI using the ancillary system. Thus, here we first focus on the following optimization problem over the unitary control:

$$F^{\text{U}}(\rho_\theta, \{M_i\})^{-1} = \inf_{(\mathbf{x}, U)} \text{Tr}(U\rho_\theta U^\dagger X_2), \quad (21)$$

$$\text{such that } \text{Tr}(U\rho_\theta U^\dagger X) = 0, \quad (22)$$

$$\text{Tr}(U\partial_\theta \rho_\theta U^\dagger X) = 1. \quad (23)$$

We obtain necessary conditions for the optimality of (\mathbf{x}, U) that will be useful later.

Lemma 4.—If (\mathbf{x}, U) is an optimal point for Eq. (21), it must satisfy

$$\frac{\text{Tr}(U\partial_\theta \rho_\theta U^\dagger X)}{\text{Tr}(U\rho_\theta U^\dagger X_2)} [X_2, U\rho_\theta U^\dagger] = 2[X, U\partial_\theta \rho_\theta U^\dagger]. \quad (24)$$

In particular, suppose $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ is pure. Let

$$|\psi_\theta^\perp\rangle := \frac{1}{\sqrt{\mathbf{n}}}(1 - |\psi_\theta\rangle\langle\psi_\theta|)|\partial_\theta \psi_\theta\rangle, \quad (25)$$

$$|\phi\rangle := U|\psi_\theta\rangle, \quad |\phi^\perp\rangle := U|\psi_\theta^\perp\rangle, \quad (26)$$

where the normalization factor $\mathbf{n} = \langle\partial_\theta \psi_\theta| (1 - |\psi_\theta\rangle\langle\psi_\theta|)|\partial_\theta \psi_\theta\rangle$. Then Eq. (24) is equivalent to the following two conditions:

- (1) $X|\phi\rangle = 1/(2\sqrt{\mathbf{n}})|\phi^\perp\rangle$.
- (2) $(\langle\phi|X_2|\phi\rangle X^2 - \langle\phi|X^2|\phi\rangle X_2)|\phi\rangle = 0$.

Proof.—Assume (\mathbf{x}, U) satisfies the constraints Eqs. (22) and (23). Then for any unitary operator V such that $\text{Tr}(U\partial_\theta \rho_\theta U^\dagger V^\dagger X V) \neq 0$,

$$\left(\frac{\mathbf{x} - \text{Tr}(U\rho_\theta U^\dagger V^\dagger X V)\mathbf{1}}{\text{Tr}(U\partial_\theta \rho_\theta U^\dagger V^\dagger X V)}, VU \right) \quad (27)$$

also satisfies the constraints Eqs. (22) and (23), where $\mathbf{1}$ is a r -dimensional vector of which each element is 1. We call

the transformation above a “ V transformation” on (\mathbf{x}, U) . After a V transformation, the target function becomes

$$\frac{\text{Tr}(U\rho_\theta U^\dagger V^\dagger X_2 V) - \text{Tr}(U\rho_\theta U^\dagger V^\dagger X V)^2}{\text{Tr}(U\partial_\theta \rho_\theta U^\dagger V^\dagger X V)^2}, \quad (28)$$

which shall be no smaller than $\text{Tr}(U\rho_\theta U^\dagger X_2)$ when (\mathbf{x}, U) is optimal. Let $V = e^{-idG}$ where dG is an arbitrary infinitesimally small Hermitian matrix. The first-order derivative of Eq. (28) with respect to dG must be zero, which then implies Eq. (24). Specifically, to simplify the notation, let $\tilde{\rho} := U\rho_\theta U^\dagger$ and $\dot{\tilde{\rho}} := U\partial_\theta \rho_\theta U^\dagger$. Then the difference between the target function after and before the V transformation must be zero up to the first order of dG , i.e.,

$$\frac{\text{Tr}(\tilde{\rho}X_2) + \text{Tr}(\tilde{\rho}(-i[dG, X_2]))}{(\text{Tr}(\dot{\tilde{\rho}}X) + \text{Tr}(\dot{\tilde{\rho}}(-i[dG, X])))^2} = \frac{\text{Tr}(\tilde{\rho}X_2)}{(\text{Tr}(\dot{\tilde{\rho}}X))^2}, \quad (29)$$

$$\begin{aligned} &\Rightarrow \frac{\text{Tr}(\dot{\tilde{\rho}}X)^2}{\text{Tr}(\tilde{\rho}X_2)} - \frac{2i\text{Tr}(\dot{\tilde{\rho}}X)\text{Tr}(\dot{\tilde{\rho}}[dG, X])}{\text{Tr}(\tilde{\rho}X_2)} \\ &\quad + \frac{i\text{Tr}(\dot{\tilde{\rho}}X)^2\text{Tr}(\dot{\tilde{\rho}}[dG, X_2])}{\text{Tr}(\tilde{\rho}X_2)^2} = \frac{\text{Tr}(\dot{\tilde{\rho}}X)^2}{\text{Tr}(\tilde{\rho}X_2)}, \end{aligned} \quad (30)$$

$$\begin{aligned} &\Rightarrow -2\text{Tr}(dG[X, \dot{\tilde{\rho}}]) + \frac{\text{Tr}(\dot{\tilde{\rho}}X)\text{Tr}(dG[X_2, \tilde{\rho}])}{\text{Tr}(\tilde{\rho}X_2)} = 0, \end{aligned} \quad (31)$$

$$\Rightarrow \frac{\text{Tr}(\dot{\tilde{\rho}}X)[X_2, \tilde{\rho}]}{\text{Tr}(\tilde{\rho}X_2)} = 2[X, \dot{\tilde{\rho}}], \quad (32)$$

where in the first step we take the inverse of both sides and ignore higher-order terms, in the second step we multiply both sides by $-i\text{Tr}(\rho X_2)$, and in the last step we use the fact that if an operator A satisfies $\text{Tr}(dGA) = 0$ for any Hermitian dG , then $A = 0$.

For pure states, Eq. (24) can be further simplified. Using the definitions of $|\phi\rangle$ and $|\phi^\perp\rangle$, we have $U\rho_\theta U^\dagger = |\phi\rangle\langle\phi|$ and

$$\begin{aligned} U\partial_\theta \rho_\theta U^\dagger &= U|\partial_\theta \psi_\theta\rangle\langle\psi_\theta|U^\dagger + \text{h.c.} \\ &= (|\partial_\theta \psi_\theta\rangle - |\psi_\theta\rangle\langle\psi_\theta|\partial_\theta \psi_\theta\rangle\langle\psi_\theta|) + \text{h.c.} \\ &= \sqrt{\mathbf{n}}(|\phi^\perp\rangle\langle\phi| + |\phi\rangle\langle\phi^\perp|), \end{aligned} \quad (33)$$

where h.c. stands for the Hermitian conjugate and we use $\partial_\theta \langle\psi_\theta|\psi_\theta\rangle = \langle\partial_\theta \psi_\theta|\psi_\theta\rangle + \langle\psi_\theta|\partial_\theta \psi_\theta\rangle = 0$ in the second step. Equation (24) becomes

$$|u\rangle\langle\phi| - |\phi\rangle\langle u| + |v\rangle\langle\phi^\perp| - |\phi^\perp\rangle\langle v| = 0, \quad (34)$$

where $|u\rangle = X|\phi^\perp\rangle - (\text{Re}[\langle\phi|X|\phi^\perp\rangle]/\langle\phi|X_2|\phi\rangle)X_2|\phi\rangle$ and $|v\rangle = X|\phi\rangle$. Equation (34) is equivalent to $|u\rangle, |v\rangle \in \text{span}\{|\phi\rangle, |\phi^\perp\rangle\}$, $\langle\phi|u\rangle, \langle\phi^\perp|v\rangle \in \mathbb{R}$, and $\langle\phi|v\rangle = \langle u|\phi^\perp\rangle$. Combining these conditions with the local unbiasedness

constraints $\langle \phi | X | \phi \rangle = 0$ and $2\sqrt{n}\text{Re}[\langle \phi | X | \phi^\perp \rangle] = 1$, the two conditions in Lemma 4 are then proven. Specifically, we first use $\langle \phi | v \rangle = \langle \phi | X | \phi \rangle = 0$ and $|v\rangle \in \text{span}\{|\phi\rangle, |\phi^\perp\rangle\}$ to derive that $X|\phi\rangle \propto |\phi^\perp\rangle$. Then using $\langle \phi^\perp | v \rangle \in \mathbb{R}$ and $2\sqrt{n}\text{Re}[\langle \phi | X | \phi^\perp \rangle] = 1$, we derive Condition (1). To derive Condition (2), we first use $\langle u | \phi^\perp \rangle = \langle \phi | v \rangle = 0$ to derive that $|u\rangle \propto |\phi\rangle$ and then $2\sqrt{n}\text{Re}[\langle \phi | X | \phi^\perp \rangle] = 1$ and Condition (2) to derive that $\langle \phi | u \rangle = \langle \phi | X | \phi^\perp \rangle - \text{Re}[\langle \phi | X | \phi^\perp \rangle] / \langle \phi | X_2 | \phi \rangle \langle \phi | X_2 | \phi \rangle = \frac{1}{2\sqrt{n}} - \frac{1}{2\sqrt{n}} = 0$. Then we have $|u\rangle = 0$, combining $|u\rangle \propto |\phi\rangle$ and $\langle \phi | u \rangle = 0$. Note that $|u\rangle = 0$ is equivalent to Condition (2) after multiplying both sides by $\langle \phi | X_2 | \phi \rangle / 2\sqrt{n}$. Finally, we note that from Condition (1) and Condition (2), the necessary condition in Eq. (24) can be recovered straightforwardly, proving the equivalence between Eq. (24) and Conditions (1) and (2) for pure states. ■

As a sanity check, consider the special case where $\{M_i = |i\rangle\langle i|\}_{i=1}^{\dim(\mathcal{H}_{S'})}$ is a projection onto an orthonormal basis of $\mathcal{H}_{S'}$. Then we have $X_2 = X^2$, so Condition (2) is trivially satisfied. Furthermore, choose (\mathbf{x}, U) such that the error observable $X = 1/2\sqrt{n}(|\phi^\perp\rangle\langle\phi| + |\phi\rangle\langle\phi^\perp|)$, so that Condition (1) is satisfied. Moreover, the variance of the estimation is

$$\langle \phi | X_2 | \phi \rangle = \langle \phi | X^2 | \phi \rangle = \frac{1}{4n} = J(\rho_\theta)^{-1}, \quad (35)$$

implying that the QFI is achievable using the above projective measurement, since $J(\rho_\theta) = 4n$ for pure states [14,82]. For general quantum measurements, the QUPFI might be strictly smaller than the QFI, in which case for the optimal choice of (\mathbf{x}, U) ,

$$\frac{1}{F^\mathbb{U}(\rho_\theta, \{M_i\})} = \langle \phi | X_2 | \phi \rangle > \langle \phi | X^2 | \phi \rangle = \frac{1}{J(\rho_\theta)}. \quad (36)$$

It is interesting to note that $X_2 \geq X^2$, for general POVM measurements. This follows directly from writing

$$X_2 - X^2 = \sum_i (x_i - X) M_i (x_i - X), \quad (37)$$

and noting that each term in the above sum is positive semidefinite.

B. Unitary controls are optimal

Using the definitions of $|\phi\rangle$ and $|\phi^\perp\rangle$ in Eq. (26), we observe that Eq. (21) can be rewritten as

$$\begin{aligned} F^\mathbb{U}(\psi_\theta, \{M_i\})^{-1} &= \inf_{(\mathbf{x}, |\phi\rangle, |\phi^\perp\rangle)} \langle \phi | X_2 | \phi \rangle, \\ \text{such that } \langle \phi | \phi^\perp \rangle &= 0, \langle \phi | X | \phi \rangle = 0, \\ \text{Re}[\langle \phi | X | \phi^\perp \rangle] &= 1/(2\sqrt{n}), \end{aligned} \quad (38)$$

where $\psi_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ is pure. Here $|\phi\rangle$ and $|\phi^\perp\rangle$ are two arbitrary normal vectors that are orthogonal. From Eq. (38), changing $|\phi\rangle$ to $|\phi\rangle/(2\sqrt{n})$ makes it clear that $F^\mathbb{U}(\rho_\theta, \{M_i\})$ can be written as the product of

$$J(\psi_\theta) = 4n \quad (39)$$

and a state-independent constant. We have

$$F^\mathbb{U}(\psi_\theta, \{M_i\}) = \gamma(\{M_i\})J(\psi_\theta), \quad (40)$$

where

$$\begin{aligned} \gamma(\{M_i\})^{-1} &= \inf_{(\mathbf{x}, |\phi\rangle, |\phi^\perp\rangle)} \langle \phi | X_2 | \phi \rangle, \\ \text{such that } \langle \phi | \phi^\perp \rangle &= 0, \langle \phi | X | \phi \rangle = 0, \\ \text{Re}[\langle \phi | X | \phi^\perp \rangle] &= 1. \end{aligned} \quad (41)$$

Or more explicitly,

$$\gamma(\{M_i\}) = \sup_{|\phi\rangle, |\phi^\perp\rangle} \sum_i \text{Re}[\langle \phi | M_i | \phi^\perp \rangle]^2 / \langle \phi | M_i | \phi \rangle. \quad (42)$$

[Note that going from Eqs. (41) and (42), we need only to optimize the target function over \mathbf{x} with a fixed $(|\phi\rangle, |\phi^\perp\rangle)$ and use standard methods for quadratic programming, e.g., Lagrange multipliers [76].] Note that Eqs. (40) and (42) were also proven using a different method in Ref. [55]. $\gamma(\{M_i\})$ is the *normalized QUPFI* for any pure states with unit QFIs and it is a function of $\{M_i\}$ that lies in $[0, 1]$, which is the ratio between the QUPFI and the QFI for any pure states. It is independent of the exact ψ_θ and can fully characterize the power of quantum measurements in terms of estimation on pure states.

Note that Eq. (40) decomposes the QUPFI into the product of the QFI, as a function of states, and the normalized QUPFI, as a function of measurements. This result is useful when experimentalists have control over input states in sensing processes. It implies when a pure input state ψ_0 undergoes unitary evolution U_θ , the optimal choices of the input state that maximizes the output FI are identical in situations with or without measurement noise.

Using Condition (1) in Lemma 4, we now prove that unitary controls are always optimal, that is, the QPFI is equal to the QUPFI when $\mathcal{H}_S \cong \mathcal{H}_{S'}$. We have the following theorem.

Theorem 5.—Consider a pure state ψ_θ and a quantum measurement $\{M_i\}$ acting on the same system. Unitary preprocessing controls are always optimal among quantum preprocessing controls for optimizing the FI, i.e.,

$$F^\mathbb{P}(\psi_\theta, \{M_i\}) = F^\mathbb{U}(\psi_\theta, \{M_i\}). \quad (43)$$

Or equivalently,

$$\gamma(\{M_i\}) = \gamma(\{M_i \otimes \mathbb{1}_A\}), \quad (44)$$

where A is an ancillary system of an arbitrary size.

Proof.—We first consider the situation where the QUPFI is attainable, that is, there always exists an (\mathbf{x}, U) such that the infimum in Eq. (21) is attainable. Using Condition (1) in Lemma 4, we can rewrite Eq. (38) as

$$\begin{aligned} F^\cup(\psi_\theta, \{M_i\})^{-1} &= \min_{(\mathbf{x}, |\phi\rangle)} \langle \phi | X_2 | \phi \rangle, \\ \text{such that } \langle \phi | X | \phi \rangle &= 0, \\ \langle \phi | X^2 | \phi \rangle &= 1/(4n), \end{aligned} \quad (45)$$

where $n = J(\psi_\theta)/4$. Let $\dim(\mathcal{H}_A) \geq \dim(\mathcal{H}_{S'})^2 = d^2$. $J(\psi_\theta) = J(\psi_\theta \otimes |0_A\rangle \langle 0_A|)$ and Proposition 2 imply

$$\begin{aligned} F^\mathbb{P}(\psi_\theta, \{M_i\})^{-1} &= F^\cup(\psi_\theta \otimes |0_A\rangle \langle 0_A|, \{M_i \otimes \mathbb{1}_A\})^{-1} \\ &= \min_{(\mathbf{x}, |\phi\rangle)} \langle \phi | X_2 \otimes \mathbb{1}_A | \phi \rangle, \\ \text{such that } \langle \phi | X \otimes \mathbb{1}_A | \phi \rangle &= 0, \\ \langle \phi | X^2 \otimes \mathbb{1}_A | \phi \rangle &= 1/(4n). \end{aligned} \quad (46)$$

It is equivalent to the optimization problem

$$\begin{aligned} \min_{(\mathbf{x}, \sigma)} \text{Tr}(\sigma X_2), \\ \text{such that } \text{Tr}(\sigma X) = 0, \text{Tr}(\sigma X^2) = 1/(4n), \end{aligned} \quad (47)$$

where σ is an arbitrary density operator and corresponds to $\text{Tr}_A(|\phi\rangle \langle \phi|)$. We will show below that for any σ^* that is optimal for Eq. (47), there exists an optimal pure state solution $|\phi^{**}\rangle \langle \phi^{**}|$ for Eq. (47). Then the optimal values of Eqs. (45) and (47) must be the same, proving Eq. (43).

Assume (\mathbf{x}^*, σ^*) is optimal for Eq. (47). Without loss of generality, we assume $\text{supp}(\sigma^*) \subseteq \text{supp}((X^*)^2)$, because otherwise σ^* projected onto the support of $(X^*)^2$ is another optimal solution because the constraints in Eq. (47) are invariant and the target function is no larger after the projection. We now show there exists another optimal solution $(\mathbf{x}^{**}, |\phi^{**}\rangle \langle \phi^{**}|)$. First, note that $X^* = \sum_i x_i^* M_i$ and $X_2^* = \sum_i (x_i^*)^2 M_i$ satisfy $\text{Tr}(\sigma^* X^*) = 0$ and $\text{Tr}(\sigma^* (X^*)^2) = 1/(4n)$ from Eq. (47), and

$$\gamma(\{M_i \otimes \mathbb{1}_A\}) X_2^* \Pi = (X^*)^2 \Pi, \quad (48)$$

where Π is the projection onto the support of σ^* . Note that Eq. (48) is true because

- (i) $\langle \phi | (X^*)^2 | \phi \rangle / \langle \phi | X_2^* | \phi \rangle = F^\cup(\psi_\theta, \{M_i \otimes \mathbb{1}_A\}) / 4n = \gamma(\{M_i \otimes \mathbb{1}_A\})$, from Condition (1) in Lemma 4 and
- (ii) $\langle \phi | X_2^* | \phi \rangle (X^*)^2 \Pi = \langle \phi | (X^*)^2 | \phi \rangle X_2^* \Pi = 0$ from Condition (2) in Lemma 4.

Let $\sigma^* = \sum_{k=1}^d \mu_k |k\rangle \langle k|$, where $\{|i\rangle\}_{i=1}^{d'}$ is orthonormal in \mathcal{H}_S . We claim that we can always choose

$$|\phi^{**}\rangle = \sum_{k=1}^d e^{i\varphi_k} \sqrt{\mu_k} |k\rangle, \quad (49)$$

such that $\langle \phi^{**} | X^* | \phi^{**} \rangle = 0$, by picking a suitable $\{\varphi_k\}_{k=1}^d$. To see this, observe that

$$\langle \phi^{**} | X^* | \phi^{**} \rangle = \sum_{k \neq k'} e^{i(\varphi_k - \varphi_{k'})} \sqrt{\mu_k \mu_{k'}} \langle k | X^* | k' \rangle, \quad (50)$$

is a real, continuous function $f(\varphi_1, \dots, \varphi_d)$ of $\{\varphi_k\}_{k=1}^d \in \mathbb{R}^d$, where we omitted the sum over $k = k'$ terms because $\text{Tr}(\sigma^* X^*) = 0$ implies that it vanishes. Note that for any fixed $\{\varphi_k\}_{k=1}^d$, the sum of all 2^d terms $f(\varphi_1 \pm \pi/2, \dots, \varphi_d \pm \pi/2)$ is zero, implying that one, or more, of these terms is zero, or that some are negative and others are positive. In the latter case, the continuity of $f(\varphi_1, \dots, \varphi_d)$ implies that its image must include zero. Therefore, we can pick a $\{\varphi_k\}_{k=1}^d$ such that $f(\varphi_1, \dots, \varphi_d) = 0$, based on which the $|\phi^{**}\rangle$ defined by Eq. (49) satisfies $\langle \phi^{**} | X^* | \phi^{**} \rangle = 0$. Furthermore, we choose

$$\mathbf{x}^{**} = \sqrt{\frac{1}{4n \langle \phi^{**} | (X^*)^2 | \phi^{**} \rangle}} \mathbf{x}^*, \quad (51)$$

so that $\langle \phi^{**} | (X^*)^2 | \phi^{**} \rangle = 1/(4n)$. Note that $\langle \phi^{**} | (X^*)^2 | \phi^{**} \rangle$ is always positive and thus the above denominator is positive because we assumed $(X^*)^2$ is positive definite on $\text{supp}(\sigma^*)$. We have now proved that $(\mathbf{x}^{**}, |\phi^{**}\rangle \langle \phi^{**}|)$ satisfies the constraints in Eq. (47). Moreover, noting that $\gamma(\{M_i \otimes \mathbb{1}_A\}) X_2^{**} \Pi = (X^{**})^2 \Pi$, the value of the target function $\langle \phi^{**} | X_2^{**} | \phi^{**} \rangle = \langle \phi^{**} | (X^{**})^2 | \phi^{**} \rangle / \gamma(\{M_i \otimes \mathbb{1}_A\}) = 1/4n \gamma(\{M_i \otimes \mathbb{1}_A\}) = F^\mathbb{P}(\psi_\theta, \{M_i\})^{-1}$ is also optimal. Therefore, $(\mathbf{x}^{**}, |\phi^{**}\rangle \langle \phi^{**}|)$ is an optimal solution for both Eqs. (45) and (47), proving Eq. (43).

When the QPFI of Eq. (21) is not attainable, we take $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \mathbb{1}/d$ and using Theorem 1, we have

$$\begin{aligned} F^\mathbb{P}(\psi_\theta, \{M_i\}) &= \lim_{\epsilon \rightarrow 0^+} F^\mathbb{P}(\psi_\theta, \{M_i^{(\epsilon)}\}) \\ &= \lim_{\epsilon \rightarrow 0^+} F^\cup(\psi_\theta, \{M_i^{(\epsilon)}\}) = F^\cup(\psi_\theta, \{M_i\}), \end{aligned} \quad (52)$$

where in the second step we use the equality between the QPFI and the QUPFI in the case where the QUPFI is attainable.

So far, we have proven that Eq. (44) is true when $\dim(\mathcal{H}_A) \geq \dim(\mathcal{H}_{S'})^2$, due to Proposition 2 and the equality between the QPFI and the QUPFI. It also holds

for any $\mathcal{H}_{A'}$ such that $\dim(\mathcal{H}_{A'}) \leq \dim(\mathcal{H}_S)^2$ because we have $\gamma(\{M_i \otimes \mathbb{1}_{A'}\}) \geq \gamma(\{M_i \otimes \mathbb{1}_{A'}\}) \geq \gamma(\{M_i\})$ by definition. ■

C. Analytical solution for binary measurements

Here we provide an analytical solution to the QPFI and the corresponding optimal preprocessing control using Proposition 2 for binary measurements where $r = 2$.

1. Measurement on a qubit

We first consider the simplest case where the measurement is on a single qubit. Let $X = x_1 M_1 + x_2 M_2$ where $M_1 = M$ and $M_2 = \mathbb{1} - M$. Without loss of generality, we assume

$$M = m_1 |1\rangle\langle 1| + m_2 |2\rangle\langle 2|, \quad (53)$$

for some $m_1, m_2 \in [0, 1]$, where $\{|1\rangle, |2\rangle\}$ is an orthonormal basis. Moreover, we assume $m_1 > m_2$ and $1 - m_1 \geq m_2$. [When $m_1 = m_2$, we must have $\gamma(\{M_i\}) = 0$ because the measurement outcome does not depend on θ .] Here m_2 and $1 - m_1$ can be interpreted as the error probabilities that state $|2\rangle$ is mistaken for $|1\rangle$, and state $|1\rangle$ is mistaken for $|2\rangle$, respectively.

Consider first the case where $1 > m_1 > m_2 > 0$, that is, the error probabilities are both nonzero. We show in Appendix E 1 that all solutions that satisfy the two necessary conditions in Lemma 4 give the same optimal FI. One optimal solution to the preprocessed state is

$$|\phi^*\rangle = \sqrt{p^*} |1\rangle + \sqrt{1 - p^*} |2\rangle, \quad (54)$$

$$|\phi^{\perp*}\rangle = \sqrt{1 - p^*} |1\rangle - \sqrt{p^*} |2\rangle, \quad (55)$$

where

$$p^* = \frac{\sqrt{m_2(1 - m_2)}}{\sqrt{m_1(1 - m_1)} + \sqrt{m_2(1 - m_2)}}. \quad (56)$$

Here the optimal unitary control U^* can be chosen as any unitary such that Eq. (26) is true for Eqs. (54) and (55). [In the following, we will only use $(|\phi^*\rangle, |\phi^{\perp*}\rangle)$ to represent the optimal preprocessing unitary with the implicit assumption that U^* can be chosen as any unitary rotating $(|\psi_\theta\rangle, |\psi_\theta^\perp\rangle)$ to $(|\phi^*\rangle, |\phi^{\perp*}\rangle)$.] Note that the symmetry transformations $|\phi^{\perp*}\rangle \mapsto -|\phi^{\perp*}\rangle$, $|1\rangle \mapsto e^{i\omega} |1\rangle$ and $|2\rangle \mapsto e^{i\omega'} |2\rangle$ for any $\omega, \omega' \in \mathbb{R}$ will generate alternative optimal solutions, and they all provide the same optimal normalized FI:

$$\gamma(\{M_i\}) = 1 - (\sqrt{m_1 m_2} + \sqrt{(1 - m_1)(1 - m_2)})^2. \quad (57)$$

Note that this result was obtained also in Ref. [55] using a different method based on the Bloch-sphere representation. Here $\sqrt{1 - \gamma(\{M_i\})}$ is exactly equal to the fidelity

between two binary probability distributions $(m_1, 1 - m_1)$ and $(m_2, 1 - m_2)$.

Take the symmetric binary measurement as an example, where $m_1 = 1 - m$, $m_2 = m$, and $m < 1/2$, and m represents the probability of a bit-flip error in the measurement. Then we have $p^* = 1/2$ (as expected from the bit-flip symmetry), and $\gamma(\{M_i\}) = 1 - 4m(1 - m)$, which is equal to 1 in the noiseless case, and drops to 0 when $m \rightarrow 1/2$.

In the case of perfect projective measurements where $1 = m_1 > m_2 = 0$, we show in Appendix E 1 that the QPFI is equal to the QFI and is attainable for any $0 < p^* < 1$. The case where $1 > m_1 > m_2 = 0$ is singular, in the sense that the QPFI is no longer attainable but only approachable. It corresponds to the situation where one type of error ($|2\rangle$ mistaken for $|1\rangle$) is zero, while the other ($|1\rangle$ mistaken for $|2\rangle$) is nonzero. In this case, we have $\gamma(\{M_i\}) = m_1$ using Eq. (57) and Theorem 1.

2. Measurement on a qudit

Next, we consider the general case where the measurement is on a qudit and we assume $\dim(\mathcal{H}_S) = d \geq 2$. Without loss of generality, we assume

$$M = \sum_{j=1}^d m_j |j\rangle\langle j|, \quad (58)$$

where $\{|j\rangle\}_{j=1}^d$ is an orthonormal basis of \mathcal{H}_S . We also assume $m_i \geq m_j$ for all $i \leq j$ without loss of generality. Here we assume $1 > m_1 > m_d > 0$, which guarantees the attainability of the QPFI (see Lemma 15 in Appendix C) and the nontriviality of quantum measurements. (The singular cases where $m_1 = 1$ or $m_d = 0$ can be derived using Theorem 1.) We show in Appendix E 2 that the optimal solution to $|\phi\rangle$ is supported on basis states corresponding to at most two different values of m_i and the problem is simplified to selecting the optimal basis states and applying the qubit-case results. We show that

$$|\phi^*\rangle = \sqrt{p^*} |1\rangle + \sqrt{1 - p^*} |d\rangle, \quad (59)$$

$$|\phi^{\perp*}\rangle = \sqrt{1 - p^*} |1\rangle - \sqrt{p^*} |d\rangle, \quad (60)$$

is an optimal solution, where

$$p^* = \frac{\sqrt{m_d(1 - m_d)}}{\sqrt{m_1(1 - m_1)} + \sqrt{m_d(1 - m_d)}}. \quad (61)$$

The normalized QPFI is given by

$$\gamma(\{M_i\}) = 1 - (\sqrt{m_1 m_d} + \sqrt{(1 - m_1)(1 - m_d)})^2. \quad (62)$$

Viewing $\{(m_i, 1 - m_i)\}_{i=1}^d$ as d binary probability distributions, the optimal strategy is always to select the two probability distributions that have the minimum fidelity (i.e., the largest distance) between each other.

D. Semianalytical solution and analytical bounds for commuting-operator measurements

Here we consider *commuting-operator measurements*, where all measurement operators commute, which is among the most common types of measurements in quantum sensing experiments, e.g., projective measurements affected by detection errors.

Assume $\dim(\mathcal{H}_S) = d \geq 2$. Without loss of generality, we assume

$$M_i = \sum_{j=1}^d m_j^{(i)} |j\rangle \langle j|, \quad (63)$$

where $\{|j\rangle\}_{j=1}^d$ is an orthonormal basis of \mathcal{H}_S and $\sum_{i=1}^r m_j^{(i)} = 1$ for all j . Again, we assume $m_j^{(i)} > 0$ for all i, j to exclude the singular cases where the QPFI is not attainable.

In order to find the optimal control, we first prove the following theorem, which states that the optimal $|\phi\rangle$ can be restricted to a two-dimensional subspace spanned by two basis states, i.e., the optimal unitary controls rotate the pure state and its derivative to a subspace spanned by two of the eigenstates of the commuting-operator measurement.

Theorem 6.—For commuting-operator measurements [Eq. (63)], there always exists an optimal solution to $(|\phi\rangle, |\phi^\perp\rangle)$ such that $|\phi\rangle = \sqrt{p} |k\rangle + \sqrt{1-p} |l\rangle$ and $|\phi^\perp\rangle = \sqrt{1-p} |k\rangle - \sqrt{p} |l\rangle$ for two basis states $|k\rangle$ and $|l\rangle$ and $0 < p < 1$.

The proof is provided in Appendix F 1. Then we see that the normalized QPFI for commuting-operator measurements will be

$$\gamma(\{M_i\}) = \max_{1 \leq k < l \leq d} \gamma_{kl}(\{M_i\}), \quad (64)$$

using Theorem 6, where

$$\gamma_{kl}(\{M_i\}) = \gamma(\{M_i\}|_{\text{span}\{|k\rangle, |l\rangle\}}), \quad (65)$$

and $\{M_i\}|_{\text{span}\{|k\rangle, |l\rangle\}}$ is the quantum measurement restricted in the subspace spanned by $|k\rangle$ and $|l\rangle$.

We show in Appendix F 2 that

$$\gamma_{kl}(\{M_i\}) = \sum_i \frac{p_{kl}^* (1 - p_{kl}^*) (m_k^{(i)} - m_l^{(i)})^2}{p_{kl}^* m_k^{(i)} + (1 - p_{kl}^*) m_l^{(i)}}, \quad (66)$$

where $p_{kl}^* \in (0, 1)$ is the unique solution to

$$\sum_{i=1}^r \frac{m_k^{(i)} (m_k^{(i)} - m_l^{(i)})^2}{(m_k^{(i)} + \frac{1-p_{kl}^*}{p_{kl}^*} m_l^{(i)})^2} = \sum_{i=1}^r \frac{m_l^{(i)} (m_k^{(i)} - m_l^{(i)})^2}{(\frac{p_{kl}^*}{1-p_{kl}^*} m_k^{(i)} + m_l^{(i)})^2} \quad (67)$$

and the corresponding optimal preprocessed state in $\text{span}\{|k\rangle, |l\rangle\}$ is

$$|\phi_{kl}^*\rangle = \sqrt{p_{kl}^*} |k\rangle + \sqrt{1-p_{kl}^*} |l\rangle, \quad (68)$$

$$|\phi_{kl}^{\perp*}\rangle = \sqrt{1-p_{kl}^*} |k\rangle - \sqrt{p_{kl}^*} |l\rangle. \quad (69)$$

(The symmetry transformations $|\phi^{\perp*}\rangle \mapsto -|\phi^{\perp*}\rangle$, $|k\rangle \mapsto e^{i\omega} |k\rangle$ and $|l\rangle \mapsto e^{i\omega'} |l\rangle$ for any $\omega, \omega' \in \mathbb{R}$ will generate alternative optimal solutions.) The optimal preprocessed state $(|\phi^*\rangle, |\phi^{\perp*}\rangle)$ in the entire Hilbert space that achieves Eq. (64) is chosen as $(|\phi_{kl}^*\rangle, |\phi_{kl}^{\perp*}\rangle)$ for (k, l) that maximizes $\gamma_{kl}(\{M_i\})$.

For the special case where $r = 2$, the problem reduces to the binary measurement problem discussed in Sec. IV C and p_{kl}^* can be found analytically. In general, however, the analytical solution to p_{kl}^* might not exist since it is a root of a high degree polynomial [Eq. (67)] and numerical methods are needed. Nonetheless, a simple analytical upper bound on $\gamma(\{M_i\})$ can still be obtained, as shown in the following theorem (see a detailed proof in Appendix F 3).

Theorem 7.—For commuting-operator measurements [Eq. (63)], the normalized QPFI $\gamma(\{M_i\})$ satisfies

$$\gamma(\{M_i\}) \leq 1 - \min_{kl} \left(\sum_i \sqrt{m_k^{(i)} m_l^{(i)}} \right)^2. \quad (70)$$

When there exists a (k, l) that minimizes $\sum_i \sqrt{m_k^{(i)} m_l^{(i)}}$ such that the set $\{m_k^{(i)}/m_l^{(i)}, 1 \leq i \leq r\}$ contains at most two elements, the inequality is tight.

To derive lower bounds on $\gamma(\{M_i\})$, one could replace p_{kl}^* with any $0 \leq p \leq 1$ in the expression Eq. (66). For example, taking $p = 1/2$, we have (as also shown in Ref. [55])

$$\gamma(\{M_i\}) \geq \max_{kl} \sum_i \frac{(m_k^{(i)} - m_l^{(i)})^2}{2(m_k^{(i)} + m_l^{(i)})} \quad (71)$$

$$\geq 1 - \min_{kl} \sum_i \sqrt{m_k^{(i)} m_l^{(i)}}, \quad (72)$$

where we use $m_k^{(i)} + m_l^{(i)} \leq (\sqrt{m_k^{(i)}} + \sqrt{m_l^{(i)}})^2$. Combining the upper and lower bounds, we observe that $\gamma(\{M_i\}) \approx 1$ when $\sum_i \sqrt{m_k^{(i)} m_l^{(i)}} \approx 0$. It means that the QPFI will be close to the QFI when there exist two basis states $|k\rangle$ and $|l\rangle$ such that the fidelity between two probability distributions $\{m_k^{(i)}\}$ and $\{m_l^{(i)}\}$ is close to zero (meaning that they are almost perfectly distinguishable).

The upper bound in Eq. (70) is saturated when the measurement is binary. Another physical example is lossy photodetection. The probability of detecting i photons given a Fock state of k ($i \leq k$) photons is $m_k^{(i)} = \binom{k}{i} (1-\eta)^i \eta^{k-i}$, where $1-\eta$ is the quantum efficiency of the photodetector. Assuming the maximal number of photons is N , it is simple to see that the optimal basis states are Fock states $|0\rangle, |N\rangle$. Since $m_0^{(0)} = 1$, only $m_0^{(0)}/m_N^{(0)}$ is nonvanishing

and thus $\gamma(\{M_i\})$ saturates the upper bound: $\gamma(\{M_i\}) = 1 - \eta^N$. [Technically, we need to assume all $m_k^{(i)} > 0$ to avoid the singularity issue, but the above statement holds because the value of $\gamma(\{M_i\})$ can be calculated by first adding a small perturbation to the detection errors (like in Theorem 1) and then taking the limit as the perturbation vanishes.]

Finally, note that although Theorems 6 and 7 do not directly tell us how to choose the two optimal basis states, such a choice may sometimes be obvious. For example, consider a n -qubit system ($\text{span}\{|1\rangle, |2\rangle\})^{\otimes n}$ measured by $\{M, \mathbb{1} - M\}^{\otimes n}$ (independently on each subsystem) and $M = (1 - m)|1\rangle\langle 1| + m|2\rangle\langle 2|$. Then using Theorem 6, due to the bit-flip symmetry and the fact that tracing out some parts of the quantum state will not increase its QPFI, it is clear that rotating $(|\psi_\theta\rangle, |\psi_\theta^\perp\rangle)$ into $\text{span}\{|1\rangle^{\otimes n}, |2\rangle^{\otimes n}\}$, or any other basis states, e.g., $\{|121\dots 1\rangle, |212\dots 2\rangle\}$ that are distinct on each qubit, must be an optimal choice. In general, it remains open if there is a simple criterion to help us select the optimal k and l besides a direct calculation of Eq. (66) [or sometimes Eq. (70)] for different k and l .

V. CLASSICALLY MIXED STATES

In this section, we consider another type of quantum states, which we called *classically mixed states*, with commuting-operator measurements. A classically mixed state is a state, which commutes with its derivative, e.g., Gibbs states whose temperature is to be estimated [83]. In this section, we use the following form of classically mixed states:

$$\zeta_\theta = \sum_{i=1}^D \lambda_{i,\theta} |i\rangle\langle i|, \quad (73)$$

where $D = \dim(\mathcal{H}_S)$, $\lambda_{i,\theta}$ are functions of θ (we will drop the subscript θ for conciseness), $\{|i\rangle\}$ is an orthonormal basis of \mathcal{H}_S that is independent of θ and we use ζ_θ to represent classically mixed states. Note that the QFI of Eq. (73) $J(\zeta_\theta) = \sum_{i=1}^D (\partial_\theta \lambda_i)^2 / \lambda_i$ is equal to the FI $J(\{\lambda_i\})$ of the classical distribution $\{\lambda_i\}_{i=1}^D$. Also, note that we assume in this section, without loss of generality, that the commuting-operator measurement $\{M_i\}$ and the classically mixed state ζ_θ share the same eigenstates $\{|i\rangle\}_{i=1}^{\max\{d,D\}}$, as it is always possible to apply a unitary rotation in the preprocessing control so that they are aligned.

We first show that optimizing the FI over quantum channels is equivalent to finding optimal stochastic matrices (which describes the transitions of a classical Markov chain) for the classical preprocessing optimization problem. Then we prove that the optimal control always corresponds to a stochastic matrix that has only elements 0 or 1, which we call a coarse-graining stochastic matrix. It implies that the QPFI is always attainable, and that

the QPFI can in some cases be strictly larger than the QUPFI. Finally, we closely examine the case of a binary measurement on a single qubit.

A. Optimization over stochastic matrices

Lemma 8.—Consider classically mixed states Eq. (73) and commuting-operator measurements Eq. (63). Then

$$F^{\mathbb{P}}(\zeta_\theta, \{M_i\}) = \sup_{P \in \mathbb{S}_{d,D}} J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}), \quad (74)$$

and when $d = D$,

$$F^{\mathbb{U}}(\zeta_\theta, \{M_i\}) \leq \sup_{P \in \mathbb{S}_{D,D}^{db}} J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}), \quad (75)$$

where $\mathbb{S}_{d,D}$ represents the set of $d \times D$ stochastic matrices of which every column vector sums up to one and $\mathbb{S}_{D,D}^{db}$ represents the set of $D \times D$ doubly stochastic matrices of which every column and row vector sums up to one, $\mathbf{m}^{(i)}$ is a column vector whose entries are $m_j^{(i)}$, λ_θ is a column vector whose entries are λ_i .

Proof.—Let $\mathcal{E}(\cdot) = \sum_j K_j(\cdot)K_j^\dagger$ be an arbitrary quantum channel, then we have

$$\text{Tr}(M_i \mathcal{E}(\zeta_\theta)) = \mathbf{m}^{(i)T} P \lambda_\theta, \quad (76)$$

where the matrix P satisfies $P_{\ell k} = \sum_j |\langle \ell | K_j | k \rangle|^2 = \sum_j |(K_j)_{\ell k}|^2$, which implies

$$J(\{\text{Tr}(M_i \mathcal{E}(\zeta_\theta))\}) = J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}). \quad (77)$$

We must have $\sum_\ell P_{\ell k} = \sum_{\ell j} \langle k | K_j^\dagger | \ell \rangle \langle \ell | K_j | k \rangle = \langle k | k \rangle = 1$, because $\sum_j K_j^\dagger K_j = \mathbb{1}$. Thus, P is a stochastic matrix. For any quantum channel, there exists a stochastic matrix such that Eq. (76) holds true, proving the left-hand side is no larger than the right-hand side in Eq. (74). Moreover, when $\mathcal{E}(\cdot) = U(\cdot)U^\dagger$ is a unitary channel, $P_{\ell k} = |U_{\ell k}|^2$ must be doubly stochastic, implying Eq. (75).

On the other hand, for any stochastic matrix P , we define $K_{(\ell,k)} = \sqrt{P_{\ell k}} | \ell \rangle \langle k |$ for $1 \leq \ell \leq d$ and $1 \leq k \leq D$. Then we have $\sum_{\ell k} K_{(\ell,k)}^\dagger K_{(\ell,k)} = \sum_{\ell k} P_{\ell k} | k \rangle \langle k | = \mathbb{1}$. And $\mathcal{E}(\cdot) = \sum_{(\ell,k)} K_{(\ell,k)}(\cdot)K_{(\ell,k)}^\dagger$ is then a quantum channel. For any stochastic matrix, there exists a quantum channel such that Eq. (76) holds true, proving the left-hand side is no smaller than the right-hand side in Eq. (74). ■

We show in Lemma 8 that the problem of optimizing preprocessing quantum controls on classically mixed states with commuting-operator measurements is equivalent to a

classical version of preprocessing optimization where

$$F^{\mathbb{P}}(\lambda_\theta, \{\mathbf{m}^{(i)}\}) := \sup_{P \in \mathbb{S}_{d,D}} J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}) \quad (78)$$

represents the classical FI with respect to a classical distribution λ_θ and a noisy measurement $\mathbf{m}^{(i)}$ satisfying $\sum_i \mathbf{m}^{(i)} = \mathbf{1}$ ($\mathbf{1}$ is a vector with all elements equal to 1), optimized over any stochastic mapping described by stochastic matrices. In particular, for perfect measurements where $(\mathbf{m}^{(i)})_j = \delta_{ij}$, $F^{\mathbb{P}}(\lambda_\theta, \{\mathbf{m}^{(i)}\}) = J(\lambda_\theta) = \sum_{i=1}^D (\partial_\theta \lambda_i)^2 / \lambda_i$ is the classical FI. Note that Theorem 10 presented later implies that the supremum of the FI over stochastic matrices is always attainable using some $P \in \mathbb{S}_{d,D}$ and it means we are allowed to replace $\sup_{P \in \mathbb{S}_{d,D}}$ by $\max_{P \in \mathbb{S}_{d,D}}$ in the definition [Eq. (78)].

B. Coarse-graining controls are optimal

We first consider the classical case and prove Eq. (78) can always be attained using some $d \times D$ stochastic matrix P where every element of P is either 0 or 1. We call this type of stochastic matrix a *coarse-graining stochastic matrix* in the sense that P sums up one or multiple entries of λ_θ to one entry in $P\lambda_\theta$, which is a coarse graining of measurement outcomes.

Lemma 9.—Given a classical probability distribution $\lambda_\theta \in \mathbb{R}^D$ and a measurement $\{\mathbf{m}^{(i)}\} \subseteq \mathbb{R}^d$ (satisfying $\sum_i \mathbf{m}^{(i)} = \mathbf{1}$). When $F^{\mathbb{P}}(\lambda_\theta, \{\mathbf{m}^{(i)}\})$ is attainable, there exists a $d \times D$ coarse-graining stochastic matrix P such that,

$$F^{\mathbb{P}}(\lambda_\theta, \{\mathbf{m}^{(i)}\}) = J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}). \quad (79)$$

Proof.—Suppose $F^{\mathbb{P}}(\lambda_\theta, \{\mathbf{m}^{(i)}\})$ is attainable and P^* is an optimal solution. We will show that there exists an optimal solution P whose every column vector contains one (and only) element equal to 1. If P^* does not satisfy this condition, without loss of generality, assume $P_{11}^* = t_1^*$ and $P_{21}^* = a_1^* - t_1^*$ where $0 < t_1^* < a_1^* \leq 1$. Let $P(t_1)$ be a stochastic matrix function of $t_1 \in [0, a_1^*]$ where $P(t_1)_{11} = t_1$, $P(t_1)_{21} = a_1^* - t_1$ and $P(t_1)_{\ell k} = (P^*)_{\ell k}$ for $(\ell, k) \neq (1, 1), (2, 1)$. We have the FI equal to

$$\begin{aligned} f(t_1) &= \sum_i \frac{(\partial_\theta (\mathbf{m}^{(i)T} P(t_1) \lambda_\theta))^2}{\mathbf{m}^{(i)T} P(t_1) \lambda_\theta} \\ &= \sum_i \frac{((m_1^{(i)} - m_2^{(i)}) \partial_\theta \lambda_1 t_1 + b^{*(i)})^2}{(m_1^{(i)} - m_2^{(i)}) \lambda_1 t_1 + a^{*(i)}}, \end{aligned}$$

where $a^{*(i)}$ and $b^{*(i)}$ are constants, independent of t_1 . The second-order derivative of $f(t_1)$ is

$$\frac{\partial^2 f(t_1)}{\partial t_1^2} = \sum_i \frac{2(m_1^{(i)} - m_2^{(i)})^2 (a^{*(i)} \partial_\theta \lambda_1 - b^{*(i)} \lambda_1)^2}{((m_1^{(i)} - m_2^{(i)}) \lambda_1 t_1 + a^{*(i)})^3}, \quad (80)$$

which is always non-negative. Therefore, $f(t_1)$ is a convex function and always attains its maximum at the boundary $t_1 = 0$ or $t_1 = a_1^*$. Repeat this argument many times, one can show that there exists an optimal solution P such that there is only one positive entry in every column. ■

Note that it is not necessarily true that the optimal coarse-graining stochastic matrix that maximizes $J(\{\mathbf{m}^{(i)T} P \lambda_\theta\})$ is a full-rank matrix. Consider the following example. Let $d = D = 3$, $r = 2$, $\lambda_\theta = (\cos^2 \theta, \frac{1}{2} \sin^2 \theta, \frac{1}{2} \sin^2 \theta)$, $\mathbf{m}^{(1)} = (1, \frac{1}{2}, 0)$, and $\mathbf{m}^{(2)} = (0, \frac{1}{2}, 1)$. Then it is clear that the following stochastic matrix is optimal:

$$P^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad (81)$$

because $J(\{\mathbf{m}^{(i)T} P^* \lambda_\theta\}) = J(\lambda_\theta) = 4$. However, it can be verified by enumeration that $J(\{\mathbf{m}^{(i)T} P \lambda_\theta\}) \leq 3$, whenever P is a permutation matrix, showing the nonoptimality of the full-rank stochastic matrices.

Using Lemma 8, we can show a similar result to Lemma 9 in the quantum case, that is, *coarse-graining channels* are optimal quantum controls.

Theorem 10.—Consider classically mixed states Eq. (73) and commuting-operator measurements Eq. (63). The QPFI is always attainable using the following type of quantum channels, which we call coarse-graining channels:

$$\mathcal{E}(\cdot) = \sum_{\ell k} P_{\ell k} |\ell\rangle \langle k| (\cdot) |k\rangle \langle \ell|, \quad (82)$$

where $P_{\ell k}$ is a $d \times D$ stochastic matrix satisfying $\sum_\ell P_{\ell k} = 1$ and $P_{\ell k} = 0$ or 1.

Proof.—By definition, there exists a sequence of channels $(\mathcal{E}_1, \dots, \mathcal{E}_n, \dots)$ such that $\lim_{n \rightarrow \infty} F(\mathcal{E}_n(\zeta_\theta), \{M_i\}) = F^{\mathbb{P}}(\zeta_\theta, \{M_i\})$. According to Eq. (77) in the proof of Lemma 8 and the arguments in Lemma 9, for every \mathcal{E}_n there exists a channel $\tilde{\mathcal{E}}_n$ of the form Eq. (82) such that $F(\mathcal{E}_n(\zeta_\theta), \{M_i\}) \leq F(\tilde{\mathcal{E}}_n(\zeta_\theta), \{M_i\})$. Therefore, $\lim_{n \rightarrow \infty} F(\tilde{\mathcal{E}}_n(\zeta_\theta), \{M_i\}) = F^{\mathbb{P}}(\zeta_\theta, \{M_i\})$. Since there are finite number of channels of the form Eq. (82), there must exist a $\mathcal{E}^* = \tilde{\mathcal{E}}_n$ for some n such that

$$F(\mathcal{E}^*(\zeta_\theta), \{M_i\}) = F^{\mathbb{P}}(\zeta_\theta, \{M_i\}), \quad (83)$$

proving the attainability of the QPFI. ■

Theorem 10 also implies that there is a gap between the QUPFI and the QPFI for general quantum states, unlike for pure states where the QUPFI is equal to the QPFI.

Theorem 11.—There exists a classically mixed state ζ_θ and a commuting-operator measurement $\{M_i\}$ such that

$$F^{\mathbb{U}}(\zeta_\theta, \{M_i\}) < F^{\mathbb{P}}(\zeta_\theta, \{M_i\}). \quad (84)$$

Proof.—Consider the example discussed below Lemma 9 and here we fix $\theta = \pi/4$. Theorem 10 implies that for $\zeta_\theta = \cos^2 \theta |1\rangle\langle 1| + \frac{1}{2} \sin^2 \theta |2\rangle\langle 2| + \frac{1}{2} \sin^2 \theta |3\rangle\langle 3|$,

$$J(\zeta_\theta) = F^{\mathbb{P}}(\zeta_\theta, \{M_i\}) = 4, \quad (85)$$

where $M_1 = |1\rangle\langle 1| + \frac{1}{2} |2\rangle\langle 2|$ and $M_2 = \frac{1}{2} |2\rangle\langle 2| + |3\rangle\langle 3|$. In general, given any stochastic matrix P , the probabilities for measurement outcomes 1 and 2 must have the form

$$p_1 = \mathbf{m}^{(1)T} P \boldsymbol{\lambda}_\theta = a \cos^2 \theta + b \sin^2 \theta, \quad (86)$$

$$p_2 = \mathbf{m}^{(2)T} P \boldsymbol{\lambda}_\theta = (1 - a) \cos^2 \theta + (1 - b) \sin^2 \theta, \quad (87)$$

for some $0 \leq a, b \leq 1$. Moreover, $J(\{p_1, p_2\}) = 4(a - b)^2 / (2 - (a + b)) / (a + b)$. And $J(\{p_1, p_2\}) = J(\zeta_\theta)$ if and only if $(a, b) = (1, 0)$ or $(0, 1)$. Noting that the situation where $(a, b) = (1, 0)$ or $(0, 1)$ is not possible if P is doubly stochastic. Applying Lemma 8, Eq. (84) is then proven. ■

The intuition behind this type of gap between the QPFI and the QUPFI stems from the fact that nonunitary operations, e.g., the coarse-graining channel, have the power of reducing the rank of quantum states, while unitary operations do not. Consequently, when certain conditions are met: (i) the noisy measurement under consideration is noiseless in a lower-dimensional subspace, e.g., $\text{span}\{|1\rangle, |3\rangle\}$ in the example above and (ii) the rank of the quantum state can be compressed without reducing its QFI, e.g., collapsing $\text{span}\{|2\rangle, |3\rangle\}$ into $\text{span}\{|3\rangle\}$, nonunitary preprocessing operations can achieve the optimal QFI. In contrast, relying solely on unitary preprocessing for high-rank states results in unavoidable measurement noise and suboptimal performance.

Finally, we note that although the implementation of general quantum preprocessing channels can sometimes be challenging with the requirement of preparing a clean ancillary system that occurs in Stinespring's dilation (see Proposition 2), the resources needed to perform coarse-graining channels can be reduced in many cases. Firstly, the ancilla size required to perform coarse-graining channels is, in principle, smaller than d^2 that is required in general cases. In fact, any coarse-graining channel defined by Eq. (82) can be simulated using a d -dimensional ancilla, e.g., by first performing a unitary operation on $\mathcal{H}_S \otimes \mathcal{H}_{A_1}$ that maps $|k\rangle_S |0\rangle_{A_1} \mapsto |k\rangle_S |\iota(k)\rangle_{A_1}$ for all k , where $\iota(k)$ corresponds to the index of the row such that $P_{\iota(k)k} = 1$, and then discarding the probe system \mathcal{H}_S . Secondly, the coarse-graining channel can also be performed on certain quantum states by resetting some parts of the system with no additional ancillas in some cases. For example, consider a two-qubit quantum state $\zeta_\theta = \cos^2 \theta |00\rangle\langle 00| + \frac{1}{2} \sin^2 \theta |10\rangle\langle 10| + \frac{1}{2} \sin^2 \theta |11\rangle\langle 11|$ and measurement operators $M_0 = |00\rangle\langle 00|$

$|00\rangle + \frac{1}{2} |01\rangle\langle 01| + \frac{1}{2} |10\rangle\langle 10|$ and $M_1 = |11\rangle\langle 11| + \frac{1}{2} |01\rangle\langle 01| + \frac{1}{2} |10\rangle\langle 10|$. The coarse-graining channel mapping $|00\rangle \mapsto |00\rangle$, $|10\rangle \mapsto |11\rangle$, and $|11\rangle \mapsto |11\rangle$ is optimal and it can be performed by first resetting the second qubit to $|0\rangle$ and then applying a CNOT gate that maps $|00\rangle \mapsto |00\rangle$ and $|10\rangle \mapsto |11\rangle$. Note that resetting qubits is usually considered much less noisy than measuring ones, e.g., in nitrogen-vacancy centers [27,28].

C. Binary measurement on a single qubit

With Theorem 10, in principle, one can find the QPFI for classically mixed states and commuting-operator measurements by exhausting all channels of the form Eq. (82) which is contained in a finite set. However, since the number of coarse-graining stochastic matrices is large, the exhaustion procedure will be too costly. Here we closely examine a special case where a classically mixed state is measured by a binary measurement on a single qubit. The time to exhausting all coarse-graining matrices is exponentially large with respect to the state dimension D . We will show that the time to find a solution can be reduced to a linear complexity by narrowing down the possible forms of the optimal controls.

To be specific, consider the binary measurement $M_1 = M = m_1 |1\rangle\langle 1| + m_2 |2\rangle\langle 2|$, $M_2 = \mathbb{1} - M$ (assuming $m_2 \leq \min\{m_1, 1 - m_1\}$), and $\zeta_\theta = \sum_{i=1}^D \lambda_i |i\rangle\langle i|$. Then using Lemma 8, we first have $F^{\mathbb{P}}(\zeta_\theta, \{M_i\}) = \max_{\mathbf{t}} f(\mathbf{t})$ where

$$f(\mathbf{t}) := J(\{p_\theta(\mathbf{t}), 1 - p_\theta(\mathbf{t})\}), \quad (88)$$

and

$$p_\theta(\mathbf{t}) := m_2 + (m_1 - m_2) \mathbf{t}^T \boldsymbol{\lambda}_\theta, \quad (89)$$

and \mathbf{t} is a column vector in $[0, 1]^D$, corresponding to the first row of the stochastic matrix P in the proof of Lemma 8. (Note that although from Theorem 10, it is possible to restrict \mathbf{t} to $\{0, 1\}^D$, and we keep the generality of \mathbf{t} by allowing it to be in $[0, 1]^D$ for later use.)

Without loss of generality, we can arrange the order of the positive elements in $\boldsymbol{\lambda}_\theta$ such that

$$(\partial_\theta \lambda_i) / \lambda_i \geq (\partial_\theta \lambda_j) / \lambda_j, \quad \forall i < j \text{ and } \lambda_{ij} > 0. \quad (90)$$

Then we assert that

$$F^{\mathbb{P}}(\zeta_\theta, \{M_i\}) = \max_{i \in [1, D-1]} \max\{f(\mathbf{1}_{\leq i}), f(\mathbf{1}_{\geq i})\}, \quad (91)$$

where $\mathbf{1}_{\leq i}$ represents the vector whose the first i elements are equal to 1 and the rest are zero. $\mathbf{1}_{\geq i+1} = \mathbf{1} - \mathbf{1}_{\leq i}$.

Now we prove Eq. (91). Choose an optimal $\mathbf{t}^* \in [0, 1]^D$ that maximizes $f(\mathbf{t})$. We prove Eq. (91) in each of the following three cases:

- (i) $\mathbf{t}^{*T} \partial_\theta \lambda_\theta = 0$. Then the QPFI is zero and Eq. (91) is trivial.
- (ii) $\mathbf{t}^{*T} \partial_\theta \lambda_\theta > 0$. If there exists $i < j$ such that $t_i^* < 1$, $t_j^* > 0$ and $\lambda_{i,j} > 0$. Then define $t_i^{**} = t_i^* + \epsilon/\lambda_i$ and $t_j^{**} = t_j^* - \epsilon/\lambda_j$ where $\epsilon = \min\{(1 - t_i^*)\lambda_i, t_j^*\lambda_j\}$ and $t_{k \neq i,j}^{**} = t_{k \neq i,j}^*$. Then we have either $t_i^{**} = 1$ or $t_j^{**} = 0$. Moreover, we have $f(\mathbf{t}^{**}) \geq f(\mathbf{t}^*)$ because $\mathbf{t}^{*T} \lambda_\theta = \mathbf{t}^{**T} \lambda_\theta$ and $\mathbf{t}^{*T} \partial_\theta \lambda_\theta \leq \mathbf{t}^{**T} \partial_\theta \lambda_\theta$. Then \mathbf{t}^{**} is also optimal. Repeating this procedure, we can always find an optimal \mathbf{t} of the form $(1 \cdots 1 t 0 \cdots 0)$ for some $t \in [0, 1]$. $\Omega(1) \leq J(\rho_\theta^{(n)}) \leq e^{o(n)}$. Using the same convexity argument as in the proof of Lemma 9, we can further show t can be taken to be 1 or 0. Equation (91) is proven.
- (iii) $\mathbf{t}^{*T} \partial_\theta \lambda_\theta < 0$. If there exists $i < j$ such that $t_i^* > 0$, $t_j^* < 1$ and $\lambda_{i,j} > 0$. Then define $t_i^{**} = t_i^* - \epsilon/\lambda_i$ and $t_j^{**} = t_j^* + \epsilon/\lambda_j$ where $\epsilon = \min\{(1 - t_j^*)\lambda_j, t_i^*\lambda_i\}$ and $t_{k \neq i,j}^{**} = t_{k \neq i,j}^*$. Moreover, we have $f(\mathbf{t}^{**}) \geq f(\mathbf{t}^*)$ because $\mathbf{t}^{*T} \lambda_\theta = \mathbf{t}^{**T} \lambda_\theta$ and $\mathbf{t}^{*T} \partial_\theta \lambda_\theta \geq \mathbf{t}^{**T} \partial_\theta \lambda_\theta$. Then \mathbf{t}^{**} is also optimal. Repeating this procedure, we can always find an optimal \mathbf{t} of the form $(0 \cdots 0 t 1 \cdots 1)$ for some $t \in [0, 1]$. Using the same convexity argument as in the proof of Lemma 9, we can further show t can be taken to be 1 or 0. Equation (91) is proven.

VI. GENERAL QUANTUM STATES

In Secs. IV and V, we have obtained fruitful results on preprocessing optimization for pure states and classically mixed states. Here, we consider the QPFI for general mixed states and derive useful upper and lower bounds on them.

A. Upper bound

Theorem 12.—Given any density operator ρ_θ and quantum measurement $\{M_i\}$, we have

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq \gamma(\{M_i\})J(\rho_\theta). \quad (92)$$

Proof.—Suppose \mathcal{H}_{A_1} and \mathcal{H}_{A_2} are ancillary systems such that $\mathcal{H}_{A_1} \otimes \mathcal{H}_S \cong \mathcal{H}_{A_2} \otimes \mathcal{H}_{S'}$ and $\dim(\mathcal{H}_{A_1}) \geq \dim(\mathcal{H}_{S'})^2$, where \mathcal{H}_S and $\mathcal{H}_{S'}$ are the systems ρ_θ and $\{M_i\}$ act on. We also define an additional environmental system \mathcal{H}_E satisfying $\dim(\mathcal{H}_E) = \dim(\mathcal{H}_S)$. Let $\psi_\theta = |\psi_\theta\rangle \langle \psi_\theta|_{ES}$ denote the purifications of ρ_θ in $\mathcal{H}_E \otimes \mathcal{H}_S$. Using the purification-based definition of QFI [17,18], we have

$$J(\rho_\theta) = \min_{\psi_\theta: \rho_\theta = \text{Tr}_E(\psi_\theta)} J(\psi_\theta). \quad (93)$$

Choose ψ_θ^* to be the optimal purification ψ_θ^* that minimizes $J(\psi_\theta)$ such that $J(\rho_\theta) = J(\psi_\theta^*)$. Then

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = F^{\mathbb{U}}(\rho_\theta \otimes |0\rangle \langle 0|_{A_1}, \{M_i \otimes \mathbb{1}_{A_2}\}) \quad (94)$$

$$\leq F^{\mathbb{U}}(\psi_\theta^* \otimes |0\rangle \langle 0|_{A_1}, \{\mathbb{1}_E \otimes M_i \otimes \mathbb{1}_{A_2}\}) \quad (95)$$

$$= \gamma(\{\mathbb{1}_E \otimes M_i \otimes \mathbb{1}_{A_2}\})J(\psi_\theta^* \otimes |0\rangle \langle 0|_{A_1}) \quad (96)$$

$$= \gamma(\{M_i\})J(\rho_\theta), \quad (97)$$

where we use Proposition 2, Eq. (40) and Theorem 5. ■

Theorem 12 provides an upper bound on the QPFI for general quantum states. In particular, it shows the ratio between the QPFI and the QFI is always upper bounded by a state-independent constant $\gamma(\{M_i\})$, which is attainable when the state is pure and gives rise to the following CRB for general quantum states under noisy measurement $\{M_i\}$:

$$\Delta \hat{\theta} \geq \frac{1}{\sqrt{N_{\text{expr}} \gamma(\{M_i\}) J(\rho_\theta)}}. \quad (98)$$

B. Lower bound

Lemma 13.—Consider a density operator ρ_θ and quantum measurement $\{M_i\}$. Assume $\{T_i\}$ is a QFI-attainable measurement, i.e., $F(\rho_\theta, \{T_i\}) = J(\rho_\theta)$. Let the quantum-classical channel $\mathcal{T}(\cdot) = \sum_i \text{Tr}(\cdot T_i) |i\rangle \langle i|_C$ where $\{|i\rangle_C\}$ is an orthonormal basis of an auxiliary system \mathcal{H}_C . Then

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \geq F^{\mathbb{P}}(\mathcal{T}(\rho_\theta), \{M_i\}). \quad (99)$$

The proof Lemma 13 is straightforward—it immediately follows from the definition of the QPFI. The equality holds true when $\{M_i\}$ is a projection onto an orthonormal basis of $\mathcal{H}_{S'}$, i.e., $\{M_i = |i\rangle \langle i|_{i=1}^{\dim(\mathcal{H}_{S'})}\}$.

Note that the equality in Lemma 13 also holds when ρ_θ is a classically mixed state and the QFI-attainable measurement is chosen to be the projective measurement onto the basis of \mathcal{H}_S so that $\mathcal{T}(\rho_\theta) = \rho_\theta$. For general mixed states, since $\mathcal{T}(\rho_\theta)$ is a classically mixed state, the results in Sec. V can be applied here to analyze $F^{\mathbb{P}}(\mathcal{T}(\rho_\theta), \{M_i\})$ and derive lower bounds for general mixed states. For example, one can divide the measurement operators into two subsets, restrict the measurement in a two-dimensional subspace, and then use our previous result of the binary measurement on a qubit for classically mixed states to derive an efficiently computable lower bound on the QPFI.

Note that unlike the upper bound (Theorem 12), there are no constant lower bounds independent of ρ_θ on the ratio between $F^{\mathbb{P}}(\rho_\theta, \{M_i\})$ and $J(\rho_\theta)$. For example, consider the single qubit case where $\rho_\theta = \cos^2 \theta |1\rangle \langle 1| + \sin^2 \theta |2\rangle \langle 2|$, $M_1 = (1 - m) |1\rangle \langle 1| + m |2\rangle \langle 2|$, and $M_2 = \mathbb{1} - M_1$ ($0 < m < 1/2$). We have, from Sec. V C, that

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = \frac{4(1 - 2m)^2 \sin^2(2\theta)^2}{1 - (1 - 2m) \cos(2\theta)^2}, \quad (100)$$

which tends to zero as $\theta \rightarrow 0$ [and the optimal preprocessing is identity when $\theta \in (0, \pi/4)$]. On the other hand, $J(\rho_\theta) = 4$ is a constant, showing that $F^{\text{IP}}(\rho_\theta, \{M_i\})/J(\rho_\theta)$ has no state-independent constant lower bounds.

VII. GLOBAL PREPROCESSING: ASYMPTOTIC LIMITS

In this section, we consider the power of global quantum preprocessing in the asymptotic limit (see Fig. 2). We consider a multipartite system $\mathcal{H}_S = \mathcal{H}^{\otimes n}$ and $\mathcal{H}_{S'} = \mathcal{H}'^{\otimes n}$ where $\dim \mathcal{H} = D$ and $\dim \mathcal{H}' = d$, a set of quantum states $\rho_\theta^{(n)}$ in $\mathcal{H}^{\otimes n}$, and quantum measurements $\{M_i\}^{\otimes n}$ that can be written as tensor products of identical measurements on each subsystem \mathcal{H}' . Arbitrary (and usually global) preprocessing quantum channels \mathcal{E} are applied before the noisy measurement. We will show that for a generic class of quantum states, the QPFI can reach the QFI asymptotically for large n . Note that the QPFI is in general not achievable [55] when \mathcal{E} can act only locally and independently on each subsystem.

A. Attaining the QFI with noisy measurements

Theorem 14.—Given a set of quantum states $\{\rho_\theta^{(n)}\}_n$ where $\rho_\theta^{(n)}$ is a function of θ and acts on $\mathcal{H}^{\otimes n}$ for each n , we have

$$\lim_{n \rightarrow \infty} \frac{F^{\text{IP}}(\rho_\theta^{(n)}, \{M_i\}^{\otimes n})}{J(\rho_\theta^{(n)})} = 1, \quad (101)$$

if for each $\rho_\theta^{(n)}$ the following are true:

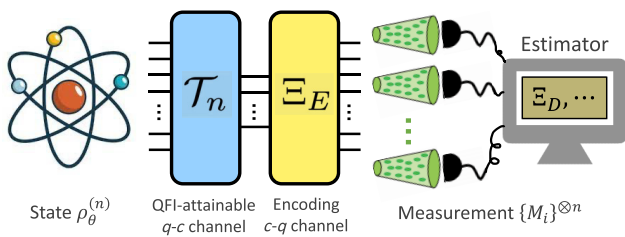


FIG. 2. A quantum state $\rho_\theta^{(n)}$ in an n -partite system is estimated using n identical noisy measurements acting on each subsystem, described by $\{M_i\}^{\otimes n}$. The QPFI can approach the QFI in the asymptotic limit $n \rightarrow \infty$ if the sufficient condition in Theorem 14 is satisfied. The optimal control is the composition of a quantum-classical channel $\mathcal{T}_n(\cdot) = \sum_{i=1}^{r_n} \text{Tr}(\cdot T_i^{(n)}) |e_i\rangle \langle e_i|$ where the measurement $\{T_i^{(n)}\}$ is asymptotically QFI attainable, and an encoding channel Ξ_E chosen as the optimal encoding channel for $\mathcal{M}^{\otimes n}$ from the HSW theorem. Note that the decoding channel Ξ_D from the HSW theorem only needs to be used in a classical postprocessing manner.

- (a) There exists a quantum measurement $\{T_i^{(n)}\}$ whose number of measurement outcomes is r_n such that

$$\lim_{n \rightarrow \infty} \frac{F(\rho_\theta^{(n)}, \{T_i^{(n)}\})}{J(\rho_\theta^{(n)})} = 1, \text{ and } \lim_{n \rightarrow \infty} \frac{\log r_n}{n} < C(\mathcal{M}), \quad (102)$$

where \log is the binary logarithm and $C(\mathcal{M})$ is the classical capacity of the quantum-classical channel $\mathcal{M}(\cdot) = \sum_i \text{Tr}(\cdot M_i) |i\rangle \langle i|_C$ ($\{|i\rangle_C\}$ is an orthonormal basis of an auxiliary system \mathcal{H}_C).

- (b) The regularity conditions are satisfied:

- (1) When $\partial_\theta \lambda_i \neq 0$, $\lambda_i = 1/e^{o(n)}$, where $\lambda_i := \text{Tr}(\rho_\theta^{(n)} T_i^{(n)})$ and $\{T_i^{(n)}\}$ is defined above.
- (2) $\Omega(1) \leq J(\rho_\theta^{(n)}) \leq e^{o(n)}$.

Theorem 14 provides a sufficient condition to attain the QFI using noisy measurements in the asymptotic limit $n \rightarrow \infty$. We will first provide a proof of Theorem 14, and return to the physical understandings of the sufficient condition later. Readers who are not interested in the technical details can skip the technical proof and advance to the discussion part.

In the proof, we will make use of a quantum-classical channel \mathcal{T}_n defined using $\{T_i^{(n)}\}$, and an encoding channel Ξ_E , such that $F(\Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)}), \{M_i\}^{\otimes n})$ approaches $J(\rho_\theta^{(n)})$ asymptotically (see Fig. 2). Intuitively speaking, the first step \mathcal{T}_n is to simulate the (asymptotically) QFI-attainable measurement $\{T_i^{(n)}\}$ on $\rho_\theta^{(n)}$ to transform it into a classically mixed state $\mathcal{T}_n(\rho)$ such that $J(\mathcal{T}_n(\rho_\theta^{(n)})) = J(\rho_\theta^{(n)})$. The second step is to choose a suitable encoding channel Ξ_E such that the classical information in $\mathcal{T}_n(\rho_\theta^{(n)})$ is fully preserved under $\mathcal{M}^{\otimes n}$, i.e., $J(\mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})) \approx J(\mathcal{T}_n(\rho_\theta^{(n)}))$, leading to the asymptotic attainability of the QFI. Here Ξ_E , along with a corresponding decoding channel Ξ_D , is chosen such that $\Xi_E \circ \mathcal{M}^{\otimes n} \circ \Xi_D$ is asymptotically equal to a completely dephasing channel with a transmission rate approximately equal to $C(\mathcal{M})$, which is guaranteed to exist using the HSW theorem [60,61].

Proof of Theorem 14.—We first choose an α such that $\lim_{n \rightarrow \infty} \log r_n/n < \alpha < C(\mathcal{M})$. According to the definition of the classical capacity of quantum channels [68], for any $\epsilon > 0$, there exists an n_0 such that for any $n > n_0$, there exist an encoding channel Ξ_E and a decoding channel Ξ_D such that

$$\|\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E - \mathcal{D}_2^{\otimes [an]}\|_{\diamond} \leq \epsilon, \quad (103)$$

where \mathcal{D}_2 is a completely dephasing qubit channel acting on qubit Hilbert space \mathcal{H}_B , i.e. $\mathcal{D}_2(\cdot) = |0\rangle \langle 0| (\cdot) |0\rangle \langle 0| + |1\rangle \langle 1| (\cdot) |1\rangle \langle 1|$ and $\|\cdot\|_{\diamond}$ is the diamond norm of a quantum channel [68] defined by $\|\Phi\|_{\diamond} =$

$\max\{\|(\Phi \otimes \mathbb{1})(X)\|_1, \|X\|_1 \leq 1\}$ (Φ and $\mathbb{1}$ act on systems of the same dimension, and $\|\cdot\|_1$ denotes the trace norm). Moreover, $\epsilon = e^{-\Omega(n)}$ (see a proof in Appendix G). For any operator σ , we have

$$\|\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E(\sigma) - \mathcal{D}_2^{\otimes[\alpha n]}(\sigma)\|_1 \leq \epsilon \|\sigma\|_1. \quad (104)$$

We also assume n_0 is large enough such that for any $n > n_0$, $r_n \leq 2^{\lfloor \alpha n \rfloor}$.

Let $\mathcal{T}_n(\cdot) := \sum_{i=1}^{r_n} \text{Tr}((\cdot)T_i^{(n)}) |e_i\rangle \langle e_i|$ where we choose $\{|e_i\rangle\}_{i=1}^{r_n}$ to be a subset of the computational basis in $\mathcal{H}_B^{\otimes[\alpha n]}$. Without loss of generality, we assume $\lambda_i = \text{Tr}(\rho_\theta^{(n)} T_i^{(n)}) > 0$ for all i (we can always exclude the terms that are equal to zero), then $\mathcal{T}_n(\rho_\theta^{(n)}) := \sum_{i=1}^{r_n} \lambda_i |e_i\rangle \langle e_i|$ and

$$\begin{aligned} F(\rho_\theta^{(n)}, \{T_i^{(n)}\}) &\geq F(\Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)}), \{M_i\}^{\otimes n}) \\ &= J(\mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})) \\ &\geq J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})), \end{aligned} \quad (105)$$

where we use the monotonicity of the QFI in the first and third inequalities and $J(\{p_{i,\theta}\}) = J(\sum_i p_{i,\theta} |i\rangle \langle i|)$ for any classical probability distribution $\{p_{i,\theta}\}$ in the second equality. Then we have

$$\frac{J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)}))}{F(\rho_\theta^{(n)}, \{T_i^{(n)}\})} \leq 1. \quad (106)$$

Next we aim to show $J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})) / F(\rho_\theta^{(n)}, \{T_i^{(n)}\})$ is lower bounded by a constant that approaches 1 for large n . First, assume $n > n_0$, we have

$$\begin{aligned} J(\mathcal{D}_2^{\otimes[\alpha n]} \circ \mathcal{T}_n(\rho_\theta^{(n)})) \\ = J(\mathcal{T}_n(\rho_\theta^{(n)})) = F(\rho_\theta^{(n)}, \{T_i^{(n)}\}) = \sum_{i=1}^{r_n} \frac{(\partial_\theta \lambda_i)^2}{\lambda_i}, \end{aligned} \quad (107)$$

where we use $\mathcal{D}_2^{\otimes[\alpha n]}(|e_i\rangle \langle e_i|) = |e_i\rangle \langle e_i|$ in the first equality. On the other hand, consider

$$\begin{aligned} J(\mathcal{D}' \circ \Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})) \\ =: \sum_{i=1}^{r_n} \frac{(\partial_\theta \eta_i)^2}{\eta_i} \leq J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})), \end{aligned} \quad (108)$$

where $\mathcal{D}'(\cdot) = \sum_i |e_i\rangle \langle e_i| (\cdot) |e_i\rangle \langle e_i|$, $\eta_i = \lambda_i + \delta_i$, and $\delta_i = \langle e_i | (\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E - \mathcal{D}_2^{\otimes[\alpha n]}) \circ \mathcal{T}_n(\rho_\theta^{(n)}) |e_i\rangle$. We will also assume n is large enough such that $\delta_i < \lambda_i$, which is possible due to Eq. (104) and the regularity condition (1).

Then we have

$$\begin{aligned} \sum_{i=1}^{r_n} \frac{(\partial_\theta \eta_i)^2}{\eta_i} &= \sum_{i=1}^{r_n} \frac{(\partial_\theta \lambda_i + \partial_\theta \delta_i)^2}{\lambda_i + \delta_i} \\ &= \sum_{i=1}^{r_n} (\partial_\theta \lambda_i + \partial_\theta \delta_i)^2 \left(\frac{1}{\lambda_i} - \frac{\delta_i}{\lambda_i^2 (1 + \xi_i)^2} \right) \\ &\geq \sum_{i=1}^{r_n} ((\partial_\theta \lambda_i)^2 + 2\partial_\theta \lambda_i \partial_\theta \delta_i) \frac{1}{\lambda_i} \left(1 - \frac{|\delta_i|}{\lambda_i} \right) \\ &\geq F(\rho_\theta^{(n)}, \{T_i^{(n)}\}) - \sum_{i=1}^{r_n} \frac{(\partial_\theta \lambda_i)^2}{\lambda_i^2} |\delta_i| - \sum_{i=1}^{r_n} \frac{2|\partial_\theta \lambda_i \partial_\theta \delta_i|}{\lambda_i} \\ &\geq F(\rho_\theta^{(n)}, \{T_i^{(n)}\}) \\ &\quad - \epsilon \max_i \left(\frac{(\partial_\theta \lambda_i)^2}{\lambda_i^2} + 2 \left| \frac{\partial_\theta \lambda_i}{\lambda_i} \right| J(\rho_\theta^{(n)})^{1/2} \right). \end{aligned} \quad (109)$$

In the second equality above, we use the Taylor expansion $1/(1 + \delta_i/\lambda_i) = 1 - \delta_i/\lambda_i/(1 + \xi_i)^2$ for some $\xi_i \in [0, \delta_i/\lambda_i]$. In the last inequality above, we use Eq. (104) to derive that $\sum_i |\delta_i| \leq \epsilon \|\rho_\theta^{(n)}\|_1 = \epsilon$ and

$$\begin{aligned} \sum_i |\partial_\theta \delta_i| &\leq \epsilon \|\partial_\theta \mathcal{T}_n(\rho_\theta^{(n)})\|_1 \\ &\leq \epsilon J(\mathcal{T}_n(\rho_\theta^{(n)}))^{1/2} \leq \epsilon J(\rho_\theta^{(n)})^{1/2}. \end{aligned} \quad (110)$$

Here we use the inequality $\|\partial_\theta \zeta_\theta\|_1 \leq J(\zeta_\theta)^{1/2}$ for any classically mixed state $\zeta_\theta = \sum_i \mu_i |i\rangle \langle i|$, which is true because $(\sum_i |\partial_\theta \mu_i|)^2 \leq \sum_i (\partial_\theta \mu_i)^2 / \mu_i$ from the Cauchy-Schwarz inequality. Note that it also holds that $\|\partial_\theta \sigma_\theta\|_1 \leq J(\sigma_\theta)^{1/2}$ for general mixed states σ_θ [84,85].

Finally, from the monotonicity of the QFI [i.e., $\sum_i (\partial_\theta \lambda_i)^2 / \lambda_i \leq J(\rho_\theta^{(n)})$] and the regularity conditions (1) and (2), we have

$$\max_i \frac{(\partial_\theta \lambda_i)^2}{\lambda_i^2} + 2 \left| \frac{\partial_\theta \lambda_i}{\lambda_i} \right| \|\partial_\theta \rho\|_1 \leq J(\rho_\theta^{(n)}) e^{o(n)}. \quad (111)$$

Taking the limit $n \rightarrow \infty$ in Eq. (109), from $\epsilon = e^{-\Omega(n)}$ and Eq. (111), we have

$$\lim_{n \rightarrow \infty} \frac{1}{F(\rho_\theta^{(n)}, r\{T_i^{(n)}\})} \sum_{i=1}^{r_n} \frac{\partial_\theta \eta_i^2}{\eta_i} \geq 1. \quad (112)$$

Combining Eqs. (106), (108), and (112), we have

$$\lim_{n \rightarrow \infty} \frac{J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)}))}{F(\rho_\theta^{(n)}, \{T_i^{(n)}\})} = 1. \quad (113)$$

Since $\lim_{n \rightarrow \infty} F(\rho_\theta^{(n)}, \{T_i^{(n)}\})/J(\rho_\theta^{(n)}) = 1$ and

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta^{(n)}, \{M_i\}^{\otimes n}) &\geq F(\Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)}), \{M_i\}^{\otimes n}) \\ &\geq J(\Xi_D \circ \mathcal{M}^{\otimes n} \circ \Xi_E \circ \mathcal{T}_n(\rho_\theta^{(n)})), \end{aligned} \quad (114)$$

we must have

$$\lim_{n \rightarrow \infty} \frac{F^{\mathbb{P}}(\rho_\theta^{(n)}, \{M_i\}^{\otimes n})}{J(\rho_\theta^{(n)})} = 1, \quad (115)$$

proving the theorem. \blacksquare

B. Discussion

Here we discuss the intuitions behind the sufficient condition in Theorem 14 and describe the relevant situations where it is satisfied. We will see that the sufficient condition is satisfied for a generic class of quantum states $\rho_\theta^{(n)}$ and noisy measurements $\{M_i\}$.

Let us first explain the meaning of the condition Eq. (102). It states that there exists an (asymptotically) QFI-attainable measurement for $\rho_\theta^{(n)}$ that has a *small* number of measurement outcomes. Specifically, the number of measurement outcomes r_n should be smaller than $2^{C(\mathcal{M})n}$ (asymptotically) where $C(\mathcal{M})$ is the classical capacity of the quantum measurement $\{M_i\}$ under consideration, i.e., Theorem 14 applies when

$$\log r_n < C(\mathcal{M})n + o(n). \quad (116)$$

The requirement [Eq. (116)] is satisfied by many practically relevant quantum states and measurements. In fact, whenever the classical capacity of \mathcal{M} is positive, $r_n = e^{o(n)}$ is a sufficient (but not necessary) condition of Eq. (116). Below we provide several typical examples where the QFI-attainable measurement with a subexponential number of outcomes exists. See Appendix A for additional details.

- (1) *Low-rank states.* For pure states, it was known that there exist two-outcome QFI-attainable measurements [14]. (Note that Ref. [55] contains another proof of Theorem 14 when $\rho_\theta^{(n)}$ is pure.) More generally, any $\rho_\theta^{(n)}$ that is supported on a subspace with a subexponential dimension also has a QFI-attainable measurement with a subexponential number of outcomes.
- (2) *Symmetric states.* The second example with a QFI-attainable measurement with a subexponential number of outcomes is symmetric (permutation-invariant) states (e.g., tensor products of n identical mixed states). According to the Schur-Weyl duality [86,87], $\mathcal{H}_S = (\mathbb{C}^D)^{\otimes n}$ can be decomposed as

$\bigoplus_\nu (\mathcal{H}_\nu(U(D)) \otimes \mathcal{H}_\nu(S_n))$, where $\mathcal{H}_\nu(U(D))$ and $\mathcal{H}_\nu(S_n)$ are irreducible representation spaces of the unitary group $U(D)$ and the permutation group S_n with index ν . Any symmetric state $\rho_\theta^{(n)}$ can be written as

$$\rho_\theta^{(n)} = \bigoplus_\nu \left(p_\nu \rho_\nu^{(n)} \otimes \frac{\mathbb{1}_\nu}{\dim(\mathcal{H}_\nu(S_n))} \right), \quad (117)$$

where $\rho_\nu^{(n)}$ are mixed states acting on $\mathcal{H}_\nu(U(D))$ and p_ν satisfies $\sum_\nu p_\nu = 1$ (both of which can be functions of θ). Then a QFI-attainable measurement with a subexponential number of outcomes $\{\bigoplus_\nu (T_i)_\nu \otimes \mathbb{1}_\nu\}$ of $\rho_\theta^{(n)}$ can be constructed from a QFI-attainable measurement $\{\bigoplus_\nu (T_i)_\nu\}$ of $\bigoplus_\nu p_\nu \rho_\nu^{(n)}$. Let us estimate the number of measurement outcomes: ν corresponds to Young diagrams (i.e., partitions of n into D parts), implying the number of different indices ν is $O(n^{D-1})$. For any ν , $\dim(\mathcal{H}_\nu(U(D)))$ is equal to the number of semistandard Young tableaux, which is at most $O(n^{D(D-1)/2})$ according to the Weyl dimension formula [88]. The number of measurement outcomes is thus upper bounded by $\sum_\nu \dim(\mathcal{H}_\nu(U(D))) = O(n^{(D-1)(D/2+1)})$.

- (3) *Gibbs states.* For classically mixed states $\rho_\theta^{(n)}$, the projection onto the eigenstates of $\rho_\theta^{(n)}$ is QFI attainable but has exponentially many measurement outcomes. However, we argue that in many cases, a subexponential number of projections onto direct sums of eigenspaces are sufficient to attain the QFI up to the leading order, so that Theorem 14 applies. For instance, consider the Gibbs state

$$\rho_\theta^{(n)} = \frac{1}{\sum_\nu e^{-\theta E_\nu}} \sum_\nu e^{-\theta E_\nu} |\nu\rangle \langle \nu|, \quad (118)$$

where $\{|\nu\rangle\}$ are energy eigenstates with eigenvalues $\{E_\nu\}$ and θ is the inverse temperature to be estimated. The QFI is equal to the variance of energy, i.e.,

$$J(\rho_\theta^{(n)}) = \sum_\nu p_\nu E_\nu^2 - \left(\sum_\nu p_\nu E_\nu \right)^2, \quad (119)$$

where $p_\nu = e^{-\theta E_\nu} / \sum_\nu e^{-\theta E_\nu}$. Assume the energy eigenvalues lie in $[0, \widehat{E}]$, where $\widehat{E} = \Theta(n)$ (which is a standard assumption in condensed-matter systems) and divide them into intervals $\{I_k = [E_k, E_{k+1})\}_{k=1}^n$, such that $E_0 = 0$, $E_n = \widehat{E}$ and $E_{k+1} - E_k = \Delta E = \widehat{E}/n^2$. Consider the projections $\{\Pi_k\}_{k=1}^n$ onto the direct sums of eigenspaces corresponding to all

eigenvalues in I_k . The FI is

$$F(\rho_\theta^{(n)}, \{\Pi_k\}) = \sum_{k=1}^n p_k E_k^2 - \left(\sum_v p_v E_v \right)^2, \quad (120)$$

where $p_k = \sum_{v: E_v \in I_k} p_v$ and $E_k = 1/p_k \sum_{v: E_v \in I_k} p_v E_v$. Then we have $J(\rho_\theta^{(n)}) - F(\rho_\theta^{(n)}, \{\Pi_k\}) \leq \sum_k p_k (E_{k+1}^2 - E_k^2) \leq 2\widehat{E}\Delta E = \Theta(1)$. Combining with the regularity condition (2), it implies that $F(\rho_\theta^{(n)}, \{\Pi_k\})$ is equal to $J(\rho_\theta^{(n)})$ up to the leading order.

Next, let us explain the intuitions behind the regularity conditions:

- (1) Regularity condition (1) states that when the probability of obtaining measurement outcome i depends on θ (i.e., $\partial_\theta \lambda_i \neq 0$), it must be no smaller than an inverse of a subexponential function of n , that is, the probability to detect i cannot be exponentially small. This is also a practically reasonable assumption as we would want to exclude the singular cases where an exponentially small signal provides a nontrivial contribution to the QFI.
- (2) Regularity condition (2) requires that the QFI of $\rho_\theta^{(n)}$ does not decrease with n asymptotically, which should be satisfied in any practically relevant cases. It also requires the QFI to be subexponential, which is a natural assumption in quantum sensing experiments [note that the Heisenberg limit implies $J(\rho_\theta^{(n)}) = O(n^2)$].

Lastly, we briefly comment on the resources required to implement optimal preprocessing controls. First, the total number of ancillary qubits required to implement the desired preprocessing channel $\Xi_E \circ \mathcal{T}_n$ is at most $O(n)$, because in general $\log(D^n r_n)$ ancillary qubits are sufficient to implement the QFI-attainable q -c channel \mathcal{T}_n and another $\log(D^n r_n)$ ancillary qubits are sufficient to implement the encoding channel Ξ_E . The gate complexity to implement \mathcal{T}_n is expected to depend on the structure of the quantum state $\rho_\theta^{(n)}$. For example, for symmetric states, the Schur transform, efficiently implementable [89], can be an important step in \mathcal{T}_n . Unitary gates that are used in aligning the output basis of \mathcal{T}_n to the input basis of the encoding channel Ξ_E should also be taken into consideration. For example, in the special case where ρ_n is a low-rank classically mixed state, \mathcal{T}_n should be a rotation that matches ρ_n eigenstates to the input basis of Ξ_E . The gate complexity to implement the optimal encoding channel Ξ_E is high in general. However, when r_n is subexponential (as we discussed above), the encoding channel does not need to be capacity achieving as it needs only to reliably transmit an exponentially small amount of information, potentially making it relatively easier to implement (the details are left for future

discussion). For example, when $r_n = 2$, a simple repetition code mapping $|0\rangle$ to $|0\rangle^{\otimes n}$ and $|1\rangle$ to $|1\rangle^{\otimes n}$ will be optimal. Finally, note that although we have shown that $\Xi_E \circ \mathcal{T}_n$ is optimal, other simpler optimal preprocessing channels may still exist. For example, for pure states, unitary controls are optimal according to Theorem 5, requiring no ancillas; and a design of an optimal preprocessing unitary is presented in Ref. [55].

C. Examples

Lastly, we present three simple but natural examples with powerful global preprocessing controls that can be efficiently implemented using $O(\log^2 n)$ -depth circuits, assuming arbitrary two-qubit gates and all-to-all connectivity (see details in Appendix H). In these three examples, we always assume $\mathcal{H} = \mathcal{H}' = \text{span}\{|0\rangle, |1\rangle\}$ are qubit systems and the quantum measurement is $\{M_i\} = \{M_0, M_1\}$, where $M_0 = (1 - m)|0\rangle\langle 0| + m|1\rangle\langle 1|$ ($0 < m < 1/2$), and $M_1 = \mathbb{1} - M_0$.

In the first two examples, our preprocessing circuits manage to achieve a FI that is asymptotically equal to the QFI for any noise rate m . In the third example, our preprocessing circuit achieves a FI that is asymptotically equal to $2/\pi$ of the QFI, which still beats local controls in the noise regime $m \geq 0.1011$. The guideline to design these circuits is to convert the quantum state to a two-level state in $\text{span}\{|0\rangle^{\otimes n}, |1\rangle^{\otimes n}\}$ whose probability (or amplitude) distribution encodes θ . Then a *majority voting postprocessing method* can be used to estimate θ with a vanishingly small measurement error. Specifically, in the majority voting postprocessing method, we partition the measurement outcomes from measuring the two-level state using $\{M_0, M_1\}^{\otimes n}$, which are represented by n -bit strings in $\{0, 1\}^n$, into two sets depending on whether the Hamming weight of the string is larger than $\lfloor n/2 \rfloor$. The FI of this binary probability distribution achieves the desired value asymptotically.

The first example is phase sensing using GHZ states [5], where

$$|\psi_\theta^{(n)}\rangle = \frac{e^{in\theta} |0\rangle^{\otimes n} + e^{-in\theta} |1\rangle^{\otimes n}}{\sqrt{2}}, \quad (121)$$

and an optimal preprocessing circuit U^G that achieves

$$F(U^G \psi_\theta^{(n)} (U^G)^\dagger, \{M_i\}^{\otimes n}) \xrightarrow{n \rightarrow \infty} J(\psi_\theta^{(n)}) = 4n^2 \quad (122)$$

is shown in Fig. 3(a), mapping $|\psi_\theta^{(n)}\rangle$ to

$$\cos(n\theta) |0\rangle^{\otimes n} + i \sin(n\theta) |1\rangle^{\otimes n}. \quad (123)$$

The majority voting postprocessing method gives an optimal estimator of θ .

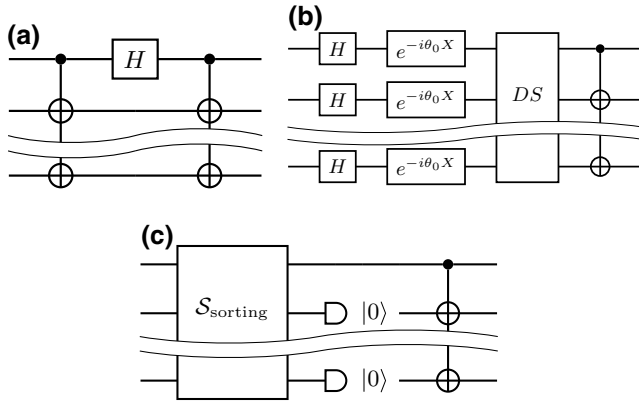


FIG. 3. Examples of preprocessing circuits (U^G or \mathcal{E}^G). (a) Phase sensing using GHZ states. The optimal circuit is composed of two $C(\text{NOT})^{n-1}$ gates, which, conditioning on the first qubit being $|1\rangle$, performs $X^{\otimes n-1}$ (X is the Pauli- X gate) on the remaining $n-1$ qubits, and a Hadamard gate H acting on the first qubit between them. The circuit depth is $O(\log n)$. (b) Phase sensing using product states. The optimal circuit is composed of a global H gate and a global Pauli- X rotation, a desymmetrization gate DS that maps W state $1/\sqrt{n}(|10\dots 0\rangle + |010\dots 0\rangle + \dots + |0\dots 01\rangle)$ to $|10\dots 0\rangle$ and $|0\dots 0\rangle$ to $|0\dots 0\rangle$, and a $C(\text{NOT})^{n-1}$ gate. The circuit depth is $O(\log n)$. (c) Phase sensing using classically mixed states. $\mathcal{S}_{\text{sorting}}$ is a sorting channel with circuit depth $O(\log^2 n)$, that uses $O(n \log^2 n)$ ancillary qubits and outputs one qubit in state Eq. (129). It first sorts the bit string and then swaps the first qubit with the $\lfloor n \sin^2 \theta_0 \rfloor$ th qubit. (The D-shape detectors mean the qubits are completely discarded.)

The second example is phase sensing using product pure states (usually known as Ramsey interferometry [90]), where

$$|\psi_\theta^{(n)}\rangle = \left(\frac{e^{i\theta} |0\rangle + e^{-i\theta} |1\rangle}{\sqrt{2}} \right)^{\otimes n}, \quad (124)$$

and an optimal preprocessing circuit of depth $O(\log n)$ that achieves

$$F(U^G \psi_\theta^{(n)} (U^G)^\dagger, \{M_i\}^{\otimes n}) \xrightarrow{n \rightarrow \infty} J(\psi_\theta^{(n)}) = 4n \quad (125)$$

is shown in Fig. 3(b). Here we assume θ_0 is a rough estimate of θ such that $|\theta - \theta_0| \ll 1/\sqrt{n}$. The first step is to implement global Hadamard gates and Pauli- X rotations such that $|\psi_\theta^{(n)}\rangle$ is mapped to $(e^{-i\theta_0 X})^{\otimes n} H^{\otimes n} |\psi_\theta^{(n)}\rangle = (\cos(\theta - \theta_0) |0\rangle + i \sin(\theta - \theta_0) |1\rangle)^{\otimes n}$. The second step is to apply a desymmetrization gate DS and a $C(\text{NOT})^{n-1}$ gate such that the state is approximately mapped to

$$\cos(\sqrt{n}(\theta - \theta_0)) |0\rangle^{\otimes n} + i \sin(\sqrt{n}(\theta - \theta_0)) |1\rangle^{\otimes n}, \quad (126)$$

with an error $O(n(\theta - \theta_0)^2)$. The majority voting postprocessing method gives an optimal estimator of θ .

The third example is phase sensing using classically mixed states [which can be seen as Eq. (124) after global Hadamard gates and completely dephasing noise], where

$$\rho_\theta^{(n)} = (\cos^2 \theta |0\rangle\langle 0| + \sin^2 \theta |1\rangle\langle 1|)^{\otimes n}. \quad (127)$$

We assume $\theta \in (0 + \varepsilon, \pi/4 - \varepsilon)$ for some constant $\varepsilon > 0$. We show a preprocessing channel \mathcal{E}^G [in Fig. 3(c)] of circuit depth $O(\log^2 n)$ that achieves

$$F(\mathcal{E}^G(\rho_\theta^{(n)}), \{M_i\}^{\otimes n}) \xrightarrow{n \rightarrow \infty} \frac{8}{\pi} n = \frac{2}{\pi} J(\rho_\theta^{(n)}). \quad (128)$$

After a sorting channel $\mathcal{S}_{\text{sorting}}$ and discarding all qubits except the first qubit, the first qubit is in state

$$p_\theta |0\rangle\langle 0| + (1 - p_\theta) |1\rangle\langle 1|, \quad (129)$$

where p_θ is the probability that after flipping n coins whose probability of getting heads are $\sin^2 \theta$, the number of heads are smaller than or equal to $\lfloor n \sin^2 \theta_0 \rfloor$ (θ_0 is a rough estimate of θ satisfying $|\theta - \theta_0| \ll 1/\sqrt{n}$). A FI asymptotically equal to $(8/\pi)n$ can then be achieved using a $C(\text{NOT})^{n-1}$ gate with $n-1$ ancillas initialized in $|0\rangle^{\otimes n-1}$ and the majority voting postprocessing method.

Note that $\rho_\theta^{(n)}$ is a symmetric state. According to the discussion in Sec. VII B, the QPFI should be asymptotically equal to the QFI, but whether there exists an efficient implementation of the optimal preprocessing circuits is unknown. Here we demonstrate the advantage of global controls by providing an efficient but suboptimal circuit in Fig. 3(c). The first part ($\mathcal{S}_{\text{sorting}}$) of our circuit can be viewed as the optimal quantum-classical channel \mathcal{T}_n in Theorem 14. The second part that encodes one qubit into n qubits is, however, suboptimal. [In order to faithfully transmit all probability distribution information, the encoding channel in the second part needs to encode $\log(n+1)$ qubits into n qubits.]

Nonetheless, our circuit in Fig. 3(c) is superior to any local preprocessing controls in the noise regime where

$$m \geq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \approx 0.101. \quad (130)$$

This can be proven noting that the optimal FI achievable using arbitrary local channels satisfies

$$\begin{aligned} & \max_{\mathcal{E}=\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \dots \otimes \mathcal{E}_n} F(\mathcal{E}(\rho_\theta^{(n)}), \{M_i\}^{\otimes n}) \\ & = F^{\mathbb{P}}(\rho_\theta^{(1)}, \{M_i\}) n \leq (1 - 2m)^2 n, \end{aligned} \quad (131)$$

from Eq. (100). Thus it is always smaller than $(8/\pi)n$ when $m \geq \frac{1}{2} - 1/\sqrt{2\pi}$. Specially, when $m \rightarrow \frac{1}{2}$, the linear constant of the locally optimized FI is vanishingly small,

while the suboptimal global one is still above a positive number.

VIII. CONCLUSIONS AND OUTLOOK

We conducted a systematic study of the preprocessing optimization problem for noisy quantum measurements in quantum metrology. The QPFI (i.e., the FI of noisy measurement statistics optimized over all preprocessing quantum channels), that we defined and investigated in depth, sets an ultimate precision bound for noisy measurement of quantum states. Our results provide, in many cases, both numerically and analytically, approaches to identifying the optimal preprocessing controls that will be of great importance in alleviating the effect of measurement noise in quantum sensing experiments.

We also considered, specifically, the asymptotic limit of the QPFI in multiprobe systems with individual measurement on each probe. We proved the convergence of the QPFI to the QFI when there exists an (asymptotically) QFI-attainable measurement with a sufficiently small number of measurement outcomes, by establishing a connection to the classical channel capacity theorem. It would be interesting to explore, in future works, if the number of outcomes for QFI-attainable measurements can be easily bounded given a quantum state.

Although we have discussed only two types of quantum preprocessing controls, CPTP maps and unitary maps, our biconvex formulation might be generalized to cover other more restricted types of quantum controls. We also narrowed the analytical forms of optimal controls for pure states and classically mixed states down to rotations onto the span of two basis states and coarse-graining channels, respectively, but it remains open whether a simple method exists to help us determine the exact operations.

Finally, there are a few important directions to extend our results to, e.g., incorporating the state preparation optimization into the QPFI optimization problem, considering the preprocessing optimization in multiparameter estimation where the incompatibility of optimal preprocessings for different parameters might become an issue, and finding optimal preprocessings for other information processing tasks beyond quantum metrology, such as state tomography and discrimination.

ACKNOWLEDGMENTS

We thank Senrui Chen, Kun Fang, Jan Kołodyński, Yaodong Li, Zi-Wen Liu, John Preskill, Alex Retzker, Mark Wilde, and Tianci Zhou for helpful discussions. The authors acknowledge funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant No. PHY-1733907). T.G. further acknowledges funding provided by the Quantum Science and Technology Scholarship of the Israel Council for Higher Education.

APPENDIX A: QFI-ATTAINABLE MEASUREMENTS

In this Appendix, we provide several simple facts about the QFI-attainable measurements for quantum states, that will be useful in the main text.

- (1) *Necessary and sufficient condition* [14]. Given a quantum state ρ_θ , $\{M_i\}$ attains the QFI if and only if

$$\forall i = 1, \dots, r, \exists c_i \in \mathbb{R}, \text{ such that} \\ M_i^{1/2} \rho_\theta^{1/2} = c_i M_i^{1/2} L_\theta \rho_\theta^{1/2}, \quad (\text{A1})$$

where L_θ is the symmetric logarithmic derivative (SLD) operator, which is a Hermitian operator defined by $\partial_\theta \rho_\theta = \frac{1}{2}(L_\theta \rho_\theta + \rho_\theta L_\theta)$ and the QFI $J(\rho_\theta) = \text{Tr}(\rho_\theta L_\theta^2)$. (Note that although the definition of L_θ is not unique when ρ_θ is not full rank, one can choose an arbitrary one and the QFI and the necessary and sufficient condition will be invariant.) Clearly, the necessary and sufficient condition Eq. (A1) is satisfied when the measurement is chosen as the rank-one projection onto eigenstates of L_θ .

- (2) *Low-rank states*. If ρ_θ and $\partial_\theta \rho_\theta$ are supported on a D' -dimensional subspace of \mathcal{H}_S ($D' \leq D$) (which should be true when ρ_θ is supported on a $\lfloor D'/2 \rfloor$ -dimensional subspace), there exists a SLD L_θ supported on the D' -dimensional subspace. Moreover, let $L_\theta = \sum_{i=1}^{D'} \ell_i |i\rangle \langle i|$ where $\ell_i > 0$ and $\{|i\rangle\}$ is an orthonormal basis for the D' -dimensional subspace, it can be verified that $\{M_i\}_{i=1}^{D'}$ (that has D' measurement outcomes) is QFI-attainable when $M_i = |i\rangle \langle i|$ for $i = 1, \dots, D' - 1$ and $M_{D'} = \mathbb{1} - \sum_{i=1}^{D'} M_i$. That means, a state of a subexponential rank must have a QFI-attainable measurement with a subexponential number of measurement outcomes. In particular, for pure states, binary measurements are sufficient to attain the QFI.
- (3) *Classically mixed states*. For classically mixed states $\rho_\theta = \sum_{i=1}^D \lambda_i |i\rangle \langle i|$, the SLD is also diagonal in the basis $\{|i\rangle\}$. It implies that the rank-one projection onto the basis is QFI attainable for classically mixed states.
- (4) *Block-diagonal states*. More generally, consider block-diagonal states $\rho_\theta = \bigoplus_{v,\theta} p_{v,\theta} \rho_{v,\theta}$, where $\rho_{v,\theta}$ are supported on different orthogonal subspaces. The SLD can also be block diagonal, implying that there exists a QFI-attainable measurement of the form $\{\bigoplus_v (T_i)_v\}$, whose number of measurement outcomes is at most the rank of ρ_θ , where $\{(T_i)_v\}$ is a quantum measurement for each fixed v .

APPENDIX B: MATHEMATICAL PROPERTIES OF THE QPFI

Here we briefly discuss the mathematical properties of the QPFI. We will always assume quantum measurements are nontrivial in this Appendix (i.e., $\exists M_i \not\propto \mathbb{1}$ for all $\{M_i\}$).

- (1) *Faithfulness.* The QPFI is faithful, that means,

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \geq 0, \quad (\text{B1})$$

and the equality holds if and only if $\partial_\theta \rho_\theta = 0$. The non-negativity follows from the faithfulness of the classical FI. In order to see when $\partial_\theta \rho_\theta \neq 0$, $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) > 0$, we use Lemma 13 in Sec. VI. Assume $\partial_\theta \rho_\theta \neq 0$. $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) > F^{\mathbb{P}}(\mathcal{T}(\rho_\theta), \{M_i\})$, where $\mathcal{T}(\rho_\theta)$ is a classically mixed state that satisfies $J(\mathcal{T}(\rho_\theta)) = J(\rho_\theta) > 0$ (from the faithfulness of the QFI). Consider the following simplification of measurements: dividing $\{M_i\}$ into two subsets and restricting them in a two-dimensional subspace such that the measurement becomes a nontrivial binary measurement on a qubit $\{\tilde{M}, \mathbb{1} - \tilde{M}\}$. Then we see that $F^{\mathbb{P}}(\mathcal{T}(\rho_\theta), \{M_i\}) \geq F^{\mathbb{P}}(\mathcal{T}(\rho_\theta), \{\tilde{M}, \mathbb{1} - \tilde{M}\}) > 0$ from the exact expression of the QPFI in Eq. (91).

- (2) *QFI as an upper bound.* The QPFI is always no larger than the QFI. Note that

$$F(\mathcal{E}(\rho_\theta), \{M_i\}) = F(\rho_\theta, \{\mathcal{E}^\dagger(M_i)\}) \leq J(\rho_\theta), \quad (\text{B2})$$

where \mathcal{E}^\dagger is the dual map of \mathcal{E} and the second inequality follows from the definition of the QFI and the fact that $\{\mathcal{E}^\dagger(M_i)\}$ is still a POVM when \mathcal{E} is a CPTP map. Taking the supremum over \mathcal{E} on both sides in the above inequality, we have proved $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq J(\rho_\theta)$.

- (3) *Monotonicity.* The QPFI is monotonic, i.e., $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \geq F^{\mathbb{P}}(\mathcal{E}(\rho_\theta), \{M_i\})$ for any CPTP map

\mathcal{E} by definition. Or equivalently, $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \geq F^{\mathbb{P}}(\rho_\theta, \{\mathcal{E}^\dagger(M_i)\})$ where \mathcal{E}^\dagger is the dual map of \mathcal{E} .

- (4) *Convexity.* The QPFI is convex in quantum states when $\{M_i\}$ is fixed. That is, for $p \in (0, 1)$ independent of θ ,

$$\begin{aligned} F^{\mathbb{P}}(p\rho_\theta + (1-p)\sigma_\theta, \{M_i\}) &= \sup_{\mathcal{E}} F(\mathcal{E}(p\rho_\theta + (1-p)\sigma_\theta), \{M_i\}) \\ &\leq \sup_{\mathcal{E}} pF(\mathcal{E}(\rho_\theta), \{M_i\}) + (1-p)F(\mathcal{E}(\sigma_\theta), \{M_i\}) \end{aligned} \quad (\text{B3})$$

$$\leq pF^{\mathbb{P}}(\rho_\theta, \{M_i\}) + (1-p)F^{\mathbb{P}}(\sigma_\theta, \{M_i\}), \quad (\text{B4})$$

where we use the convexity of the classical FI in the second step. Similarly, we also have

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta, \{pM_i + (1-p)M'_i\}) &\leq pF^{\mathbb{P}}(\rho_\theta, \{M_i\}) + (1-p)F^{\mathbb{P}}(\rho_\theta, \{M'_i\}). \end{aligned} \quad (\text{B6})$$

- (5) *Additivity.* When quantum states under consideration are pure, the QPFI is additive because of Eq. (40) and the additivity of the QFI, i.e., $F^{\mathbb{P}}(\psi_\theta \otimes \psi'_\theta, \{M_i\}) = F^{\mathbb{P}}(\psi_\theta, \{M_i\}) + F^{\mathbb{P}}(\psi'_\theta, \{M_i\})$ when ψ_θ and ψ'_θ are pure.

For general mixed states, there is no general inequality relation between $F^{\mathbb{P}}(\rho_\theta \otimes \sigma_\theta, \{M_i\})$ and $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) + F^{\mathbb{P}}(\sigma_\theta, \{M_i\})$. Consider the simple example where one (or two identical) qubit state $\rho_\theta = \cos^2 \theta |1\rangle\langle 1| + \sin^2 \theta |2\rangle\langle 2|$ ($\theta \in (0, \pi/2)$), is measured using the binary measurement on a qubit: $M_1 = (1-m)|1\rangle\langle 1| + m|2\rangle\langle 2|$ and $M_2 = \mathbb{1} - M_1$ ($0 < m < 1/2$). We have, from Eq. (91), that

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = \frac{4(1-2m)^2 \sin(2\theta)^2}{1 - (1-2m) \cos(2\theta)^2}, \quad (\text{B7})$$

and

$$F^{\mathbb{P}}(\rho_\theta \otimes \rho_\theta, \{M_i\}) = \max \left\{ \frac{16(1-2m)^2 \cos(\theta)^6 \sin(\theta)^2}{(1-m)m + (1-2m)^2 (\cos \theta)^4 (1 - (\cos \theta)^4)}, \frac{16(1-2m)^2 \sin(\theta)^2 \cos(\theta)^6}{(1-m)m + (1-2m)^2 (\sin \theta)^4 (1 - (\sin \theta)^4)} \right\}. \quad (\text{B8})$$

Consider the limit $m \rightarrow 0$, we have $F^{\mathbb{P}}(\rho_\theta \otimes \rho_\theta, \{M_i\}) < 2F^{\mathbb{P}}(\rho_\theta, \{M_i\})$ (which is expected because $\{M_i\}$ when $m = 0$ is QFI attainable for ρ_θ and the QFI is additive). On the other hand, one can immediately find cases where $F^{\mathbb{P}}(\rho_\theta \otimes \rho_\theta, \{M_i\}) > 2F^{\mathbb{P}}(\rho_\theta, \{M_i\})$, e.g., when $m = 0.1$ and

$\theta = \pi/8$. For any fixed $m > 0$, there is a threshold of θ above which the sign of $F^{\mathbb{P}}(\rho_\theta \otimes \rho_\theta, \{M_i\}) - 2F^{\mathbb{P}}(\rho_\theta, \{M_i\})$ changes from positive to negative.

Finally, we can consider multiple states under multiple measurements. We have, by definition, $F^{\mathbb{P}}(\rho_\theta \otimes \sigma_\theta, \{M_i\} \otimes \{M'_i\}) \geq F^{\mathbb{P}}(\rho_\theta, \{M_i\}) + F^{\mathbb{P}}(\sigma_\theta,$

$\{M_i'\}$) and the inequality can be strict (see the convergence to the QFI in the asymptotic limit in Sec. VII).

APPENDIX C: ATTAINABILITY OF THE QPFI

Here we prove several results that are related to the attainability of the QPFI and the QUPFI.

We first show the existence of the optimal controls (or unitaries) for generic noisy measurement that has nonzero noise in all subspaces.

Lemma 15.—For arbitrary quantum states ρ_θ and noisy measurements $\{M_i\}$ such that $\min_i \lambda_{\min}(M_i) > 0$, the supremums in Eqs. (7) and (10) are attainable. Here $\lambda_{\min}(\cdot)$ represents the minimum eigenvalue of an operator.

Proof.—By definition, there exists a sequence of quantum channels $(\mathcal{E}_1, \dots, \mathcal{E}_n, \dots)$ such that

$$F(\mathcal{E}_n(\rho_\theta), \{M_i\}) \geq F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta_n, \quad (\text{C1})$$

where $\lim_{n \rightarrow \infty} \eta_n = 0$. Since the set of quantum channels is bounded and closed, there exists a limiting point of $(\mathcal{E}_1, \dots, \mathcal{E}_n, \dots)$ that we denote by \mathcal{E} . Without loss of generality, we assume the sequence converges and $\mathcal{E} = \lim_{n \rightarrow \infty} \mathcal{E}_n$. And

$$\lim_{n \rightarrow \infty} F(\mathcal{E}_n(\rho_\theta), \{M_i\}) = \lim_{n \rightarrow \infty} \sum_i \frac{(\text{Tr}(\mathcal{E}_n(\partial_\theta \rho_\theta) M_i))^2}{\text{Tr}(\mathcal{E}_n(\rho_\theta) M_i)} \quad (\text{C2})$$

$$= \sum_i \frac{\lim_{n \rightarrow \infty} (\text{Tr}(\mathcal{E}_n(\partial_\theta \rho_\theta) M_i))^2}{\lim_{n \rightarrow \infty} \text{Tr}(\mathcal{E}_n(\rho_\theta) M_i)} \quad (\text{C3})$$

$$= \sum_i \frac{(\text{Tr}(\mathcal{E}(\partial_\theta \rho_\theta) M_i))^2}{\text{Tr}(\mathcal{E}(\rho_\theta) M_i)} = F(\mathcal{E}(\rho_\theta), \{M_i\}), \quad (\text{C4})$$

where we use $\text{Tr}(\mathcal{E}_n(\rho_\theta) M_i) > \min_i \lambda_{\min}(M_i) > 0$ for all n . Then we must have $F(\mathcal{E}(\rho_\theta), \{M_i\}) \geq F^{\mathbb{P}}(\rho_\theta, \{M_i\})$ using Eq. (C1). Since $F(\mathcal{E}(\rho_\theta), \{M_i\}) \leq F^{\mathbb{P}}(\rho_\theta, \{M_i\})$ by definition, we have

$$F(\mathcal{E}(\rho_\theta), \{M_i\}) = F^{\mathbb{P}}(\rho_\theta, \{M_i\}), \quad (\text{C5})$$

proving the existence of the optimal channels. The existence of the optimal unitaries can also be proven analogously. ■

As a corollary of Lemma 15, we show that for any measurement whose QPFI (or QUPFI) may not be attainable, there always exists a measurement in its neighborhood such that its QPFI (or QUPFI) is attainable and close to that of the original measurement.

Lemma 16.—For any quantum state ρ_θ , quantum measurement $\{M_i\}$ and $\eta > 0$, there always exists $\{M_i^{(\epsilon)}\}$ and a

constant $c > 0$ such that the corresponding QPFI and the QUPFI are attainable, and when $\epsilon < c$,

$$F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}) \geq F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta, \quad (\text{C6})$$

$$F^{\mathbb{U}}(\rho_\theta, \{M_i^{(\epsilon)}\}) \geq F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta. \quad (\text{C7})$$

Proof.—Assume $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta \geq F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta > 0$. By definition, we can pick \mathcal{E} and U such that

$$F(\mathcal{E}(\rho_\theta), \{M_i\}) \geq F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta/2, \quad (\text{C8})$$

$$F(U\rho_\theta U^\dagger, \{M_i\}) \geq F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta/2. \quad (\text{C9})$$

Let

$$c' = \min \left\{ \min_{i: \text{Tr}(\mathcal{E}(\rho_\theta) M_i) \neq 0} \frac{\text{Tr}(\mathcal{E}(\rho_\theta) M_i)}{\text{Tr}(M_i)}, \min_{i: \text{Tr}(U(\rho_\theta) U^\dagger M_i) \neq 0} \frac{\text{Tr}(U(\rho_\theta) U^\dagger M_i)}{\text{Tr}(M_i)} \right\}, \quad (\text{C10})$$

define [assuming $d = \dim(\mathcal{H}_{S'})$]

$$M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \frac{\mathbb{1}}{d}, \quad (\text{C11})$$

and assume ϵ is small enough such that

$$1 - \epsilon - \frac{\epsilon}{dc'}(1 - \epsilon) > \frac{F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta}{F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta/2}. \quad (\text{C12})$$

Note that a similar construction of $\rho^{(\epsilon)}$, instead of $M_i^{(\epsilon)}$, was used in Ref. [91] to remove singularity of the QFI. Using Lemma 15, it is clear that the QPFI and the QUPFI for $\{M_i^{(\epsilon)}\}$ are attainable. Furthermore, we have

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}) &\geq \sum_{i: \text{Tr}(\mathcal{E}(\rho_\theta) M_i^{(\epsilon)}) \neq 0} \frac{(\text{Tr}(\mathcal{E}(\partial_\theta \rho_\theta) M_i^{(\epsilon)}))^2}{\text{Tr}(\mathcal{E}(\rho_\theta) M_i^{(\epsilon)})} \\ &\geq \sum_{i: \text{Tr}(\mathcal{E}(\rho_\theta) M_i) \neq 0} \frac{(1 - \epsilon)^2 (\text{Tr}(\mathcal{E}(\partial_\theta \rho_\theta) M_i))^2}{(1 - \epsilon) \text{Tr}(\mathcal{E}(\rho_\theta) M_i) + \text{Tr}(M_i) \epsilon / r} \end{aligned} \quad (\text{C13})$$

$$\geq \sum_{i: \text{Tr}(\mathcal{E}(\rho_\theta) M_i) \neq 0} \frac{(\text{Tr}(\mathcal{E}(\partial_\theta \rho_\theta) M_i))^2}{\text{Tr}(\mathcal{E}(\rho_\theta) M_i)} \left(1 - \epsilon - \frac{\epsilon}{dc'}(1 - \epsilon)\right) \quad (\text{C14})$$

$$\begin{aligned} &\geq F(\mathcal{E}(\rho_\theta), \{M_i\}) \frac{F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta}{F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta/2} \\ &\geq F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta, \end{aligned} \quad (\text{C15})$$

proving Eq. (C6). Equation (C7) is also true, similarly. When $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) - \eta \leq 0$ or $F^{\mathbb{U}}(\rho_\theta, \{M_i\}) - \eta \leq 0$, the results also follow trivially. ■

Finally, we are ready to provide a proof of Theorem 1, which shows a way to calculate the QPFI by considering the limit of the QPFI for a set of generic noisy measurements in its neighborhood. (Note that the theorem stated below also holds for the QUPFI.)

Theorem 1.—Let $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \mathbb{1}/d$, where $d = \dim(\mathcal{H}_{S'})$ and $0 < \epsilon < 1$. Then

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) = \lim_{\epsilon \rightarrow 0^+} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}), \quad (\text{C9})$$

and the QPFI $F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\})$ is attainable for any $\epsilon \in (0, 1]$.

Proof.—For any $\eta > 0$, following Lemma 16, we have

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}) + \eta, \quad (\text{C16})$$

when ϵ is small enough, where $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \mathbb{1}/d$. Take the limit $\epsilon \rightarrow 0^+$ on both sides, we have $F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq \liminf_{\epsilon \rightarrow 0^+} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}) + \eta$ for any $\eta > 0$, implying

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \leq \liminf_{\epsilon \rightarrow 0^+} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}). \quad (\text{C17})$$

On the other hand, consider a quantum channel for $0 < \epsilon < 1$,

$$\mathcal{E}_\epsilon(\sigma) = (1 - \epsilon)\sigma + \epsilon \text{Tr}(\sigma) \frac{\mathbb{1}}{d}, \quad (\text{C18})$$

Then we have $\text{Tr}(\mathcal{E}_\epsilon(\sigma)M_i) = \text{Tr}(\sigma \mathcal{E}_\epsilon^\dagger(M_i)) = \text{Tr}(\sigma M_i^{(\epsilon)})$ for any σ . By definition, we have

$$\begin{aligned} F^{\mathbb{P}}(\rho_\theta, \{M_i\}) &= \sup_{\mathcal{E}} F(\mathcal{E}(\rho_\theta), \{M_i\}) \\ &\geq \sup_{\mathcal{E}} F(\mathcal{E}_\epsilon(\rho_\theta), \{M_i\}) \\ &= F^{\mathbb{P}}(\rho_\theta, \{\mathcal{E}_\epsilon^\dagger(M_i)\}) = F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}). \end{aligned} \quad (\text{C19})$$

Take the limit $\epsilon \rightarrow 0^+$ on both sides, we have

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\}) \geq \limsup_{\epsilon \rightarrow 0^+} F^{\mathbb{P}}(\rho_\theta, \{M_i^{(\epsilon)}\}). \quad (\text{C20})$$

The theorem then follows from Eqs. (C17) and (C20). ■

APPENDIX D: GLOBAL OPTIMIZATION ALGORITHM FOR BICONVEX OPTIMIZATION PROBLEMS

In Sec. III, we showed that the QPFI can be obtained from the following biconvex optimization problem [Eq. (20)]:

$$F^{\mathbb{P}}(\rho_\theta, \{M_i\})^{-1} = \inf_{(\mathbf{x}, \Omega)} \text{Tr}((X_2 \otimes \rho_\theta^T)\Omega), \quad (\text{D1})$$

such that $\Omega \geq 0$,

$$\text{Tr}_{S'}(\Omega) = \mathbb{1}_S, \quad \text{Tr}((X \otimes \rho_\theta^T)\Omega) = 0,$$

$$\text{Tr}((X \otimes \partial_\theta \rho_\theta^T)\Omega) = 1.$$

The constraints on Ω guarantee any feasible Ω is contained in a convex compact set R_2 [the absolute value of each entry of Ω should not be larger than $\dim(\mathcal{H}_S)$]. We could also set a convex compact region R_1 on \mathbf{x} , so that the following optimization problem generates the same optimal value as Eq. (20):

$$\min_{(\mathbf{x}, \Omega)} \text{Tr}((X_2 \otimes \rho_\theta^T)\Omega),$$

such that $\Omega \geq 0$,

$$\text{Tr}_{S'}(\Omega) = \mathbb{1}_S, \quad \text{Tr}((X \otimes \rho_\theta^T)\Omega) = 0,$$

$$\text{Tr}((X \otimes \partial_\theta \rho_\theta^T)\Omega) = 1,$$

$$\mathbf{x} \in R_1, \quad \Omega \in R_2. \quad (\text{D2})$$

As discussed in Sec. III, this is possible when the size of R_1 is sufficiently large, in normal cases when the infimum in Eq. (20) is attainable. Otherwise, the optimal value of Eq. (D2) can still approach that of Eq. (20) for sufficiently large size of R_1 .

Here we describe the global optimization algorithm [81] for Eq. (D2) that is guaranteed to converge to the global optimum of Eq. (D2) in finite steps. One may seek [59] for a general survey on algorithms from biconvex optimization.

We first rewrite Eq. (D2) as

$$\min_{(\mathbf{x}, \Omega) \in R_1 \times R_2} f(\mathbf{x}, \Omega),$$

$$\text{such that } \Omega \geq 0, \quad \forall i, h_i(\mathbf{x}, \Omega) = 0, \quad (\text{D3})$$

where $f(\mathbf{x}, \Omega)$ is the biconvex target function and $h_i(\mathbf{x}, \Omega)$ are bi-affine functions. The global optimization algorithm finds the global optimum of Eq. (D2) by solving a set of primal problems and relaxed dual problems, which generate upper and lower bounds on the optimum, respectively. The upper and lower bounds converge to the global optimum up to a small error in finite steps. The algorithm is described as follows.

Step 1: Initialization.

Define initial upper and lower bounds (f^U, f^L) on the global optimum, where f^U and $-f^L$ can be chosen as two very large numbers. Set the counter $K = 1$. Set a convergence tolerance parameter ε . Choose a starting point \mathbf{x}^1 . Define three empty sets $\mathfrak{R}^{\text{feas}}$ (set of feasible problems), $\mathfrak{R}^{\text{infeas}}$ (set of infeasible problems), \mathfrak{S} (set of candidates of lower bound).

Step 2: Primal problem.

- (1) Consider the primal problem for $\mathbf{x} = \mathbf{x}^K$ if it is feasible (that is, if there exists some $\Omega \in R_2$ that satisfies the constraints):

$$P(\mathbf{x}^K) = \min_{\Omega \in R_2} f(\mathbf{x}^K, \Omega),$$

$$\text{such that } \Omega \geq 0, \quad \forall i, h_i(\mathbf{x}^K, \Omega) = 0. \quad (\text{D4})$$

The strong duality theorem [76] indicates that $P(\mathbf{x}^K)$ can be solved through

$$P(\mathbf{x}^K) = \max_{\mathbf{y}, Z \geq 0} \min_{\Omega \in R_2} L(\mathbf{x}^K, \Omega, \mathbf{y}, Z), \quad (\text{D5})$$

where the Lagrange function

$$L(\mathbf{x}, \Omega, \mathbf{y}, Z) := f(\mathbf{x}, \Omega) + \sum_i y_i h_i(\mathbf{x}, \Omega) - \text{Tr}(\Omega Z), \quad (\text{D6})$$

Z is a semidefinite positive matrix acting on $\mathcal{H}_{S'} \otimes \mathcal{H}_S$ and \mathbf{y} is a vector of real numbers.

Solve Eq. (D5) and store the optimal values $(\Omega^K, \mathbf{y}^K, Z^K)$. Set $f^U = \min\{f^U, P(\mathbf{x}^K)\}$ and $\mathfrak{R}^{\text{feas}} = \mathfrak{R}^{\text{feas}} \cup \{K\}$.

- (2) If Eq. (D4) is infeasible, solve the relaxed primal problem for $\mathbf{x} = \mathbf{x}^K$ instead:

$$\delta(\mathbf{x}^K) = \min_{\Omega \in R_2, \alpha \geq 0} \alpha,$$

$$\text{such that } \Omega + \alpha \mathbb{1}_{S'S} \geq 0, \quad \forall i, h_i(\mathbf{x}^K, \Omega) = 0. \quad (\text{D7})$$

The strong duality theorem implies

$$\begin{aligned} \delta(\mathbf{x}^K) &= \max_{\mathbf{y}, Z \geq 0} \min_{\Omega \in R_2, \alpha \geq 0} \alpha + \sum_i y_i h_i(\mathbf{x}, \Omega) \\ &\quad - \text{Tr}((\Omega + \alpha \mathbb{1}_{S'S})Z) \\ &= \max_{\mathbf{y}, Z \geq 0, \text{Tr}(Z) \leq 1} \min_{\Omega \in R_2} L_1(\mathbf{x}^K, \Omega, \mathbf{y}, Z), \end{aligned} \quad (\text{D8})$$

where the Lagrange function $L_1(\mathbf{x}, \Omega, \mathbf{y}, Z) := \sum_i y_i h_i(\mathbf{x}, \Omega) - \text{Tr}(\Omega Z)$.

Solve Eq. (D8) and store the optimal values $(\Omega^K, \mathbf{y}^K, Z^K)$. Let $\mathfrak{R}^{\text{infeas}} = \mathfrak{R}^{\text{infeas}} \cup \{K\}$.

Step 3: Determine the current region of \mathbf{x} .

Suppose Ω is parameterized by a vector of real numbers Ω_i . Since Ω is contained in a compact set, Ω_i has upper and lower bounds that we denote by Ω_i^U and Ω_i^L . Consider the partial derivatives of the Lagrange functions defined by $g_i^k(\mathbf{x}) := \partial/\partial\Omega_i L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}$ for $k \in \mathfrak{R}^{\text{feas}}$ and $g_i^k(\mathbf{x}) := \partial/\partial\Omega_i L_1(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}$ for $k \in \mathfrak{R}^{\text{infeas}}$. Define the set of indices for connected variables $I_k := \{i | g_i^k(\mathbf{x}) \text{ is a nontrivial function of } \mathbf{x}\} = \{i | g_i^k(\mathbf{x}) = 0, \forall \mathbf{x}\}$ (the last equality follows from the KKT conditions [76]) and Ω_i is called a connected variable of the Lagrange functions if $i \in I_k$. We can also define the linearized Lagrange functions $L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}} := L(\mathbf{x}, \Omega^k, \mathbf{y}^k, Z^k) + \sum_{i \in I_k} g_i^k(\mathbf{x})(\Omega_i - \Omega_i^k)$ and $L_1(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}} := L_1(\mathbf{x}, \Omega^k, \mathbf{y}^k, Z^k) + \sum_{i \in I_k} g_i^k(\mathbf{x})(\Omega_i - \Omega_i^k)$. The linearized functions $L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}$ and $L_1(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}$ are functions of the connected variables only and independent of Ω_i if $i \notin I_k$.

Let $\mathfrak{B}^k := \prod_{i \in I_k} \{\Omega_i^L, \Omega_i^U\}$ be the set of combinations of upper and lower bounds on Ω_i for all $i \in I_k$. We abuse the notation a bit and use $\Omega \in \mathfrak{B}^k$ to denote the case where the part of connected variables Ω_{I_k} in Ω is contained in \mathfrak{B}^k and the other part is arbitrary. We will see that the other part is irrelevant in our calculations and can be ignored. In this sense, there are in total $2^{|I_k|}$ number of $\Omega \in \mathfrak{B}^k$, which is finite. We also define $R(k, \Omega)$ to be a region of \mathbf{x} as a function of $\Omega \in \mathfrak{B}^k$ defined by

$$R(k, \Omega) := \{\mathbf{x} | \forall i \in I_k, g_i^k(\mathbf{x}) \leq_{\Omega_i} 0\}, \quad (\text{D9})$$

where “ \leq_{Ω_i} ” represents “ \leq ” if $\Omega_i = \Omega_i^U$, and “ \geq ” if $\Omega_i = \Omega_i^L$.

Let $\mathfrak{B}^{(k,K)} = \{\Omega \in \mathfrak{B}^k | \mathbf{x}^K \in R(k, \Omega)\}$. The relaxed dual problem in the next step will be solved in the region of \mathbf{x} that is contained in $\bigcap_{k=1}^{K-1} \bigcap_{\Omega \in \mathfrak{B}^{(k,K)}} R(k, \Omega)$.

Step 4: Relaxed dual problem.

Determine the set of indices for connected variables I_K . Note that for any k , $L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}$ is a function of the connected variables only and is fixed if the connected variables Ω_{I_k} of Ω is fixed. Therefore, we will also write $L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}} = L(\mathbf{x}, \Omega_{I_k}, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}$.

For each $\Omega_\star \in \mathfrak{B}^K = \prod_{i \in I_K} \{\Omega_i^L, \Omega_i^U\}$ (there are $2^{|I_K|}$ different Ω_\star in total), solve the following relaxed dual problem:

$$\begin{aligned}
& \min_{\mathbf{x} \in \mathcal{R}_1, \mu} \mu \\
\text{such that} & \left. \begin{aligned} \mu &\geq L(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}, \\ \mathbf{x} &\in R(k, \Omega), \end{aligned} \right\} \forall \Omega \in \mathfrak{B}^{(k,K)}, 1 \leq k \leq K-1, k \in \mathcal{R}^{\text{feas}}, \\
& \left. \begin{aligned} 0 &\geq L_1(\mathbf{x}, \Omega, \mathbf{y}^k, Z^k)|_{\Omega^k}^{\text{lin}}, \\ \mathbf{x} &\in R(k, \Omega), \end{aligned} \right\} \forall \Omega \in \mathfrak{B}^{(k,K)}, 1 \leq k \leq K-1, k \in \mathcal{R}^{\text{infeas}}, \\
& \left. \begin{aligned} \mu &\geq L(\mathbf{x}, \Omega_*, \mathbf{y}^K, Z^K)|_{\Omega^K}^{\text{lin}}, \\ \mathbf{x} &\in R(k, \Omega_*), \end{aligned} \right\} K \in \mathcal{R}^{\text{feas}}, \\
& \left. \begin{aligned} 0 &\geq L_1(\mathbf{x}, \Omega_*, \mathbf{y}^K, Z^K)|_{\Omega^K}^{\text{lin}}, \\ \mathbf{x} &\in R(k, \Omega_*), \end{aligned} \right\} K \in \mathcal{R}^{\text{infeas}}. \tag{D10}
\end{aligned}$$

For each Ω_* , store the solution (μ_*, \mathbf{x}_*) of Eq. (D10) in \mathfrak{S} .

Step 5: **Select a new lower bound and determine \mathbf{x}^{K+1} .**

From the set \mathfrak{S} , select the minimum μ^{\min} and the corresponding \mathbf{x}^{\min} . Set $f^L = \mu^{\min}$ and $\mathbf{x}^{K+1} = \mathbf{x}^{\min}$. Delete $(\mu^{\min}, \mathbf{x}^{\min})$ from the set of candidates of lower bound \mathfrak{S} .

Step 6: **Check for convergence.**

Check if $f^L > f^U - \varepsilon$, if yes, STOP; otherwise, set $K = K + 1$ and return to Step 2.

The global optimization algorithm described above works in a branch-and-bound way where \mathbf{x} is partitioned into different regions and different candidates of lower bounds of the global optimum are explored in each iteration. The subproblems that are solved in each iteration are semidefinite programs [Eqs. (D5) and (D8)] and quadratically constrained quadratic programs [Eq. (D10)], which can be solved efficiently (for a moderate system dimension) using algorithms for convex optimization [76]. The running time of the entire algorithm depends largely on the number of subproblems that are solved in each iteration,

which is exponential in the number of connected variables. Methods that can reduce this complexity were discussed in Ref. [81].

APPENDIX E: BINARY MEASUREMENTS ON PURE STATES

1. Measurement on a qubit

Here consider a binary measurement on a single qubit where $X = x_1 M_1 + x_2 M_2$, $M_1 = M$ and $M_2 = \mathbb{1} - M$. Without loss of generality, we assume

$$M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad m_1, m_2 \in [0, 1] \text{ and } m_1 > m_2. \tag{E1}$$

Our goal is to find (\mathbf{x}, U) such that the two necessary conditions in Lemma 4 are satisfied.

Let $|\phi\rangle = \sqrt{p}|1\rangle + \sqrt{1-p}|2\rangle$ where $0 \leq p \leq 1$ (any additional phases in the amplitudes of $|\phi\rangle$ do not change the results). Condition (1) translates into

$$(\sqrt{p} \quad \sqrt{1-p}) \begin{pmatrix} x_1 m_1 + x_2(1-m_1) & 0 \\ 0 & x_1 m_2 + x_2(1-m_2) \end{pmatrix} \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = 0, \tag{E2}$$

$$(\sqrt{p} \quad \sqrt{1-p}) \begin{pmatrix} (x_1 m_1 + x_2(1-m_1))^2 & 0 \\ 0 & (x_1 m_2 + x_2(1-m_2))^2 \end{pmatrix} \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = \frac{1}{4n}, \tag{E3}$$

and Condition (2) is trivially true when $p = 0, 1$, otherwise translates into

$$\frac{\langle \phi | X^2 | \phi \rangle}{\langle \phi | X_2 | \phi \rangle} = \frac{(x_1 m_1 + x_2(1-m_1))^2}{x_1^2 m_1 + x_2^2(1-m_1)} = \frac{(x_1 m_2 + x_2(1-m_2))^2}{x_1^2 m_2 + x_2^2(1-m_2)} = \gamma(\{M_i\}) \leq 1, \tag{E4}$$

where we use

$$F^{\mathbb{U}}(\psi_\theta, \{M_i\})^{-1} = \langle \phi | X_2 | \phi \rangle = \frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} \cdot \langle \phi | X^2 | \phi \rangle = (\gamma(\{M_i\}) J(\psi_\theta))^{-1}, \tag{E5}$$

and $J(\psi_\theta) = 4n$.

Solving the system of equations [Eqs. (E2)–(E5)], we find the following results.

- (1) When $1 > m_1 > m_2 > 0$, according to Lemma 15, the QPFI is attainable. Moreover, the only solution satisfying the necessary conditions [Eqs. (E2)–(E5)] is

$$\frac{x_2}{x_1} = -\sqrt{\frac{m_1 m_2}{(1 - m_1)(1 - m_2)}}, \quad (\text{E6})$$

$$p = \frac{\sqrt{m_2(1 - m_2)}}{\sqrt{m_1(1 - m_1)} + \sqrt{m_2(1 - m_2)}}, \quad (\text{E7})$$

which must be the optimal solution. It gives

$$\gamma(\{M_i\}) = 1 - \left(\sqrt{m_1 m_2} + \sqrt{(1 - m_1)(1 - m_2)}\right)^2. \quad (\text{E8})$$

- (2) When $1 = m_1 > m_2 = 0$, we must have $X^2 = X_2$, implying $\gamma(\{M_i\}) = 1$. The QFI is achievable as long as a solution to $px_1 + (1 - p)x_2 = 0$ and $px_1^2 + (1 - p)x_2^2 = 1/(4n)$ exists, which means that any $0 < p < 1$ is optimal.
- (3) When $1 > m_1 > m_2 = 0$, the necessary conditions [Eqs. (E2)–(E5)] have no solutions. Thus, this is a singular case where the QPFI is not attainable. And we have from Theorem 1 that

$$\gamma(\{M_i\}) = \lim_{\epsilon \rightarrow 0^+} \gamma(\{M_i^{(\epsilon)}\}) \quad (\text{E9})$$

$$= \lim_{\epsilon \rightarrow 0^+} \left(1 - \frac{\epsilon}{2}\right) m_1 + \frac{\epsilon}{2} - \left(1 - \frac{\epsilon}{2}\right) m_1 \epsilon - 2\sqrt{\frac{\epsilon}{2} \left(1 - \frac{\epsilon}{2}\right) \left(1 - \frac{\epsilon}{2}\right) m_1 \left(1 - \left(1 - \frac{\epsilon}{2}\right) m_1\right)} = m_1. \quad (\text{E10})$$

2. Measurement on a qudit

Now consider a d -dimensional system with $d > 2$ and a binary measurement $M_1 = M$ and $M_2 = \mathbb{1} - M$ on it where

$$M = \begin{pmatrix} m_1 & & & \\ & m_2 & & \\ & & \ddots & \\ & & & m_d \end{pmatrix}, \quad (\text{E11})$$

and $1 > m_1 \geq \dots \geq m_d > 0$.

Let $|\phi\rangle = \sum_{i=1}^d \phi_i |i\rangle$. We now show that the support of $|\phi\rangle$: $\text{supp}\{|\phi\rangle\} = \{i : \phi_i \neq 0\}$ must correspond to at most two different values of m_i when $|\phi\rangle$ is optimal. We prove this by contradiction. Without loss of generality, assume $|\phi_{1,2,3}| > 0$ and $m_1 > m_2 > m_3$. Condition (2) implies that

$$\frac{\langle 1|X^2|1\rangle}{\langle 1|X_2|1\rangle} = \frac{\langle 2|X^2|2\rangle}{\langle 2|X_2|2\rangle} = \frac{\langle 3|X^2|3\rangle}{\langle 3|X_2|3\rangle}, \quad (\text{E12})$$

$$\Rightarrow \frac{(x_1 m_1 + x_2(1 - m_1))^2}{x_1^2 m_1 + x_2^2(1 - m_1)} = \frac{(x_1 m_2 + x_2(1 - m_2))^2}{x_1^2 m_2 + x_2^2(1 - m_2)} = \frac{(x_1 m_3 + x_2(1 - m_3))^2}{x_1^2 m_3 + x_2^2(1 - m_3)}, \quad (\text{E13})$$

$$\Rightarrow -\sqrt{\frac{m_1 m_2}{(1 - m_1)(1 - m_2)}} = -\sqrt{\frac{m_1 m_3}{(1 - m_1)(1 - m_3)}} = -\sqrt{\frac{m_2 m_3}{(1 - m_2)(1 - m_3)}}, \quad (\text{E14})$$

which contradicts $m_1 > m_2 > m_3$. Thus, we conclude that the support of $|\phi\rangle$ must correspond to at most two different values of m_i .

Therefore, we have

$$\gamma(\{M_i\}) = \max_{1 \leq k \leq l \leq d} \gamma_{kl}(\{M_i\}) = \max_{kl} 1 - \left(\sqrt{m_k m_l} + \sqrt{(1 - m_k)(1 - m_l)}\right)^2. \quad (\text{E15})$$

In fact, assume $m_1 \geq m_2 \geq \dots \geq m_d$, we must have

$$\gamma(\{M_i\}) = 1 - \left(\sqrt{m_1 m_d} + \sqrt{(1 - m_1)(1 - m_d)}\right)^2. \quad (\text{E16})$$

The reason is that when $m_k \geq m_l$, increasing m_k or decreasing m_l while the other element is fixed will only increase $\gamma_{kl}(\{M_i\})$. We see that by computing the derivative of $\gamma_{kl}(\{M_i\})$ with respect to m_k . We have

$$\frac{\partial}{\partial m_k} \gamma_{kl}(\{M_i\}) = 1 - 2m_l - \frac{\sqrt{(1-m_l)m_l}}{\sqrt{(1-m_k)m_k}}(1-2m_k) \geq 0, \quad (\text{E17})$$

when $m_k \geq m_l$ because $1 - 2m_l/\sqrt{(1-m_l)m_l} = \sqrt{1-m_l/m_l} - \sqrt{m_l/1-m_l} \geq 1 - 2m_k/\sqrt{(1-m_k)m_k}$.

APPENDIX F: COMMUTING-OPERATOR MEASUREMENTS ON PURE STATES

We take one step further from binary measurements and consider the commuting-operator measurements where

$$M_i = \begin{pmatrix} m_1^{(i)} & & & \\ & m_2^{(i)} & & \\ & & \ddots & \\ & & & m_d^{(i)} \end{pmatrix}, \quad (\text{F1})$$

and $\sum_i M_i = \mathbb{1}$. We also assume $m_j^{(i)} > 0$ for all i, j .

1. Proof of Theorem 6

We first prove Theorem 6:

Theorem 6.—For commuting-operator measurements [Eq. (63)], there always exists an optimal solution to $(|\phi\rangle, |\phi^\perp\rangle)$ such that $|\phi\rangle = \sqrt{p}|k\rangle + \sqrt{1-p}|l\rangle$ and $|\phi^\perp\rangle = \sqrt{1-p}|k\rangle - \sqrt{p}|l\rangle$ for two basis states $|k\rangle$ and $|l\rangle$ and $0 < p < 1$.

Proof.—Assume $(\mathbf{x}, |\phi\rangle)$ satisfies Condition (2) in Lemma 4, where we write $x_i = y(x + a_i)$, $\mathbf{x} \neq 0$ and the support of $|\phi\rangle$ contains $|1\rangle$ and $|2\rangle$. Then according to Condition (2), we must have

$$\begin{aligned} \frac{(x + \langle a \rangle_1)^2}{x^2 + 2 \langle a \rangle_1 x + \langle a^2 \rangle_1} &= \frac{(x + \langle a \rangle_2)^2}{x^2 + 2 \langle a \rangle_2 x + \langle a^2 \rangle_2} \\ &= \frac{\langle \phi | X^2 | \phi \rangle}{\langle \phi | X_2 | \phi \rangle}, \end{aligned} \quad (\text{F2})$$

where $\langle a \rangle_j = \sum_i a_i m_j^{(i)}$ and $\langle a^2 \rangle_j = \sum_i a_i^2 m_j^{(i)}$. We also define $\langle \Delta a^2 \rangle_j = \langle a^2 \rangle_j - \langle a \rangle_j^2$.

Assume $\langle a \rangle_1 > \langle a \rangle_2$, we have the following possible solutions of Eq. (F2).

(1) When $\langle \Delta a^2 \rangle_1 = \langle \Delta a^2 \rangle_2$,

$$x = -\frac{1}{2}(\langle a \rangle_1 + \langle a \rangle_2), \text{ and } \frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left(\frac{\sqrt{\langle \Delta a^2 \rangle_1} + \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2. \quad (\text{F3})$$

(2) When $\langle \Delta a^2 \rangle_1 - \langle \Delta a^2 \rangle_2 \neq 0$, we have either

$$x = \frac{-\langle a \rangle_2 \langle \Delta a^2 \rangle_1 + \langle a \rangle_1 \langle \Delta a^2 \rangle_2 - (\langle a \rangle_1 - \langle a \rangle_2) \sqrt{\langle \Delta a^2 \rangle_1 \langle \Delta a^2 \rangle_2}}{\langle \Delta a^2 \rangle_1 - \langle \Delta a^2 \rangle_2} \quad (\text{F4})$$

and

$$\frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left(\frac{\sqrt{\langle \Delta a^2 \rangle_1} + \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2, \quad (\text{F5})$$

or

$$x = \frac{-\langle a \rangle_2 \langle \Delta a^2 \rangle_1 + \langle a \rangle_1 \langle \Delta a^2 \rangle_2 + (\langle a \rangle_1 - \langle a \rangle_2) \sqrt{\langle \Delta a^2 \rangle_1 \langle \Delta a^2 \rangle_2}}{\langle \Delta a^2 \rangle_1 - \langle \Delta a^2 \rangle_2} \quad (\text{F6})$$

and

$$\frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left(\frac{\sqrt{\langle \Delta a^2 \rangle_1} - \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2. \quad (\text{F7})$$

Next we show that there always is an optimal solution such that its support contains only two elements. Without loss of generality, assume $|\phi^*\rangle = \sum_{i=1}^d \phi_i^* |i\rangle$ is an optimal solution (which is guaranteed to exist thanks to Lemma 15). The corresponding error vector \mathbf{x}^* is written as $x_i^* = y^*(x^* + a_i^*)$. We have from Condition (1) in Lemma 4 that $\sum_{i=1}^d |\phi_i^*|^2 y^*(x^* + \langle a^* \rangle_i) = 0$ and $\sum_{i=1}^d |\phi_i^*|^2 (y^*(x^* + \langle a^* \rangle_i))^2 = 1/(4n)$. Clearly, we must have some $i \neq j$, such that $|\phi_{ij}^*| > 0$, $\langle a^* \rangle_i \neq \langle a^* \rangle_j$ and $(x^* + \langle a^* \rangle_i)(x^* + \langle a^* \rangle_j) < 0$. Without loss of generality, we assume $i = 1, j = 2$ and $\langle a^* \rangle_1 > \langle a^* \rangle_2$. Then $(\mathbf{x}^*, |\phi^*\rangle)$ must satisfy either Eqs. (F3) or (F4) and (F5) from the previous discussion. [Note that if Eqs. (F6) and (F7) cannot be true because $(x^* + \langle a^* \rangle_i)(x^* + \langle a^* \rangle_j) < 0$.]

We then assert that $|\phi^{**}\rangle = \sqrt{p} |1\rangle + \sqrt{1-p} |2\rangle$ ($0 \leq p \leq 1$) is also an optimal solution, when p satisfies

$$\begin{aligned} & (\sqrt{p} \quad \sqrt{1-p}) \begin{pmatrix} x^* + \langle a^* \rangle_1 & 0 \\ 0 & x^* + \langle a^* \rangle_2 \end{pmatrix} \\ & \times \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = 0. \end{aligned} \quad (\text{F8})$$

Using Eqs. (F3)–(F5), it is easy to see that

$$p = \frac{\sqrt{\langle \Delta(a^*)^2 \rangle_2}}{\sqrt{\langle \Delta(a^*)^2 \rangle_2} + \sqrt{\langle \Delta(a^*)^2 \rangle_1}}. \quad (\text{F9})$$

Let $(\mathbf{x}^*, |\phi^*\rangle)$ be an optimal solution when $|\phi\rangle, |\phi^\perp\rangle$ are restricted in $\text{span}\{|k\rangle, |l\rangle\}$ and $|\phi^*\rangle = \sqrt{p_{kl}} |k\rangle + \sqrt{1-p_{kl}} |l\rangle$ (we also assume $\langle a^* \rangle_k > \langle a^* \rangle_l$). Using Eq. (17), we see that the optimal a_i^*

$$\begin{aligned} y^*(x^* + a_i^*) &= x_i^* = \frac{y^*}{\gamma_{kl}(\{M_i\})} \frac{\langle \phi^* | M_i X^* | \phi^* \rangle}{\langle \phi^* | M_i | \phi^* \rangle} = \frac{y^*}{\gamma_{kl}(\{M_i\})} \frac{p_{kl} m_k^{(i)} (x^* + \langle a^* \rangle_k) + (1-p_{kl}) m_l^{(i)} (x^* + \langle a^* \rangle_l)}{p_{kl} m_k^{(i)} + (1-p_{kl}) m_l^{(i)}} \\ &= \frac{p_{kl} y^* (x^* + \langle a^* \rangle_k)}{\gamma_{kl}(\{M_i\})} \frac{m_k^{(i)} - m_l^{(i)}}{p_{kl} m_k^{(i)} + (1-p_{kl}) m_l^{(i)}}, \end{aligned} \quad (\text{F12})$$

where we use

$$\langle \phi^* | X^* | \phi^* \rangle = p_{kl} y^* (x^* + \langle a^* \rangle_k) + (1-p_{kl}) y^* (x^* + \langle a^* \rangle_l) = 0, \quad (\text{F13})$$

in the last step. From Eq. (F12), we have,

$$(y^*)^2 ((x^*)^2 + 2 \langle a^* \rangle_k x^* + \langle (a^*)^2 \rangle_k) = \left(\frac{p_{kl} y^* (x^* + \langle a^* \rangle_k)}{\gamma_{kl}(\{M_i\})} \right)^2 \sum_i \frac{(m_k^{(i)} - m_l^{(i)})^2 m_k^{(i)}}{(p_{kl} m_k^{(i)} + (1-p_{kl}) m_l^{(i)})^2}, \quad (\text{F14})$$

$$(y^*)^2 ((x^*)^2 + 2 \langle a^* \rangle_l x^* + \langle (a^*)^2 \rangle_l) = \left(\frac{p_{kl} y^* (x^* + \langle a^* \rangle_k)}{\gamma_{kl}(\{M_i\})} \right)^2 \sum_i \frac{(m_k^{(i)} - m_l^{(i)})^2 m_l^{(i)}}{(p_{kl} m_k^{(i)} + (1-p_{kl}) m_l^{(i)})^2}. \quad (\text{F15})$$

According to Condition (2),

$$\frac{(x^*)^2 + 2 \langle a^* \rangle_k x^* + \langle (a^*)^2 \rangle_k}{(x^* + \langle a^* \rangle_k)^2} = \frac{(x^*)^2 + 2 \langle a^* \rangle_l x^* + \langle (a^*)^2 \rangle_l}{(x^* + \langle a^* \rangle_l)^2}. \quad (\text{F16})$$

We take $x_i^{**} = y^{**}(x^* + a_i^*)$, where y^{**} is solved from

$$\begin{aligned} & (\sqrt{p} \quad \sqrt{1-p}) \begin{pmatrix} (x^* + \langle a^* \rangle_1)^2 & 0 \\ 0 & (x^* + \langle a^* \rangle_2)^2 \end{pmatrix} \\ & \times \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = \frac{1}{4n(y^{**})^2}. \end{aligned} \quad (\text{F10})$$

Both equations are derived from Condition (1). Now we have a new solution $(\mathbf{x}^{**}, |\phi^{**}\rangle)$ such that y^{**} and $|\phi^{**}\rangle$ are solved by the equations above. Note that we still let $a_i^{**} = a_i^*$ and $x^{**} = x^*$. The new solution have the same FI as the original, i.e., $\langle \phi | X_2 | \phi \rangle$ does not change, because $\langle \phi | X_2 | \phi \rangle / \langle \phi | X^2 | \phi \rangle$ is independent of y [due to Eqs. (F3) and (F5)] and $\langle \phi | X^2 | \phi \rangle = 1/(4n)$ is invariant. The new solution is thus supported on a two-dimensional subspace spanned by $\{|1\rangle, |2\rangle\}$, proving Theorem 6. ■

2. Optimal solution for commuting-operator measurements

Now we proceed to compute general $\gamma(\{M_i\})$ for commuting-operator measurements. First, consider the optimization for measurements restricted in a two-dimensional subspace spanned by $|k\rangle, |l\rangle$ for some $k \neq l$, i.e.,

$$(M_i)_{kl} = m_k^{(i)} |k\rangle \langle k| + m_l^{(i)} |l\rangle \langle l|, \quad (\text{F11})$$

and $\sum_i (M_i)_{kl} = \mathbb{1}_{\text{span}\{|k\rangle, |l\rangle\}}$.

From Eqs. (F13)–(F16), we have

$$p_{kl}^2 \left(\sum_i \frac{(m_k^{(i)} - m_l^{(i)})^2 m_k^{(i)}}{(p_{kl} m_k^{(i)} + (1 - p_{kl}) m_l^{(i)})^2} \right) = (1 - p_{kl})^2 \left(\sum_i \frac{(m_k^{(i)} - m_l^{(i)})^2 m_l^{(i)}}{(p_{kl} m_k^{(i)} + (1 - p_{kl}) m_l^{(i)})^2} \right), \quad (\text{F17})$$

It will give us a unique solution to p_{kl} because the left-hand side is a monotonically increasing function in $[0, \sum_i (m_k^{(i)} - m_l^{(i)})^2]$ of $p_{kl} \in [0, 1]$ and the right-hand side is a monotonically decreasing function in $[0, \sum_i (m_k^{(i)} - m_l^{(i)})^2]$ of $p_{kl} \in [0, 1]$. However, a simple analytical solution to p_{kl} from Eq. (F17) might not exist because it is a root of a high degree polynomial. Then we have

$$\gamma_{kl}(\{M_i\}) = \sum_i \frac{(\text{Re}[\langle \phi^* | M_i | \phi^{\perp*} \rangle])^2}{\langle \phi^* | M_i | \phi^* \rangle} = \sum_i \frac{p_{kl}^* (1 - p_{kl}^*) (m_k^{(i)} - m_l^{(i)})^2}{p_{kl}^* m_k^{(i)} + (1 - p_{kl}^*) m_l^{(i)}}, \quad (\text{F18})$$

where p_{kl}^* is the unique solution to Eq. (F17).

Finally,

$$\gamma(\{M_i\}) = \max_{kl} \gamma_{kl}(\{M_i\}), \quad (\text{F19})$$

using Theorem 6. Note that although Eq. (F17) might only be solvable numerically in practice for a multiple-outcome measurement. Our solution for pure states and commuting-operator measurements still has a huge simplification compared to the original biconvex problem for general states and measurements.

3. Proof of Theorem 7

Here we prove a simple upper bound on the normalized QPFI:

Theorem 7.—For commuting-operator measurements [Eq. (63)], the normalized QPFI $\gamma(\{M_i\})$ satisfies

$$\gamma(\{M_i\}) \leq 1 - \min_{kl} \left(\sum_i \sqrt{m_k^{(i)} m_l^{(i)}} \right)^2. \quad (\text{F70})$$

When there exists a (k, l) that minimizes $\sum_i \sqrt{m_k^{(i)} m_l^{(i)}}$ such that the set $\{m_k^{(i)}/m_l^{(i)}, 1 \leq i \leq r\}$ contains at most two elements, the inequality is tight.

Proof.—From the discussions in Appendices F 1 and F 2, we have that

$$\gamma(\{M_i\}) = \max_{kl} \left(\max_{0 \leq p \leq 1} \sum_i \frac{p(1-p)(m_k^{(i)} - m_l^{(i)})^2}{p m_k^{(i)} + (1-p)m_l^{(i)}} \right) = \max_{kl} \sum_i \frac{p_{kl}^* (1 - p_{kl}^*) (m_k^{(i)} - m_l^{(i)})^2}{p_{kl}^* m_k^{(i)} + (1 - p_{kl}^*) m_l^{(i)}}. \quad (\text{F20})$$

Equation (70) is then proven, noting that for any $p \in [0, 1]$,

$$\sum_i \frac{p(1-p)(m_k^{(i)} - m_l^{(i)})^2}{p m_k^{(i)} + (1-p)m_l^{(i)}} - 1 \quad (\text{F21})$$

$$= \sum_i \frac{p(1-p)(m_k^{(i)} - m_l^{(i)})^2 - (p m_k^{(i)} + (1-p)m_l^{(i)})^2}{p m_k^{(i)} + (1-p)m_l^{(i)}} \quad (\text{F22})$$

$$= \sum_i \frac{(1-2p)(p m_k^{(i)2} - (1-p)m_l^{(i)2}) - 4p(1-p)m_k^{(i)} m_l^{(i)}}{p m_k^{(i)} + (1-p)m_l^{(i)}} \quad (\text{F23})$$

$$= \sum_i \frac{(1-2p) \left((p m_k^{(i)} + (1-p)m_l^{(i)})(m_k^{(i)} - m_l^{(i)}) - (2p-1)m_k^{(i)} m_l^{(i)} \right) - 4p(1-p)m_k^{(i)} m_l^{(i)}}{p m_k^{(i)} + (1-p)m_l^{(i)}} \quad (\text{F24})$$

$$= - \left(\sum_i \frac{m_k^{(i)} m_l^{(i)}}{p m_k^{(i)} + (1-p) m_l^{(i)}} \right) \left(\sum_i p m_k^{(i)} + (1-p) m_l^{(i)} \right) \leq - \left(\sum_i \sqrt{m_k^{(i)} m_l^{(i)}} \right)^2, \quad (\text{F25})$$

where in the first equality we use $\sum_i p m_k^{(i)} + (1-p) m_l^{(i)} = 1$, in the last equality we multiply the expression by a factor of $1 = \sum_i p m_k^{(i)} + (1-p) m_l^{(i)}$, and the last inequality follows from Cauchy-Schwarz.

Assume (k, l) minimizes $\sum_i \sqrt{m_k^{(i)} m_l^{(i)}}$. Then the equality above holds when $\exists p$, such that for any i, j ,

$$\frac{m_k^{(i)} m_l^{(i)}}{(p m_k^{(i)} + (1-p) m_l^{(i)})^2} = \frac{m_k^{(j)} m_l^{(j)}}{(p m_k^{(j)} + (1-p) m_l^{(j)})^2}. \quad (\text{F26})$$

When there are at most two different i and j (i.e., $r = 2$), such a p always exists, and the upper bound is tight [which also follows directly from Eq. (62)]. In general, when the set

$$\left\{ \frac{m_k^{(i)}}{m_l^{(i)}}, 1 \leq i \leq r \right\} \quad (\text{F27})$$

contains at most two distinct elements, the upper bound is tight and the optimal preprocessed state can be chosen as

$$|\phi^*\rangle = \sqrt{p_{kl}^*} |k\rangle + \sqrt{1-p_{kl}^*} |l\rangle, \quad |\phi^{\perp*}\rangle = \sqrt{1-p_{kl}^*} |1\rangle - \sqrt{p_{kl}^*} |l\rangle, \quad (\text{F28})$$

where $\{m_k^{(i)}/m_l^{(i)}, 1 \leq i \leq r\} = \{m_{kl,1}, m_{kl,2}\}$ and $p_{kl}^* = 1/1 + \sqrt{m_{kl,1} m_{kl,2}}$. Note that in this case $\{m_k^{(i)}/m_l^{(i)}, 1 \leq i \leq r\}$ must contain at least two distinct elements—otherwise, $\{M_i\}$ must be trivial.

Alternatively, we can also prove Eq. (70) directly from its original definition [Eq. (42)] without using Theorem 6. We have

$$\gamma(\{M_i\}) = \max_{\substack{\langle \phi | \phi \rangle = 1, \langle \phi^\perp | \phi^\perp \rangle = 1, \\ \langle \phi | \phi^\perp \rangle = 0}} \sum_i \frac{(\text{Re}[\langle \phi | M_i | \phi^\perp \rangle])^2}{\langle \phi | M_i | \phi \rangle} = \max_{\substack{\sum_k a_k^2 = 1, \sum_k b_k^2 \leq 1, \\ \sum_k a_k b_k = 0}} \sum_i \frac{\left(\sum_k m_k^{(i)} a_k b_k \right)^2}{\sum_k m_k^{(i)} a_k^2}, \quad (\text{F29})$$

where $a_k = \langle k | \phi \rangle$ (which we assume to be real, without loss of generality) and $b_k = \text{Re}[\langle k | \phi^\perp \rangle]$ for all $1 \leq k \leq d$. Clearly, the maximum is reached at $\sum_k b_k^2 = 1$. For any $\{a_k\}$ and $\{b_k\}$ satisfying $\sum_k a_k^2 = 1$, $\sum_k b_k^2 = 1$ and $\sum_k a_k b_k = 0$, we have

$$\sum_i \frac{\left(\sum_k m_k^{(i)} a_k b_k \right)^2}{\sum_k m_k^{(i)} a_k^2} = \sum_i \frac{\sum_k (m_k^{(i)})^2 a_k^2 b_k^2 + \sum_{k \neq l} m_k^{(i)} m_l^{(i)} a_k b_k a_l b_l}{\sum_k m_k^{(i)} a_k^2} \quad (\text{F30})$$

$$= \sum_i \frac{\left(\sum_k m_k^{(i)} a_k^2 \right) \left(\sum_l m_l^{(i)} b_l^2 \right) + \frac{1}{2} \sum_{k \neq l} m_k^{(i)} m_l^{(i)} (2 a_k b_k a_l b_l - a_k^2 b_l^2 - a_l^2 b_k^2)}{\sum_k m_k^{(i)} a_k^2} \quad (\text{F31})$$

$$= 1 - \frac{1}{2} \sum_{k \neq l} (a_k b_l - a_l b_k)^2 \sum_i \frac{m_k^{(i)} m_l^{(i)}}{\sum_k m_k^{(i)} a_k^2} \quad (\text{F32})$$

$$\leq 1 - \min_{k \neq l} \left(\sum_i \frac{m_k^{(i)} m_l^{(i)}}{\sum_k m_k^{(i)} a_k^2} \right) \left(\sum_{i,k} m_k^{(i)} a_k^2 \right) \leq 1 - \min_{kl} \left(\sum_i \sqrt{m_k^{(i)} m_l^{(i)}} \right)^2, \quad (\text{F33})$$

where we use $\sum_{i,l} m_l^{(i)} b_l^2 = 1$ in the second step, along with the identities $\sum_k a_k^2 = \sum_l b_l^2 = 1$, and $\sum_k a_k b_k = 0$ to simplify $\frac{1}{2} \sum_{k \neq l} (a_k b_l - a_l b_k)^2 = \frac{1}{2} (\sum_{kl} a_k^2 b_l^2 + a_l^2 b_k^2 - 2 a_k b_k a_l b_l) = 1$ in the third step, which allows us to minimize the remaining expression, multiplying by $1 = \sum_{i,k} m_k^{(i)} a_k^2$ to leverage the Cauchy-Schwarz inequality in the last step. Equation (70) is then proven. \blacksquare

APPENDIX G: CLASSICAL CAPACITY OF QUANTUM CHANNELS

In this section, we prove the following lemma, which is generally known to be true in quantum information theory (see, e.g., Ref. [92]):

Lemma 17.—Consider a quantum channel Φ , its classical channel capacity $C(\Phi)$ and a constant α satisfying $0 < \alpha < C(\Phi)$. Then for all but finitely many positive integers n , there exist channels Ξ_E and Ξ_D , such that

$$\|\Xi_D \circ \Phi^{\otimes n} \circ \Xi_E - \mathcal{D}_2^{\otimes \lfloor \alpha n \rfloor}\|_{\diamond} \leq e^{-\beta n},$$

for some $\beta > 0$, with \mathcal{D}_2 and $\|\cdot\|_{\diamond}$ defined in the proof of Theorem 14.

Lemma 17 essentially states that, fixing any α that is smaller than the classical channel capacity of Φ , in the large n limit, $\Phi^{\otimes n}$ with suitable encoding and decoding channels can be used to transmit classical binary information reliably at a rate α with an exponentially small error with respect to n . The proof of Lemma 17 follows almost exactly from the proof of the HSW theorem [60,61], with a slight refinement in the error analysis using Hoeffding's inequality, i.e.,

Lemma 18.—(Hoeffding's inequality [93]). Let X_1, \dots, X_n be independent random variables such that $0 \leq X_i \leq 1$ for $i = 1, \dots, n$. Then for any $\varepsilon > 0$,

$$Y_n = \frac{X_1 + \dots + X_n}{n} \quad \text{satisfies}$$

$$\Pr(|Y_n - \mathbb{E}[Y_n]| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}, \quad (\text{G1})$$

to show that the error is exponentially small. Therefore, we will only provide the error analysis part for the proof of Lemma 17 that is different from the standard proof of the HSW theorem and skip the steps that can readily be found in standard quantum information theory textbooks [68,94].

Following the proof of Theorem 8.27 in Ref. [68], it is clear that in order to prove Lemma 17, it is sufficient to prove the following lemma (which is a refinement of Theorem 8.26 in Ref. [68]):

Lemma 19.—Let $\eta = (p(a), \sigma_a)$ be an ensemble of quantum states satisfying $\sum_a p(a) = 1$, where σ_a are density operators and $a \in \Sigma$ (Σ is an alphabet whose order is equal to the square of the dimension of the system the σ_a act on). Let

$$\alpha < \chi(\eta) := H\left(\sum_{a \in \Sigma} p(a) \sigma_a\right) - \sum_{a \in \Sigma} p(a) H(\sigma_a), \quad (\text{G2})$$

where $H(\sigma_a)$ is the von Neumann entropy of σ_a and $m = \lfloor \alpha n \rfloor$. For all but finite number of n , there exists a function $f : \{0, 1\}^m \rightarrow \Sigma^n$ and a quantum measurement

$\{M_{\mathbf{b}}\}_{\mathbf{b} \in \{0,1\}^m}$ such that

$$\text{Tr}(M_{\mathbf{b}} \sigma_{f(\mathbf{b})}) > 1 - e^{-\beta n}, \quad (\text{G3})$$

for every $\mathbf{b} = b_1 \cdots b_m \in \{0, 1\}^m$, $\sigma_{f(\mathbf{b})} = \sigma_{f(\mathbf{b})_1} \otimes \cdots \otimes \sigma_{f(\mathbf{b})_m}$ and some $\beta > 0$.

Proof.—Choose a sufficiently small ε such that $\alpha < \chi(\eta) - 3\varepsilon$. Following the proof of Theorem 8.26 in Ref. [68], there exists a function $f : \{0, 1\}^m \rightarrow \Sigma^n$ and a quantum measurement $\{M_{\mathbf{b}}\}_{\mathbf{b} \in \{0,1\}^m}$ such that

$$\text{Tr}(M_{\mathbf{b}} \sigma_{f(\mathbf{b})}) > 1 - \delta, \quad (\text{G4})$$

for every $\mathbf{b} = b_1 \cdots b_m \in \{0, 1\}^m$ where

$$\delta = 4 \left(3 - 2 \text{Tr}(\Pi \sigma^{\otimes n}) - \sum_{\mathbf{a} \in \Sigma^n} p(a_1) \cdots p(a_n) \text{Tr}(\Lambda_{\mathbf{a}} \sigma_{\mathbf{a}}) \right) + 2^{m+4-n(\chi(\eta)-2\varepsilon)}. \quad (\text{G5})$$

Here $\mathbf{a} = a_1 \cdots a_n \in \Sigma^n$, $\sigma = \sum_{a \in \Sigma} p(a) \sigma_a$, $\sigma_{\mathbf{a}} = \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n}$, Π is the projection onto the ε -typical subspace with respect to σ , $\Lambda_{\mathbf{a}}$ is the projection onto the ε -typical subspace conditioned on $\mathbf{a} = a_1 \cdots a_n$. Specifically, let $\sigma = \sum_a p'(a) |u_a\rangle \langle u_a|$ where $\{|u_a\rangle, a \in \Sigma\}$ is an orthonormal basis and $p(a) \sigma_a = \sum_{c \in \Gamma} p(a, c) |u_{ac}\rangle \langle u_{ac}|$ where $\{|u_{ac}\rangle, a \in \Gamma\}$ is an orthonormal basis for each $a \in \Sigma$. Let $p(a) = \sum_{c \in \Gamma} p(a, c)$ and $H(p(a))$ be the Shannon entropy of $p(a)$. Then the definitions of Π and $\Lambda_{\mathbf{a}}$ are

$$\Pi = \sum_{\mathbf{a} \in T_{\varepsilon}} |u_{a_1}\rangle \langle u_{a_1}| \otimes \cdots \otimes |u_{a_n}\rangle \langle u_{a_n}|, \quad (\text{G6})$$

$$\Lambda_{\mathbf{a}} = \sum_{\mathbf{c} \in K_{\mathbf{a}, \varepsilon}} |u_{a_1 c_1}\rangle \langle u_{a_1 c_1}| \otimes \cdots \otimes |u_{a_n c_n}\rangle \langle u_{a_n c_n}|, \quad (\text{G7})$$

where T_{ε} is the set of \mathbf{a} satisfying $2^{-n(H(p'(a))+\varepsilon)} < p'(a_1) \cdots p'(a_n) < 2^{-n(H(p'(a))-\varepsilon)}$ and $K_{\mathbf{a}, \varepsilon}$ is the set of \mathbf{c} satisfying $2^{-n(H(p(a,c))-H(p(a))+\varepsilon)} < p(a_1, c_1) \cdots p(a_n, c_n) / p(a_1) \cdots p(a_n) < 2^{-n(H(p(a,c))-H(p(a))-\varepsilon)}$ for any \mathbf{a} satisfying $p(a_1) \cdots p(a_n) > 0$. We have

$$\text{Tr}(\Pi \sigma^{\otimes n}) = \sum_{\mathbf{a} \in T_{\varepsilon}} p'(a_1) \cdots p'(a_n), \quad (\text{G8})$$

$$\sum_{\mathbf{a} \in \Sigma^n} \text{Tr}(\Lambda_{\mathbf{a}} \sigma_{\mathbf{a}}) = \sum_{\mathbf{a} \in \Sigma^n} \sum_{\mathbf{c} \in K_{\mathbf{a}, \varepsilon}} p(a_1, c_1) \cdots p(a_n, c_n). \quad (\text{G9})$$

Define two random variables $X : \Sigma \rightarrow [0, x_{\text{upp}}]$ and $Y : \Sigma \times \Gamma \rightarrow [0, y_{\text{upp}}]$, where $x_{\text{upp}} = \max_{a: p'(a) \neq 0} -\log(p'(a))$

and $y_{\text{upp}} = \max_{a,b:p(a,c) \neq 0} -\log(p(a,c)) + \log(p(a))$, as

$$X(a) = -\log(p'(a)) \text{ if } p'(a) > 0, \text{ and } 0 \text{ otherwise,} \quad (\text{G10})$$

$$Y(a,b) = -\log(p(a,c)) + \log(p(a)) \text{ if } p(a,c) > 0, \text{ and } 0 \text{ otherwise.} \quad (\text{G11})$$

Let X_1, \dots, X_n be n independent random variables each identically distributed to X . Using the Hoeffding inequality, we have

$$\Pr\left(\left|\frac{X_1 + \dots + X_n}{n} - H(p'(a))\right| \geq \varepsilon\right) \leq 2 \exp\left(-2n\varepsilon^2/x_{\text{upp}}^2\right). \quad (\text{G12})$$

On the other hand, according to the definition of T_ε , we have

$$\Pr\left(\left|\frac{X_1 + \dots + X_n}{n} - H(p'(a))\right| \geq \varepsilon\right) = 1 - \sum_{\mathbf{a} \in T_\varepsilon} p'(a_1) \dots p'(a_n). \quad (\text{G13})$$

It implies $\text{Tr}(\Pi\sigma^{\otimes n}) \geq 1 - 2 \exp(-2n\varepsilon^2/x_{\text{upp}}^2)$ using Eq. (G8). Similarly, using the Hoeffding inequality for independent random variables distributed to Y and Eq. (G9), we have $\sum_{\mathbf{a} \in \Sigma^n} \text{Tr}(\Lambda_{\mathbf{a}}\sigma_{\mathbf{a}}) \geq 1 - 2 \exp(-2n\varepsilon^2/y_{\text{upp}}^2)$. Plugging in these bounds in Eq. (G5), we have

$$\delta \leq 16e^{-2n\varepsilon^2/x_{\text{upp}}^2} + 8e^{-2n\varepsilon^2/y_{\text{upp}}^2} + 2^{4-n\varepsilon}, \quad (\text{G14})$$

proving the lemma. \blacksquare

APPENDIX H: EXAMPLES OF GLOBAL PREPROCESSING CONTROLS

1. Achievable FIs

We first calculate the asymptotic limits of the achievable FI using circuits shown in Fig. 3.

The first example is phase sensing using GHZ states. The initial state is

$$|\psi_\theta^{(n)}\rangle = \frac{e^{i\theta} |0\rangle^{\otimes n} + e^{-i\theta} |1\rangle^{\otimes n}}{\sqrt{2}}. \quad (\text{H1})$$

The optimal circuit U^G is composed of a $\text{C}(\text{NOT})^{n-1}$ gate, a Hadamard gate and another $\text{C}(\text{NOT})^{n-1}$ gate. The final state is

$$|\psi_\theta^{(n),\text{final}}\rangle = U^G |\psi_\theta^{(n)}\rangle = \cos(n\theta) |0\rangle^{\otimes n} + i \sin(n\theta) |1\rangle^{\otimes n}. \quad (\text{H2})$$

The noisy measurement is $\{M_i\}^{\otimes n} = \{M_0, M_1\}^{\otimes n}$ where $M_0 = (1-m)|0\rangle\langle 0| + m|1\rangle\langle 1|$ ($0 < m < 1/2$) and $M_1 =$

$\mathbb{1} - M_0$. The probability of getting a measurement result $\mathbf{b} \in \{0, 1\}^n$ is

$$\Pr(\mathbf{b}) = \Pr(b_1 b_2 \dots b_n) = \cos(n\theta)^2 m^{|\mathbf{b}|} (1-m)^{n-|\mathbf{b}|} + \sin(n\theta)^2 m^{n-|\mathbf{b}|} (1-m)^{|\mathbf{b}|}, \quad (\text{H3})$$

where $|\mathbf{b}|$ is the weight of \mathbf{b} . We divide the measurement outcomes into two sets where the outcome is 1 when $|\mathbf{b}|$ is larger than $\lfloor n/2 \rfloor$ and the outcome is 0 when $|\mathbf{b}|$ is smaller than or equal to $\lfloor n/2 \rfloor$. We call this a *majority voting postprocessing method*. Let

$$M_{0,\text{mj}} = \sum_{|\mathbf{b}| \leq \lfloor n/2 \rfloor} M_{b_1} \otimes \dots \otimes M_{b_n}, \quad \text{and} \\ M_{1,\text{mj}} = \mathbb{1} - M_{0,\text{mj}} = \sum_{|\mathbf{b}| > \lfloor n/2 \rfloor} M_{b_1} \otimes \dots \otimes M_{b_n}. \quad (\text{H4})$$

Then the probability of getting outcome 0 is

$$\Pr(0)_{\text{mj}} = \text{Tr}(\psi_\theta^{(n),\text{final}} M_{0,\text{mj}}) \\ = \cos(n\theta)^2 \Pr\left(Y_n \leq \frac{\lfloor n/2 \rfloor}{n}\right) \\ + \sin(n\theta)^2 \Pr\left(Y_n > 1 - \frac{\lfloor n/2 \rfloor}{n}\right). \quad (\text{H5})$$

Here we let X_1, \dots, X_n be independent identically distributed (IID) random variables such that $X_i = 0$ with probability $1-m$ and $X_i = 1$ with probability m and $Y_n = (X_1 + \dots + X_n)/n$. According to the Hoeffding's inequality (Lemma 18),

$$\Pr(|Y_n - m| > \varepsilon) \leq 2e^{-2n\varepsilon^2}. \quad (\text{H6})$$

For a sufficiently large n , we have $\lfloor n/2 \rfloor/n - m = \Omega(1)$, and

$$1 - 2e^{-2n(\lfloor n/2 \rfloor/n - m)^2} \leq \Pr\left(Y_n \leq \frac{\lfloor n/2 \rfloor}{n}\right) \leq 1, \quad (\text{H7})$$

$$0 \leq \Pr\left(Y_n > 1 - \frac{\lfloor n/2 \rfloor}{n}\right) \\ \leq 2e^{-2n(1 - \lfloor n/2 \rfloor/n - m)^2}. \quad (\text{H8})$$

Without loss of generality, we assume that $\pi/6n \leq \theta \leq \pi/3n$. Otherwise, we can always first find a rough estimate of θ_0 such that $|\theta - \theta_0| \leq \pi/12n$ and then insert a Pauli- X rotation $e^{-i\theta_0 X}$ after the Hadamard gate, such that the final state becomes

$$\cos(n(\theta - \theta_0)) |0\rangle^{\otimes n} + i \sin(n(\theta - \theta_0)) |1\rangle^{\otimes n}, \quad (\text{H9})$$

and $\pi/6n \leq \theta - \theta_0 \leq \pi/3n$. Then one can estimate $\theta - \theta_0$ using the majority voting post-processing method.

Assuming $\pi/6n \leq \theta \leq \pi/3n$, the achievable FI is

$$F(\psi_\theta^{(n),\text{final}}, \{M_{0,\text{mj}}, M_{1,\text{mj}}\}) = \frac{(\partial_\theta \text{Pr}(0)_{\text{mj}})^2}{\text{Pr}(0)_{\text{mj}}(1 - \text{Pr}(0)_{\text{mj}})} \quad (\text{H10})$$

$$= \frac{4n^2}{1 + \frac{(2-b_n)b_n - a_n^2}{a_n^2 \sin^2(2n\theta)} + \frac{(2-2b_n a_n) \cos(2n\theta)}{a_n^2 \sin^2(2n\theta)}}, \quad (\text{H11})$$

where $a_n = \Pr(Y_n \leq \lfloor n/2 \rfloor / n) - \Pr(Y_n > 1 - \lfloor n/2 \rfloor / n)$ and $b_n = \Pr(Y_n \leq \lfloor n/2 \rfloor / n) + \Pr(Y_n > 1 - \lfloor n/2 \rfloor / n)$. Noting that $\sin^2(2n\theta) \in [3/4, 1]$, $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 1$, we have

$$F(\psi_\theta^{(n),\text{final}}, \{M_{0,\text{mj}}, M_{1,\text{mj}}\}) \xrightarrow{n \rightarrow \infty} 4n^2. \quad (\text{H12})$$

To illustrate the importance of global preprocessing controls, we also compute the FI when we measure $\psi_\theta^{(n)}$ using a noisy local Pauli- X operator measurement, which is equivalent to applying a transversal Hadamard gate $H^{\otimes n}$ as the preprocessing control. Let $M'_0 = (1-m)|+\rangle\langle+| + m|-\rangle\langle-|$ and $M'_1 = m|+\rangle\langle+| + (1-m)|-\rangle\langle-|$, where $|\pm\rangle = |0\rangle \pm |1\rangle / \sqrt{2}$ is the basis of Pauli- X operator. We have

$$F(\psi_\theta^{(n)}, \{M'_0, M'_1\}^{\otimes n}) = F(H^{\otimes n} \psi_\theta^{(n)} H^{\otimes n}, \{M_i\}^{\otimes n}) \quad (\text{H13})$$

$$= \frac{(\partial_\theta p_{\text{odd}})^2}{p_{\text{odd}}} + \frac{(\partial_\theta p_{\text{even}})^2}{p_{\text{even}}}$$

$$= \frac{4n^2 \sin^2(2n\theta)^2}{\sin^2(2n\theta)^2 - 1 + (1-2m)^{-2n}}, \quad (\text{H14})$$

where p_{odd} (or p_{even}) is the probability of obtaining measurement outcomes that form an odd (or even) parity bit string, and

$$p_{\text{odd}} = \frac{1}{2} (1 + (1-2m)^n) \sin^2 n\theta + \frac{1}{2} (1 - (1-2m)^n) \cos^2 n\theta, \quad (\text{H15})$$

$$p_{\text{even}} = \frac{1}{2} (1 - (1-2m)^n) \sin^2 n\theta + \frac{1}{2} (1 + (1-2m)^n) \cos^2 n\theta. \quad (\text{H16})$$

Note that when $m = 0$, the FI $F(H^{\otimes n} \psi_\theta^{(n)} H^{\otimes n}, \{M_i\}^{\otimes n}) = 4n^2$ is equal to the QFI. However, when $m > 0$, the FI $F(H^{\otimes n} \psi_\theta^{(n)} H^{\otimes n}, \{M_i\}^{\otimes n}) = e^{-\Omega(n)}$ is exponentially small, demonstrating the necessity of performing global preprocessing controls. In general, it was illustrated in Ref. [55] that in the presence of noisy measurements, Heisenberg scaling cannot be achieved with local control.

The second example is phase sensing using product states, where

$$|\psi_\theta^{(n)}\rangle = \left(\frac{e^{i\theta}|0\rangle + e^{-i\theta}|1\rangle}{\sqrt{2}} \right)^{\otimes n}. \quad (\text{H17})$$

The optimal circuit is composed of a global Hadamard gate, a global Pauli- X rotation of angle $\theta_0/2$, a desymmetrization gate DS and a C(NOT)^{n-1} gate. Here θ_0 is chosen such that $1/n < |\theta - \theta_0| \ll 1/\sqrt{n}$ (assuming n is large enough). After the first two gates, the quantum state becomes

$$(\cos(\theta - \theta_0)|0\rangle + i \sin(\theta - \theta_0)|1\rangle)^{\otimes n}. \quad (\text{H18})$$

The desymmetrization gate DS is a unitary gate such that

$$|0\rangle^{\otimes n} \mapsto |0\rangle^{\otimes n}, \quad |W\rangle = \frac{1}{\sqrt{n}} (|10 \dots 0\rangle + |010 \dots 0\rangle + \dots + |0 \dots 01\rangle) \mapsto |10 \dots 0\rangle. \quad (\text{H19})$$

After DS and C(NOT)^{n-1} gates, the final quantum state is

$$|\psi_\theta^{(n),\text{final}}\rangle = \cos^n(\theta - \theta_0) |0\rangle^{\otimes n} + i\sqrt{n} \sin(\theta - \theta_0) \times \cos^{n-1}(\theta - \theta_0) |1\rangle^{\otimes n} + \dots, \quad (\text{H20})$$

where we omit in “...” a state perpendicular to $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$ whose norm is of magnitude $O(n(\theta - \theta_0)^2)$. Using the majority voting method as in the first example, we have

$$\text{Pr}(0)_{\text{mj}} = \text{Tr}(\psi_\theta^{(n),\text{final}} M_{0,\text{mj}}) = \cos^{2n}(\theta - \theta_0) \text{Pr}\left(Y_n \leq \frac{\lfloor n/2 \rfloor}{n}\right) \quad (\text{H21})$$

$$+ n \sin^2(\theta - \theta_0) \cos^{2(n-1)}(\theta - \theta_0) \text{Pr}\left(Y_n > 1 - \frac{\lfloor n/2 \rfloor}{n}\right) + O(n^2(\theta - \theta_0)^4), \quad (\text{H22})$$

$$\text{Pr}(1)_{\text{mj}} = \text{Tr}(\psi_\theta^{(n),\text{final}} M_{1,\text{mj}}) = \cos^{2n}(\theta - \theta_0) \text{Pr}\left(Y_n > \frac{\lfloor n/2 \rfloor}{n}\right) \quad (\text{H23})$$

$$+ n \sin^2(\theta - \theta_0) \cos^{2(n-1)}(\theta - \theta_0) \text{Pr}\left(Y_n \leq 1 - \frac{\lfloor n/2 \rfloor}{n}\right) + O(n^2(\theta - \theta_0)^4). \quad (\text{H24})$$

For a sufficiently large n , we have $\lfloor n/2 \rfloor/n - m = \Omega(1)$, and the achievable FI is

$$F(\psi_\theta^{(n),\text{final}}, \{M_{0,\text{mj}}, M_{1,\text{mj}}\}) = \frac{(\partial_\theta \Pr(0)_{\text{mj}})^2}{\Pr(0)_{\text{mj}}(1 - \Pr(0)_{\text{mj}})} \quad (\text{H25})$$

$$= \frac{(2n \sin(\theta - \theta_0) \cos^{2n-1}(\theta - \theta_0) + e^{-\Omega(n)} + O(n^2(\theta - \theta_0)^3))^2}{n \cos^{4n-2}(\theta - \theta_0) \sin(\theta - \theta_0)^2 + e^{-\Omega(n)} + O(n^2(\theta - \theta_0)^4)}. \quad (\text{H26})$$

Note that the first terms in the numerator and the denominator must be dominant terms because we have assumed $|\theta - \theta_0| \ll 1/n$. Therefore, we have,

$$F(\psi_\theta^{(n),\text{final}}, \{M_{0,\text{mj}}, M_{1,\text{mj}}\}) \xrightarrow{n \rightarrow \infty} 4n. \quad (\text{H27})$$

The third example is phase sensing using classically mixed states, where

$$\rho_\theta^{(n)} = (\cos^2 \theta |0\rangle\langle 0| + \sin^2 \theta |1\rangle\langle 1|)^{\otimes n}. \quad (\text{H28})$$

Our preprocessing circuit is composed of a sorting channel $\mathcal{S}_{\text{sorting}}$, discarding $n - 1$ qubits and preparing $|0\rangle^{\otimes n-1}$ and a C(NOT)^{n-1} gate. In particular, the sorting channel $\mathcal{S}_{\text{sorting}}$ first performs a sorting network such that

$$|b_1 b_2 \dots b_n\rangle \mapsto |1^{|b|} 0^{n-|b|}\rangle, \quad (\text{H29})$$

where $\mathbf{b} \in \{0, 1\}^n$. The second step of the sorting channel is to swap the first qubit with the $\lfloor n \sin^2 \theta_0 \rfloor$ th qubit, where θ_0 is chosen such that $|\theta - \theta_0| \ll 1/\sqrt{n}$, i.e., $|\theta - \theta_0| \sqrt{n} = o(1)$. Note that $\mathcal{S}_{\text{sorting}}$ is not a unitary channel. It can be viewed as an optimal quantum-classical channel as in Theorem 14 that converts the state into a $(n + 1)$ -level quantum system with no sensitivity loss. We will

later specify a circuit implementation of $\mathcal{S}_{\text{sorting}}$ using $O(n \log^2 n)$ ancillary qubits.

After discarding all but the first probe qubit, we have

$$p_\theta |0\rangle\langle 0| + (1 - p_\theta) |1\rangle\langle 1|, \quad (\text{H30})$$

where $p_\theta = \Pr(Y_n^\theta \leq \lfloor n \sin^2 \theta_0 \rfloor/n)$ is the probability that after flipping n biased coins whose head probability is $\sin^2 \theta$, the number of heads are smaller than or equal to $\lfloor n \sin^2 \theta_0 \rfloor$. Here we let $X_1^\theta, \dots, X_n^\theta$ be IID random variables such that $X_i^\theta = 0$ with probability $\cos^2(\theta)$ and $X_i^\theta = 1$ with probability $\sin^2(\theta)$ and $Y_n^\theta = (X_1^\theta + \dots + X_n^\theta)/n$. In particular, we use the Berry-Esseen theorem, which states the following.

Lemma 20.—(Berry-Esseen theorem [95,96]). Let B_1, \dots, B_n be IID random variables such that $\mathbb{E}[B_i] = 0$, $\mathbb{E}[B_i^2] = \sigma^2 > 0$, and $\mathbb{E}[|B_i|^3] < \infty$. F_n is the cumulative distribution function of $A_n \sqrt{n}/\sigma$, where $A_n = (B_1 + B_2 + \dots + B_n)/n$ and Φ is the cumulative distribution function of the standard normal distribution, i.e., $\Phi(x) = \int_{-\infty}^x (1/\sqrt{2\pi}) e^{-(1/2)y^2} dy$. Then for all x and n ,

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sigma^3 \sqrt{n}} \quad (\text{H31})$$

for some constant C .

Then we have

$$p_\theta = \Pr(Y_n^\theta \leq \lfloor n \sin^2 \theta_0 \rfloor/n) = F_n \left(\frac{\sqrt{n} \left(\frac{\lfloor n \sin^2 \theta_0 \rfloor}{n} - \sin^2 \theta \right)}{|\cos(\theta) \sin(\theta)|} \right) \quad (\text{H32})$$

$$= \Phi \left(\frac{\sqrt{n} \left(\frac{\lfloor n \sin^2 \theta_0 \rfloor}{n} - \sin^2 \theta \right)}{|\cos(\theta) \sin(\theta)|} \right) + O \left(\frac{1}{\sqrt{n}} \right) = \frac{1}{2} + o(1), \quad (\text{H33})$$

where we use $\Phi(0) = 1/2$ and $\lfloor n \sin^2 \theta_0 \rfloor/n - \sin^2 \theta = o(1/\sqrt{n})$. On the other hand,

$$\partial_\theta p_\theta = \partial_\theta \Pr(Y_n^\theta \leq \lfloor n \sin^2 \theta_0 \rfloor/n) \quad (\text{H34})$$

$$= \partial_\theta \sum_{0 \leq n_h \leq \lfloor n \sin^2 \theta_0 \rfloor} \binom{n}{n_h} \sin^{2n_h} \theta \cos^{2(n-n_h)} \theta \quad (\text{H35})$$

$$= \sum_{0 \leq n_h \leq \lfloor n \sin^2 \theta_0 \rfloor} \binom{n}{n_h} (2n_h \sin^{2n_h-1} \theta \cos^{2(n-n_h)+1} \theta - 2(n-n_h) \sin^{2n_h+1} \theta \cos^{2(n-n_h)-1} \theta) \quad (\text{H36})$$

$$= -2n \binom{n-1}{\lfloor n \sin^2 \theta_0 \rfloor} \sin^{2\lfloor n \sin^2 \theta_0 \rfloor+1} \theta \cos^{2(n-\lfloor n \sin^2 \theta_0 \rfloor)-1} \theta, \quad (\text{H37})$$

and asymptotically we have (using \simeq to mean asymptotically equivalence),

$$\partial_\theta p_\theta \simeq -\frac{2n}{\sqrt{2\pi}} \frac{(n-1)^{n-\frac{1}{2}}}{\lfloor n \sin^2 \theta_0 \rfloor^{n \sin^2 \theta_0 + \frac{1}{2}} (n-1 - \lfloor n \sin^2 \theta_0 \rfloor)^{n-\frac{1}{2}-\lfloor n \sin^2 \theta_0 \rfloor}} \quad (\text{H38})$$

$$\simeq -\frac{2\sqrt{n}}{\sqrt{2\pi}} \left(\frac{(n-1) \sin^2 \theta}{(n-1) \sin^2 \theta + c_n} \right)^{(n-1) \sin^2 \theta + c_n} \left(\frac{(n-1) \cos^2 \theta}{(n-1) \cos^2 \theta - c_n} \right)^{(n-1) \cos^2 \theta - c_n} \quad (\text{H39})$$

$$= -\frac{2\sqrt{n}}{\sqrt{2\pi}} \exp \left(-((n-1) \sin^2 \theta + c_n) \ln \left(1 + \frac{c_n}{(n-1) \sin^2 \theta} \right) - ((n-1) \cos^2 \theta - c_n) \ln \left(1 - \frac{c_n}{(n-1) \cos^2 \theta} \right) \right) \quad (\text{H40})$$

$$= -\frac{2\sqrt{n}}{\sqrt{2\pi}} \exp \left(-\frac{2c_n^2}{(n-1) \sin^2(2\theta)} + O\left(\frac{c_n^3}{n^2}\right) \right) \simeq -\frac{2\sqrt{n}}{\sqrt{2\pi}}, \quad (\text{H41})$$

where we use Stirling's formula [97] in the first step, use $c_n := -(n-1) \sin^2 \theta + \lfloor n \sin^2 \theta_0 \rfloor$ in the second step, use the Taylor expansion $\ln(1+x) = x - x^2/2 + O(x^3)$ and $c_n = o(\sqrt{n})$ in the last step.

After resetting the discard qubits to be $|0\rangle^{\otimes n-1}$ and performing the $\text{C}(\text{NOT})^{n-1}$ gate, we have

$$\rho_\theta^{(n), \text{final}} = p_\theta |0\rangle^{\otimes n} \langle 0|^{\otimes n} + (1-p_\theta) |1\rangle^{\otimes n} \langle 1|^{\otimes n}. \quad (\text{H42})$$

Using the majority voting method, we have

$$\Pr(0)_{\text{mj}} = \text{Tr}(\rho_\theta^{(n), \text{final}} M_{0, \text{mj}}) = p_\theta \Pr \left(Y_n \leq \frac{\lfloor n/2 \rfloor}{n} \right) + (1-p_\theta) \Pr \left(Y_n > 1 - \frac{\lfloor n/2 \rfloor}{n} \right), \quad (\text{H43})$$

$$\Pr(1)_{\text{mj}} = \text{Tr}(\rho_\theta^{(n), \text{final}} M_{1, \text{mj}}) = p_\theta \Pr \left(Y_n > \frac{\lfloor n/2 \rfloor}{n} \right) + (1-p_\theta) \Pr \left(Y_n \leq 1 - \frac{\lfloor n/2 \rfloor}{n} \right). \quad (\text{H44})$$

For a sufficiently large n , we have $\lfloor n/2 \rfloor/n - m = \Omega(1)$, and the achievable FI is

$$F(\rho_\theta^{(n), \text{final}}, \{M_{0, \text{mj}}, M_{1, \text{mj}}\}) = \frac{(\partial_\theta \Pr(0)_{\text{mj}})^2}{\Pr(0)_{\text{mj}}(1 - \Pr(0)_{\text{mj}})} \quad (\text{H45})$$

$$= \frac{(\partial_\theta p_\theta + e^{-\Omega(n)})^2}{(p_\theta + e^{-\Omega(n)})(1 - p_\theta + e^{-\Omega(n)})}. \quad (\text{H46})$$

Therefore, we have,

$$F(\rho_\theta^{(n), \text{final}}, \{M_{0, \text{mj}}, M_{1, \text{mj}}\}) \xrightarrow{n \rightarrow \infty} \frac{8}{\pi} n. \quad (\text{H47})$$

Note that the reason that the QFI is not achieved here lies in the second step where all but one probe qubit are discarded.

If, instead, we can perfectly encode the entire $(n+1)$ -level system that we obtain in the first step to a n -qubit state that is immune to measurement errors, the QFI will be achieved.

2. Gate complexity

Finally, we discuss the gate complexities of implementing $\text{C}(\text{NOT})^{n-1}$, DS , and $\mathcal{S}_{\text{sorting}}$. We will show that, assuming arbitrary two-qubit gates and all-to-all connectivity, the $\text{C}(\text{NOT})^{n-1}$ and the desymmetrization gate DS can be implemented using $O(n)$ gates in depth $O(\log n)$. The $\mathcal{S}_{\text{sorting}}$ channel can be implemented using $O(n \log^2 n)$ gates and $O(n \log^2 n)$ ancillary qubits in depth $O(\log^2 n)$.

We first investigate the implementation of $\text{C}(\text{NOT})^{n-1}$ gates and DS gates. Note that in experimental platforms like Rydberg atoms where long-range interactions are

available, the $C(\text{NOT})^{n-1}$ gate can be implemented in a single step [98]. However, we focus on the standard quantum circuit model here where only two-qubit gates are allowed.

Reference [99] included detailed quantum circuits for $C(\text{NOT})^{n-1}$ gates and DS gates using $O(n)$ gates in depth $O(\log n)$. For completeness, we briefly discuss these circuits here, in the case where $n = 2^k$.

To implement $C(\text{NOT})^{n-1}$, one starts with a Hadamard gate on the first qubit, and then implement $C_1(\text{NOT})_2$ (which means a CNOT gate where the control qubit is the first qubit and the target qubit is the second) in the first step, $C_1(\text{NOT})_3$ and $C_2(\text{NOT})_4$ in the second step, and so on. The circuit continues in the same way. In the final step, i.e., the k th step, $C_l(\text{NOT})_{2^{k-1+l}}$ for $l = 1, 2, \dots, 2^{k-1}$ are implemented. One can verify the above $O(\log n)$ -depth circuit implements a $C(\text{NOT})^{n-1}$ gate using $O(n)$ single- or two-qubit gates.

To implement DS , one can equivalently consider the circuit implementation of DS^\dagger and then conjugate and reverse the orders of each gate. DS^\dagger is a gate that prepares W states, where

$$|10 \dots 0\rangle \mapsto \frac{1}{\sqrt{n}} (|10 \dots 0\rangle + |010 \dots 0\rangle + \dots + |0 \dots 01\rangle), \quad |00 \dots 0\rangle \mapsto |00 \dots 0\rangle. \quad (\text{H48})$$

To implement DS^\dagger , one starts with a Pauli- X gate on the first qubit, then performs a two-qubit gate that is a composition of a C_1H_2 (controlled-Hadamard) gate and then a $C_2\text{NOT}_1$ gate (again, we use subscripts l to denote the l th qubit) in the first step. $C_1H_3 + C_3\text{NOT}_1$ and $C_2H_4 + C_4\text{NOT}_2$ in the second step, and so on. The circuit continues

in the same way. In the final step, i.e., the k th step, $C_lH_{2^{k-1+l}} + C_{2^{k-1+l}}\text{NOT}_l$ for $l = 1, 2, \dots, 2^{k-1}$ are implemented. One can verify the above $O(\log n)$ -depth circuit implements a DS^\dagger gate using $O(n)$ single- or two-qubit gates.

Finally, we discuss the implementation of S_{sorting} which can be decomposed into a sorting network that implements

$$|b_1 b_2 \dots b_n\rangle \mapsto |1^{|b|} 0^{n-|b|}\rangle, \quad (\text{H49})$$

for $\mathbf{b} \in \{0, 1\}^n$ and a SWAP gate that swaps the first qubit with the $\lfloor n \sin^2 \theta_0 \rfloor$ th qubit. Now we discuss the implementation of the sorting network [Eq. (H49)], which directly follows from a classical sorting network because our input state is a classically mixed state and the sorting channel is incoherent. To be specific, we define a comparator to be a two-qubit quantum channel such that

$$|ij\rangle \mapsto \begin{cases} |ij\rangle & \text{when } i \geq j, \\ |ji\rangle & \text{when } j > i. \end{cases} \quad (\text{H50})$$

It can be implemented using a unitary gate acting on two probe qubits and one ancillary qubit such that

$$|ij\rangle |0\rangle \mapsto \begin{cases} |ij\rangle |0\rangle & \text{when } i \geq j, \\ |ji\rangle |1\rangle & \text{when } j > i, \end{cases} \quad (\text{H51})$$

and discarding the ancillary qubit afterwards. Our sorting network [Eq. (H49)] then follows from a classical sorting network, replacing all its classical comparators with the two-qubit sorting channels described above.

Here we use a classical sorting network called a bitonic sorter [100] that uses $O(n \log^2 n)$ comparators in depth

Input: *arr*: the array to be sort.

Output: *arr*: the sorted array in descending order.

```

1: arr = BITONICSORT(arr, "descending")
2: function BITONICSORT(arr, direction)
3:   n = the length of arr
4:   if n > 1 then
5:     BITONICSORT(arr[1:n/2], "ascending")
6:     BITONICSORT(arr[n/2+1:n], "descending")
7:     MERGE(arr, direction)
8:   end if
9: end function
10: function MERGE(arr, direction)
11:   n = the length of arr
12:   if n > 1 then
13:     for i = 1, 2, ..., n/2 do
14:       Exchange arr[i] and arr[i+n/2] if they are not in the right order
15:     end for
16:     MERGE(arr[1:n/2], direction)
17:     MERGE(arr[n/2+1:n], direction)
18:   end if
19: end function

```

ALGORITHM 1. Bitonic sort algorithm.

$O(\log^2 n)$. Note that it is also possible to construct sorting networks of depth $O(\log n)$ [and size $O(n \log^2 n)$] [101], although the linear constant is large, making it impractical. We briefly summarize the bitonic sort algorithm in the following pseudocode. Note that here we assume $n = 2^k$ (we can always add more qubits in prepared in $|0\rangle$ to make n a power of 2).

The bitonic sort algorithm can be divided into two steps: (1) forming a bitonic sequence; (2) sorting a bitonic sequence. A bitonic sequence of length n is defined to be a sequence \mathbf{b} where there is an index i such that (b_1, \dots, b_i) is monotonically nondecreasing, and (b_{i+1}, \dots, b_n) is monotonically nonincreasing, or a sequence that can be cyclically shifted into the above sequence. MERGE(*arr*, *direction*) sorts a bitonic sequence *arr* in the required order. It can be proven that the MERGE function contains $O(\log n)$ parallel computing steps; and the BITONICSORT function contains $O(\log^2 n)$ parallel computing steps. As a result, one can see that the bitonic sorter uses $O(n \log^2 n)$ comparators in depth $O(\log^2 n)$.

-
- [1] Christian L. Degen, Friedemann Reinhard, and Paola Cappellaro, Quantum sensing, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [2] Luca Pezze, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein, Quantum metrology with nonclassical states of atomic ensembles, *Rev. Mod. Phys.* **90**, 035005 (2018).
- [3] Stefano Pirandola, B. Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd, Advances in photonic quantum sensing, *Nat. Photon.* **12**, 724 (2018).
- [4] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, Advances in quantum metrology, *Nat. Photon.* **5**, 222 (2011).
- [5] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, Quantum Metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [6] Carlton M. Caves, Quantum-mechanical noise in an interferometer, *Phys. Rev. D* **23**, 1693 (1981).
- [7] Dietrich Leibfried, Murray D. Barrett, T. Schaetz, Joseph Britton, J. Chiaverini, Wayne M. Itano, John D. Jost, Christopher Langer, and David J. Wineland, Toward Heisenberg-limited spectroscopy with multiparticle entangled states, *Science* **304**, 1476 (2004).
- [8] Morgan W. Mitchell, Jeff S. Lundeen, and Aephraem M. Steinberg, Super-resolving phase measurements with a multiphoton entangled state, *Nature* **429**, 161 (2004).
- [9] Mankei Tsang, Ranjith Nair, and Xiao-Ming Lu, Quantum Theory of Superresolution for Two Incoherent Optical Point Sources, *Phys. Rev. X* **6**, 031033 (2016).
- [10] Raphael Kaubruegger, Denis V. Vasilyev, Marius Schulte, Klemens Hammerer, and Peter Zoller, Quantum Variational Optimization of Ramsey Interferometry and Atomic Clocks, *Phys. Rev. X* **11**, 041045 (2021).
- [11] Christian D. Marciniak, Thomas Feldker, Ivan Pogorelov, Raphael Kaubruegger, Denis V. Vasilyev, Rick van Bijnen, Philipp Schindler, Peter Zoller, Rainer Blatt, and Thomas Monz, Optimal metrology with programmable quantum sensors, *Nature* **603**, 604 (2022).
- [12] Alexander S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, Netherlands, 1982).
- [13] Carl Wilhelm Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [14] Samuel L. Braunstein and Carlton M. Caves, Statistical Distance and the Geometry of Quantum States, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [15] Jing Liu, Haidong Yuan, Xiao-Ming Lu, and Xiaoguang Wang, Quantum Fisher information matrix and multiparameter estimation, *J. Phys. A: Math. Theor.* **53**, 023001 (2019).
- [16] Dénes Petz and Csaba Sudár, Geometries of quantum states, *J. Math. Phys.* **37**, 2662 (1996).
- [17] Akio Fujiwara and Hiroshi Imai, A fibre bundle over manifolds of quantum channels and its application to quantum statistics, *J. Phys. A: Math. Theor.* **41**, 255304 (2008).
- [18] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, *Nat. Phys.* **7**, 406 (2011).
- [19] Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Mădălin Guță, The elusive Heisenberg limit in quantum-enhanced metrology, *Nat. Commun.* **3**, 1063 (2012).
- [20] Haidong Yuan and Chi-Hang Fred Fung, Fidelity and fisher information on quantum channels, *New J. Phys.* **19**, 113039 (2017).
- [21] Akira Sone, M. Cerezo, Jacob L. Beckey, and Patrick J. Coles, Generalized measure of quantum fisher information, *Phys. Rev. A* **104**, 062602 (2021).
- [22] Anian Altherr and Yuxiang Yang, Quantum Metrology for non-Markovian Processes, *Phys. Rev. Lett.* **127**, 060501 (2021).
- [23] M. V. Gurudev Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. S. Zibrov, P. R. Hemmer, and M. D. Lukin, Quantum register based on individual electronic and nuclear spin qubits in diamond, *Science* **316**, 1312 (2007).
- [24] R. Hanson, V. V. Dobrovitski, A. E. Feiguin, O. Gywat, and D. D. Awschalom, Coherent dynamics of a single spin interacting with an adjustable spin bath, *Science* **320**, 352 (2008).
- [25] J. M. Taylor, Paola Cappellaro, L. Childress, Liang Jiang, Dmitry Budker, P. R. Hemmer, Amir Yacoby, R. Walsworth, and M. D. Lukin, High-sensitivity diamond magnetometer with nanoscale resolution, *Nat. Phys.* **4**, 810 (2008).
- [26] Liang Jiang, J. S. Hodges, J. R. Maze, Peter Maurer, J. M. Taylor, D. G. Cory, P. R. Hemmer, Ronald Lee Walsworth, Amir Yacoby, and Alexander S. Zibrov, *et al.*, Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae, *Science* **326**, 267 (2009).
- [27] Marcus W. Doherty, Neil B. Manson, Paul Delaney, Fedor Jelezko, Jörg Wrachtrup, and Lloyd C. L. Hollenberg, The nitrogen-vacancy colour centre in diamond, *Phys. Rep.* **528**, 1 (2013).
- [28] Thomas Unden, Priya Balasubramanian, Daniel Louzon, Yuval Vinkler, Martin B. Plenio, Matthew Markham,

- Daniel Twitchen, Alastair Stacey, Igor Lovchinsky, and Alexander O. Sushkov, *et al.*, Quantum Metrology Enhanced by Repetitive Quantum Error Correction, *Phys. Rev. Lett.* **116**, 230502 (2016).
- [29] Simon Schmitt, Tuvia Gefen, Daniel Louzon, Christian Osterkamp, Nicolas Staudenmaier, Johannes Lang, Matthew Markham, Alex Retzker, Liam P. McGuinness, and Fedor Jelezko, Optimal frequency measurements with quantum probes, *npj Quantum Inf.* **7**, 55 (2021).
- [30] Philip Krantz, Morten Kjaergaard, Fei Yan, Terry P. Orlando, Simon Gustavsson, and William D. Oliver, A quantum engineer's guide to superconducting qubits, *Appl. Phys. Rev.* **6**, 021318 (2019).
- [31] A. H. Myerson, D. J. Szwer, S. C. Webster, D. T. C. Allcock, M. J. Curtis, G. Imreh, J. A. Sherman, D. N. Stacey, A. M. Steane, and D. M. Lucas, High-Fidelity Readout of Trapped-Ion Qubits, *Phys. Rev. Lett.* **100**, 200502 (2008).
- [32] Colin D. Bruzewicz, John Chiaverini, Robert McConnell, and Jeremy M. Sage, Trapped-ion quantum computing: Progress and challenges, *Appl. Phys. Rev.* **6**, 021314 (2019).
- [33] Lorenza Viola, Emanuel Knill, and Seth Lloyd, Dynamical Decoupling of Open Quantum Systems, *Phys. Rev. Lett.* **82**, 2417 (1999).
- [34] Eric M. Kessler, Igor Lovchinsky, Alexander O. Sushkov, and Mikhail D. Lukin, Quantum Error Correction for Metrology, *Phys. Rev. Lett.* **112**, 150802 (2014).
- [35] W. Dür, M. Skotiniotis, Florian Froewis, and B. Kraus, Improved Quantum Metrology Using Quantum Error Correction, *Phys. Rev. Lett.* **112**, 080801 (2014).
- [36] Gilad Arrad, Yuval Vinkler, Dorit Aharonov, and Alex Retzker, Increasing Sensing Resolution With Error Correction, *Phys. Rev. Lett.* **112**, 150801 (2014).
- [37] Jing Liu and Haidong Yuan, Quantum parameter estimation with optimal control, *Phys. Rev. A* **96**, 012117 (2017).
- [38] Kaoru Yamamoto, Suguru Endo, Hideaki Hakoshima, Yuichiro Matsuzaki, and Yuuki Tokunaga, Error-mitigated quantum metrology, *ArXiv:2112.01850* (2021).
- [39] Francesco Albarelli, Matteo A. C. Rossi, Dario Tamascelli, and Marco G. Genoni, Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment, *Quantum* **2**, 110 (2018).
- [40] Frank Verstraete, Michael M. Wolf, and J. Ignacio Cirac, Quantum computation and quantum-state engineering driven by dissipation, *Nat. Phys.* **5**, 633 (2009).
- [41] Liang Jiang, Ana Maria Rey, Oriol Romero-Isart, Juan José García-Ripoll, Anna Sanpera, and Mikhail D. Lukin, Preparation of decoherence-free cluster states with optical superlattices, *Phys. Rev. A* **79**, 022309 (2009).
- [42] Mattias T. Johnsson, Nabomita Roy Mukty, Daniel Burgarth, Thomas Volz, and Gavin K. Brennen, Geometric Pathway to Scalable Quantum Sensing, *Phys. Rev. Lett.* **125**, 190403 (2020).
- [43] Sisi Zhou, Mengzhen Zhang, John Preskill, and Liang Jiang, Achieving the Heisenberg limit in quantum metrology using quantum error correction, *Nat. Commun.* **9**, 78 (2018).
- [44] Sisi Zhou and Liang Jiang, Asymptotic Theory of Quantum Channel Estimation, *PRX Quantum* **2**, 010343 (2021).
- [45] Qiushi Liu, Zihao Hu, Haidong Yuan, and Yuxiang Yang, Optimal Strategies of Quantum Metrology with a Strict Hierarchy, *Phys. Rev. Lett.* **130**, 070803 (2023).
- [46] Emily Davis, Gregory Bentsen, and Monika Schleier-Smith, Approaching the Heisenberg Limit Without Single-Particle Detection, *Phys. Rev. Lett.* **116**, 053601 (2016).
- [47] Florian Fröwis, Pavel Sekatski, and Wolfgang Dür, Detecting Large Quantum Fisher Information with Finite Measurement Precision, *Phys. Rev. Lett.* **116**, 090801 (2016).
- [48] Tommaso Macri, Augusto Smerzi, and Luca Pezzè, Loschmidt echo for quantum metrology, *Phys. Rev. A* **94**, 010102 (2016).
- [49] Samuel P. Nolan, Stuart S. Szigeti, and Simon A. Haine, Optimal and Robust Quantum Metrology using Interaction-Based Readouts, *Phys. Rev. Lett.* **119**, 193601 (2017).
- [50] Simon A. Haine, Using interaction-based readouts to approach the ultimate limit of detection-noise robustness for quantum-enhanced metrology in collective spin systems, *Phys. Rev. A* **98**, 030303 (2018).
- [51] Martin Koppenhöfer, Peter Groszkowski, and A. A. Clerk, Squeezed superradiance enables robust entanglement-enhanced metrology even with highly imperfect readout, *ArXiv:2304.05471* (2023).
- [52] Onur Hosten, Radha Krishnakumar, Nils J. Engelsens, and Mark A. Kasevich, Quantum phase magnification, *Science* **352**, 1552 (2016).
- [53] Daniel Linnemann, Helmut Strobel, Wolfgang Muesel, Jonas Schulz, Robert J. Lewis-Swan, Karen V. Kheruntsyan, and Markus K. Oberthaler, Quantum-Enhanced Sensing Based on Time Reversal of Nonlinear Dynamics, *Phys. Rev. Lett.* **117**, 013001 (2016).
- [54] Zeyang Li, Simone Colombo, Chi Shu, Gustavo Velez, Saúl Pilatowsky-Cameo, Roman Schmied, Soonwon Choi, Mikhail Lukin, Edwin Pedrozo-Peñañiel, and Vladan Vuletić, Improving metrology with quantum scrambling, *Science* **380**, 1381 (2023).
- [55] Yink Loong Len, Tuvia Gefen, Alex Retzker, and Jan Kołodźński, Quantum metrology with imperfect measurements, *Nat. Commun.* **13**, 6971 (2022).
- [56] Filip B. Maciejewski, Zoltán Zimborás, and Michał Oszmaniec, Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography, *Quantum* **4**, 257 (2020).
- [57] Michael R. Geller, Rigorous measurement error correction, *Quantum Sci. Technol.* **5**, 03LT01 (2020).
- [58] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. McKay, and Jay M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [59] Jochen Gorski, Frank Pfeuffer, and Kathrin Klamroth, Biconvex sets and optimization with biconvex functions: A survey and extensions, *Math. Methods Oper. Res.* **66**, 373 (2007).

- [60] Alexander S. Holevo, The capacity of the quantum channel with general signal states, *IEEE Trans. Inf. Theory* **44**, 269 (1998).
- [61] Benjamin Schumacher and Michael D. Westmoreland, Sending classical information via noisy quantum channels, *Phys. Rev. A* **56**, 131 (1997).
- [62] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, New York, 2010).
- [63] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Volume I Estimation Theory* (Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1993).
- [64] Erich L. Lehmann and George Casella, *Theory of Point Estimation* (Springer, New York, New York, 2006).
- [65] Hisashi Kobayashi, Brian L. Mark, and William Turin, *Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance* (Cambridge University Press, New York, New York, 2011).
- [66] Luigi Seveso, Francesco Albarelli, Marco G. Genoni, and Matteo G. A. Paris, On the discontinuity of the quantum fisher information for quantum statistical models with parameter dependent rank, *J. Phys. A: Math. Theor.* **53**, 02LT01 (2019).
- [67] C. Huerta Alderete, Max Hunter Gordon, Frédéric Sauvage, Akira Sone, Andrew T. Sornborger, Patrick J. Coles, and M. Cerezo, Inference-Based Quantum Sensing, *Phys. Rev. Lett.* **129**, 190501 (2022).
- [68] John Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, England, 2018).
- [69] Tuvia Gefen, Maxim Khodas, Liam P. McGuinness, Fedor Jelezko, and Alex Retzker, Quantum spectroscopy of single spins assisted by a classical clock, *Phys. Rev. A* **98**, 013844 (2018).
- [70] K. S. Cujia, Jens M. Boss, Konstantin Herb, Jonathan Zopes, and Christian L. Degen, Tracking the precession of single nuclear spins by weak measurements, *Nature* **571**, 230 (2019).
- [71] Matthias Pfender, Ping Wang, Hitoshi Sumiya, Shinobu Onoda, Wen Yang, Durga Bhaktavatsala Rao Dasari, Philipp Neumann, Xin-Yu Pan, Junichi Isoya, and Ren-Bao Liu, *et al.*, High-resolution spectroscopy of single nuclear spins via sequential weak measurements, *Nat. Commun.* **10**, 594 (2019).
- [72] D. Cohen, T. Gefen, L. Ortiz, and A. Retzker, Achieving the ultimate precision limit with a weakly interacting quantum probe, *npj Quantum Inf.* **6**, 83 (2020).
- [73] O. E. Barndorff-Nielsen and R. D. Gill, Fisher information in quantum statistics, *J. Phys. A: Math. Gen.* **33**, 4481 (2000).
- [74] Masahito Hayashi, Comparison between the Cramer-Rao and the mini-max approaches in quantum channel estimation, *Commun. Math. Phys.* **304**, 689 (2011).
- [75] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi, Attaining the ultimate precision limit in quantum state estimation, *Commun. Math. Phys.* **368**, 223 (2019).
- [76] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, New York, 2004).
- [77] Carl Helstrom, The minimum variance of estimates in quantum signal detection, *IEEE Trans. Inf. Theory* **14**, 234 (1968).
- [78] Luca Pezzé and Augusto Smerzi, Entanglement, Nonlinear Dynamics, and the Heisenberg Limit, *Phys. Rev. Lett.* **102**, 100401 (2009).
- [79] B. M. Escher, Quantum noise-to-sensibility ratio, *ArXiv: 1212.2533* (2012).
- [80] Katarzyna Macieszczak, Martin Fraas, and Rafał Demkowicz-Dobrzański, Bayesian quantum frequency estimation in presence of collective dephasing, *New J. Phys.* **16**, 113002 (2014).
- [81] Christodoulos A. Floudas, *Deterministic Global Optimization: Theory, Methods and Applications* (Springer Science & Business Media, Dordrecht, Netherlands, 2013), Vol. 37.
- [82] William K. Wootters, Statistical distance and Hilbert space, *Phys. Rev. D* **23**, 357 (1981).
- [83] Stephen J. Blundell and Katherine M. Blundell, *Concepts in Thermal Physics* (Oxford University Press, New York, New York, 2010).
- [84] Michał Oszmaniec, Remigiusz Augusiak, Christian Gogolin, Jan Kołodyński, Antonio Acín, and Maciej Lewenstein, Random Bosonic States for Robust Quantum Metrology, *Phys. Rev. X* **6**, 041044 (2016).
- [85] Marcin Jarzyna and Jan Kołodyński, Geometric approach to quantum statistical inference, *IEEE J. Sel. Areas Inf. Theory* **1**, 367 (2020).
- [86] William Fulton and Joe Harris, *Representation Theory: a First Course* (Springer Science & Business Media, New York, New York, 2013), Vol. 129.
- [87] Masahito Hayashi, *Group Representation for Quantum Theory* (Springer, Cham, Switzerland, 2017).
- [88] Roe Goodman and Nolan R. Wallach, *Representations and Invariants of the Classical Groups* (Cambridge University Press, New York, New York, 2000).
- [89] Dave Bacon, Isaac L. Chuang, and Aram W. Harrow, Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms, *Phys. Rev. Lett.* **97**, 170502 (2006).
- [90] Norman F. Ramsey, A molecular beam resonance method with separated oscillating fields, *Phys. Rev.* **78**, 695 (1950).
- [91] Dominik Safránek, Simple expression for the quantum fisher information matrix, *Phys. Rev. A* **97**, 042322 (2018).
- [92] Masahito Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, *Phys. Rev. A* **76**, 062301 (2007).
- [93] Wassily Hoeffding, in *The collected works of Wassily Hoeffding* (Springer, 1994), p. 409.
- [94] Mark M. Wilde, *Quantum Information Theory* (Cambridge University Press, New York, New York, 2013).
- [95] Andrew C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Am. Math. Soc.* **49**, 122 (1941).
- [96] Carl-Gustav Esseen, A moment inequality with an application to the central limit theorem, *Scand. Actuar. J.* **1956**, 160 (1956).
- [97] Joel Spencer, *Asymptopia* (American Mathematical Soc., Providence, Rhode Island, 2014), Vol. 71.

- [98] M. Müller, I. Lesanovsky, H. Weimer, H. P. Büchler, and P. Zoller, Mesoscopic Rydberg Gate Based on Electromagnetically Induced Transparency, *Phys. Rev. Lett.* **102**, 170502 (2009).
- [99] Diogo Cruz, Romain Fournier, Fabien Gremion, Alix Jeannerot, Kenichi Komagata, Tara Tosić, Jarla Thiesbrummel, Chun Lam Chan, Nicolas Macris, and Marc-André Dupertuis, *et al.*, Efficient quantum algorithms for GHZ and W states, and implementation on the IBM quantum computer, *Adv. Quantum Technol.* **2**, 1900015 (2019).
- [100] Behrooz Parhami, *Introduction to Parallel Processing: Algorithms and Architectures* (Springer Science & Business Media, New York, New York, 2006).
- [101] Miklós Ajtai, János Komlós, and Endre Szemerédi, in *Proceedings of the fifteenth annual ACM symposium on Theory of computing* (1983), p. 1.