Compression of metrological quantum information in the presence of noise

Flavio Salvati,[1] Wilfred Salmon,[2,3] Crispin H. W. Barnes,[1] and David R. M. Arvidsson-Shukur[3]

[1] *Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[2] *DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom*
[3] *Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

(Received 31 July 2023; accepted 16 July 2024; published 15 August 2024)

In quantum metrology, information about unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ is accessed by measuring probe states $\hat{\rho}_{\boldsymbol{\theta}}$. In experimental settings where copies of $\hat{\rho}_{\boldsymbol{\theta}}$ can be produced rapidly (e.g., in optics), the information-extraction bottleneck can stem from high postprocessing costs or detector saturation. In these regimes, it is desirable to compress the information encoded in $\hat{\rho}_{\boldsymbol{\theta}}^{\otimes n}$ into $m < n$ copies of a postselected state: $\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}\,\otimes m}$. Remarkably, recent works have shown that, in the absence of noise, compression can be lossless, for m/n arbitrarily small. Here, we fully characterize the family of filters that enable lossless compression. Further, we study the effect of noise on quantum-metrological information amplification. Motivated by experiments, we consider a popular family of filters, which we show is optimal for qubit probes. Further, we show that, for the optimal filter in this family, compression is still lossless if noise acts after the filter. However, in the presence of depolarizing noise before filtering, compression is lossy. In both cases, information extraction can be implemented significantly better than simply discarding a constant fraction of the states, even in the presence of strong noise.

DOI: 10.1103/PhysRevA.110.022426

I. INTRODUCTION

Quantum metrology is a promising application of quantum technologies. By using measurement probes made of quantum states, quantum metrology exploits nonclassical effects, such as entanglement and squeezing, to make high-resolution measurements of physical parameters. Using quantum resources, one can reduce errors in measurements compared to classical strategies [1–11].

In metrology, one can improve signal-to-noise ratios by preparing and measuring an increasing number of probes. However, measurement devices often have limited sensitivity and suffer from a dead time, i.e., the time needed to reset a detector after triggering it. Moreover, each measurement might generate large overheads in terms of postprocessing. Even if probes are "cheap" to produce, they may be expensive to measure.

These problems can be mitigated by compressing metrological information into fewer states, using a strategy known as *postselection*. Postselection is essentially the application of a filter. As recently demonstrated in a quantum-optics experiment [12], postselection of quantum probes can compress information beyond classically achievable limits [13]. In particular, postselected quantum metrology can allow detectors to operate at lower intensities while retaining the vast majority of the information encoded in the original high-intensity beam of probe states. This can reduce saturation of sensitive components, as well as alleviate computational costs associated with postprocessing. The most common instance of postselected metrology is *weak-value amplification* [14], which has found many applications [15–25].

In recent papers [12,26] Jenne, Arvidsson-Shukur, and Lupu-Gladstein (JAL) *et al.* introduced a postselection filter that can compress information contained in $\hat{\rho}_{\boldsymbol{\theta}}^{\otimes n}$ into $\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}\,\otimes m}$, where $m \leqslant n$. Moreover, m/n can be made arbitrarily small and the compression can happen without any loss of information. These remarkable properties of the JAL filter rely on the quantum metrology experiments' being completely noise free. However, noise in real experiments leads to natural limits on compression [12]. Until now, there exists no thorough investigation of the effect of noise on general multiparameter postselected quantum metrology.

In this paper, we provide such an investigation. First, we give a thorough review of the quantum-Fisher-information matrix (QFIM), which quantifies a metrology protocol's ability to estimate multiple parameters, and we also review previous results in postselected metrology. Then, we find the family of optimal postselection filters for noiseless multiparameter quantum metrology and show that it contains the JAL filter. Next, we analyze and quantify the effect of noise on postselected metrology protocols.

We consider both photon loss and depolarizing noise, applied either before or after the probes have been postselected. Photon loss is an important factor that limits the transmission of a photonic signal in optical fibers across large distances. Depolarizing noise represents a ubiquitous and well-understood model for noise in quantum systems.

We show that photon loss, acting before or after postselection, leads to a decrease of information equal to the fraction of lost photons. The optimal filter remains the same as in the noiseless case.

For depolarizing noise, we consider two regimes: first when postprocessing costs are dominant and one wishes to maximize the information per measured probe and second when detector saturation dominates and one wishes to maximize the rate of information arriving at the detector. We analyze the performance of an experimentally motivated [12] family of filters (including the JAL filter) in both of these regimes. This family, we show, is optimal for qubit probes. When noise acts after postselection, we find that the JAL filter remains optimal in all dimensions. Then, we provide an explicit example demonstrating that the JAL filter can be suboptimal when noise acts before postselection. Finally, we provide a filter that always outperforms the naive strategy of discarding a constant fraction of states, even in the presence of arbitrarily strong noise.

II. PRELIMINARIES

A. Local estimation theory

A typical quantum estimation problem consists of recovering the value of M continuous parameters $\boldsymbol{\theta} := (\theta_1, \theta_2, \dots, \theta_M)$ encoded in a parametrized quantum state $\hat{\rho}_{\boldsymbol{\theta}}$. We focus on the well-established field of local estimation [27], where one considers small deviations in parameters, as opposed to global estimation [28–30] where the entire parameter space is considered. A general scheme consists of the following three steps [2].

(1) Prepare a parameter-independent probe state $\hat{\rho}_0$.

(2) Evolve the state with a parameter-dependent unitary operation $\hat{U}(\boldsymbol{\theta})$:

$$\hat{\rho}_{\boldsymbol{\theta}} = \hat{U}(\boldsymbol{\theta})\rho_0\hat{U}(\boldsymbol{\theta})^{\dagger}. \quad (1)$$

(3) Extract information by means of a suitable measurement. The most general measurement procedure is a positive-operator valued measure (POVM) [1], described by a collection $\{\hat{F}_k\}$ of positive semidefinite operators ($\hat{F}_k \geqslant 0$) that sum to unity ($\sum_k \hat{F}_k = \hat{1}$). One then observes outcome k with probability:

$$p(k|\boldsymbol{\theta}) = \mathrm{Tr}[\hat{F}_k\hat{\rho}_{\boldsymbol{\theta}}]. \quad (2)$$

All the information about the parameters $\boldsymbol{\theta}$ is then encapsulated in the probability distribution $p(k|\boldsymbol{\theta})$.

The parameters are estimated through an estimator $\hat{\boldsymbol{\theta}}(k) := (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_M)$—a map from the space of measurement outcomes to the space of possible values of the parameters. If we are in a limit of small deviations, we assume that the estimator $\hat{\boldsymbol{\theta}}$ is locally unbiased, that is,

$$\sum_k [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(k)]p(k|\boldsymbol{\theta}) = 0, \quad \sum_k \hat{\boldsymbol{\theta}}_i(k)\partial_j p(k|\boldsymbol{\theta}) = \delta_{ij}, \quad (3)$$

where $i, j = 1, \dots, M$ and $\partial_j := \frac{\partial}{\partial \boldsymbol{\theta}_j}$.

The accuracy of $\hat{\boldsymbol{\theta}}(k)$ is quantified by its covariance matrix, given by

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) := \sum_k [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(k)][\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(k)]^{\top} p(k|\boldsymbol{\theta}). \quad (4)$$

The covariance matrix obeys the Cramér-Rao bound (CRB) [31,32]

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geqslant I(\boldsymbol{\theta})^{-1}, \quad (5)$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix (FIM):

$$I(\boldsymbol{\theta})_{i,j} := \sum_k p(k|\boldsymbol{\theta})[\partial_i \ln p(k|\boldsymbol{\theta})][\partial_j \ln p(k|\boldsymbol{\theta})]. \quad (6)$$

In a quantum experiment, the choice of measurement in step (3) affects the probabilities $p(k|\boldsymbol{\theta})$ and hence $I(\boldsymbol{\theta})$. The quantum Fisher information matrix (QFIM) [33–35] is defined by

$$\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})_{i,j} = \mathrm{Tr}[\hat{\Lambda}_i\, \partial_j \hat{\rho}_{\boldsymbol{\theta}}], \quad (7)$$

where $\hat{\Lambda}_i$ is the symmetric logarithmic derivative, implicitly defined by $\partial_i\hat{\rho}_{\boldsymbol{\theta}} = \frac{1}{2}(\hat{\Lambda}_i\hat{\rho}_{\boldsymbol{\theta}} + \hat{\rho}_{\boldsymbol{\theta}}\hat{\Lambda}_i)$ [36]. The inverse QFIM lower bounds the inverse classical Fisher information matrix:

$$I(\boldsymbol{\theta})^{-1} \geqslant \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})^{-1}, \quad (8)$$

for any choice of measurement. In general, Eq. (8) is not saturable. However, there is always a measurement that gets within a factor of 2 of the QFIM in the asymptotic limit: see Ref. [37] for details [38]. The QFIM can thus replace the FIM in Eq. (5), leading to the quantum Cramér-Rao bound (QCRB):

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geqslant \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})^{-1}. \quad (9)$$

In practice, one receives $N > 1$ copies of the quantum state $\hat{\rho}_{\boldsymbol{\theta}}$ and thus has access to the state $\hat{\rho}_{\boldsymbol{\theta}}^{\otimes N}$. One finds that $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\otimes N}) = N\,\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$, implying that

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geqslant \frac{1}{N}\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})^{-1}. \quad (10)$$

Hence one can decrease the variance of the estimate by increasing the number of measurements N or by designing a setup that increases the quantum Fisher information $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$.

Finally, we consider the choice of $\hat{\rho}_0$ in step (1). Since $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$ is convex [39], the maximum QFIM is always achieved by using pure probe states [40,41].

B. Postselection

A common issue in quantum-metrology experiments is that one can create states $\hat{\rho}_{\boldsymbol{\theta}}$ faster than the best detectors can measure them. Thus one must filter, or postselect, a fraction of the states to arrive at the detector. Ideally, the filter should be tuned such that it only lets through a small number of postselected states, each carrying a large information content. We now describe how postselected metrology protocols work. (See Fig. 1 for a schematic overview.)

The encoded state $\hat{\rho}_{\boldsymbol{\theta}}$ is measured with a two-outcome POVM $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$. If the measurement yields outcome $\hat{F}_2$, the state is discarded. If it yields outcome $\hat{F} = \hat{F}_1$, the state is retained. In this way, $\hat{F}$ acts as a filter, where we postselect on passing the filter. The experiment outputs the information-compressed states $\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{ps}} = |\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}\rangle\langle\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}|$ with success probability $P_{\boldsymbol{\theta}}^{\mathrm{ps}}$, where

$$|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}\rangle = \frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{P_{\boldsymbol{\theta}}^{\mathrm{ps}}}}, \quad P_{\boldsymbol{\theta}}^{\mathrm{ps}} = \mathrm{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}], \quad (11)$$

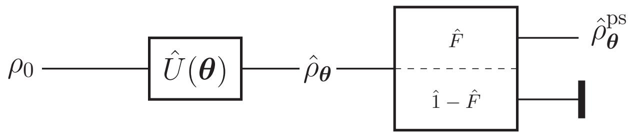

FIG. 1. State $\hat{\rho}_0$ is evolved by the unitary $\hat{U}(\boldsymbol{\theta})$ into $\hat{\rho}_{\boldsymbol{\theta}}$. A postselective measurement $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$ destroys the state, unless outcome $\hat{F}$ happens. The output is the postselected state $\hat{\rho}_{\boldsymbol{\theta}}^{\rm ps}$. This state is finally measured by a detector.

and $\hat{K}$ is a Kraus operator of the generalized measurement used to implement the POVM, i.e., $\hat{F} = \hat{K}^\dagger \hat{K}$. One can check that (see Ref. [26] for details)

$$\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\rm ps}\big)_{i,j} = 4{\rm Re}\bigg[\frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - \frac{1}{(P_{\boldsymbol{\theta}}^{\rm ps})^2}\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle\bigg]. \quad (12)$$

References [12,26] introduce a filter $\hat{F}$ that can arbitrarily compress metrological information in the absence of noise: let $\boldsymbol{\theta}_0$ denote an initial estimate of the true parameters of interest $\boldsymbol{\theta}$ and let $\boldsymbol{\delta} := \boldsymbol{\theta} - \boldsymbol{\theta}_0$. The JAL filter is

$$\hat{F} = (t^2 - 1)\hat{\rho}_{\boldsymbol{\theta}_0} + \hat{1}, \quad (13)$$

where $t \in [0, 1]$. Substituting the JAL filter in Eq. (12) gives [26]

$$\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\rm ps}\big)_{i,j} = \frac{1}{t^2}\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})_{i,j} + O(|\boldsymbol{\delta}|^2/t^2), \quad (14)$$

$$P_{\boldsymbol{\theta}}^{\rm ps} = t^2 + O(|\boldsymbol{\delta}|^2/t^2). \quad (15)$$

Physically, the JAL filter transmits the expected state $\hat{\rho}_{\boldsymbol{\theta}_0}$ with probability t^2 and always transmits any state orthogonal to $\hat{\rho}_{\boldsymbol{\theta}_0}$. Hence information can be distilled by choosing a small t^2, so that only the states orthogonal to the expected state are transmitted. The maximum information amplification is unbounded in the limit $t^2 \to 0$, provided that $|\boldsymbol{\delta}|^2 \ll t^2$. The JAL filter in Eq. (13) increases all of the entries of the QFIM by a factor of $1/t^2$. Remarkably, this protocol is lossless in the limit $\boldsymbol{\delta} \to 0$:

$$P_{\boldsymbol{\theta}}^{\rm ps}\,\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\rm ps}\big) = \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}) + O(|\boldsymbol{\delta}|^2). \quad (16)$$

III. OPTIMAL FILTER FOR NOISELESS POSTSELECTION

In this section, we characterize the most general optimal filter for noiseless multiparameter quantum metrology. Then, we show that the JAL filter is in the family of optimal filters; it is the canonical choice.

Suppose that we have a process that produces a state $\hat{\rho}_{\boldsymbol{\theta}}^{(2)}$ from a state $\hat{\rho}_{\boldsymbol{\theta}}^{(1)}$ with probability $P_{\boldsymbol{\theta}}^{\rm ps}$, for example, by postselection. The process is said to uniformly amplify the information contained in $\hat{\rho}_{\boldsymbol{\theta}}^{(1)}$ if, for all i and j, the ratio

$$\mathcal{A}_{i,j}\big(\hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)}\big) := \frac{\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(2)}\big)_{i,j}}{\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(1)}\big)_{i,j}} \quad (17)$$

is the same. In this case, we drop the i, j label in $\mathcal{A}$ and call $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)})$ the *information amplification*. Note by the data processing inequality [42] that

$$P_{\boldsymbol{\theta}}^{\rm ps}\mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(2)}\big) \leqslant \mathcal{I}\big(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(1)}\big) \quad (18)$$

and thus $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)}) \leqslant 1/P_{\boldsymbol{\theta}}^{\rm ps}$. Thus we also define the *compression efficiency* (which appears in [43]),

$$\eta\big(P_{\boldsymbol{\theta}}^{\rm ps}, \hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)}\big) := P_{\boldsymbol{\theta}}^{\rm ps}\,\mathcal{A}\big(\hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)}\big). \quad (19)$$

The compression efficiency equals the ratio of the total expected information content before and after the process. If $\eta(P_{\boldsymbol{\theta}}, \hat{\rho}_{\boldsymbol{\theta}}^{(2)}, \hat{\rho}_{\boldsymbol{\theta}}^{(1)}) = 1$, the process is said to be lossless; otherwise, it is said to be lossy. For a chosen postselection probability $P_{\boldsymbol{\theta}}^{\rm ps}$, a filter $\hat{F}$ is optimal if, in the limit as $\boldsymbol{\delta} \to 0$, postselection using $\hat{F}$ is lossless. By the above, for a given postselection probability $P_{\boldsymbol{\theta}}^{\rm ps}$, the optimal filter gives the maximum possible information amplification.

Instead of insisting that information amplification be *uniform*, one could measure amplification by minimizing the scalar quantity: ${\rm Tr}[\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(2)})^{-1}W]$ for some positive semidefinite weight matrix $W \geqslant 0$. By Eq. (9), this gives a lower bound on ${\rm Tr}[W{\rm Cov}(\hat{\boldsymbol{\theta}})]$. However, if amplification is not uniform, then Eq. (18) implies $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(2)})^{-1} > P_{\boldsymbol{\theta}}^{\rm ps}\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(1)})^{-1}$. Thus (as the Fisher information matrix is symmetric) there exists a vector $\boldsymbol{u}$ such that $\boldsymbol{u}^T\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(2)})^{-1}\boldsymbol{u} > P_{\boldsymbol{\theta}}^{\rm ps}\boldsymbol{u}^T\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{(1)})^{-1}\boldsymbol{u}$. Taking $W = \boldsymbol{u}\boldsymbol{u}^T$ shows that uniform amplification is needed if optimal performance is required for any choice of W.

Theorem. Suppose that one filters the state $|\psi_{\boldsymbol{\theta}}\rangle$ with a two-outcome POVM $\{\hat{K}^\dagger\hat{K}, \hat{1} - \hat{K}^\dagger\hat{K}\}$. Suppose further that $\mathcal{I}_{i,i}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) \neq 0$, for every $i = 1, \ldots, M$ (so that there is information to compress). Let $\mathcal{U} = {\rm span}\{|\psi_{\boldsymbol{\theta}}\rangle, |\partial_i\psi_{\boldsymbol{\theta}}\rangle : i = 1, \ldots, M\}$ and $\hat{\Pi}_u$ the orthogonal projection onto $\mathcal{U}$ [44]. Then, (1) the postselected Fisher information depends only on $\hat{F}_u := \hat{\Pi}_u\hat{K}^\dagger\hat{K}\hat{\Pi}_u$ (and the state before the filter) and (2) for a fixed postselection probability $P_{\boldsymbol{\theta}}^{\rm ps}$, the POVM is optimal iff $\hat{F}_u = (P_{\boldsymbol{\theta}}^{\rm ps} - 1)|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}| + \hat{\Pi}_u$.

Proof. See Appendix A.

The simplest example of an optimal postselection filter is the JAL filter:

$$\hat{F} = (t^2 - 1)\hat{\rho}_{\boldsymbol{\theta}} + \hat{1}. \quad (20)$$

Note that one does not need to know $\mathcal{U}$ to implement the JAL filter, making it the canonical choice.

Finally, let us discuss intuitively what makes a filter optimal. Provided that $\boldsymbol{\delta}$ is small, the states $|\psi_{\boldsymbol{\theta}_0}\rangle$ contribute vanishingly small information and should therefore be filtered away. This is achieved by setting $P_{\boldsymbol{\theta}}^{\rm ps} \to 0$, for $\boldsymbol{\theta} \sim \boldsymbol{\theta}_0$. In this limit, all the information is carried by the states in the subspace $|\psi_{\boldsymbol{\theta}_0}\rangle^\perp$. If the dimension of $\mathcal{U}$ satisfies $\dim\mathcal{U} \equiv u < d$, we can locally find a basis in which only u dimensions are parameter dependent, while $d - u$ are parameter independent. Therefore, the optimal postselection filter is unique only within the u dimensional subspace. It can be arbitrary outside of that subspace. Because the information is contained only within the u-dimensional subspace, if the filter is to be lossless, it should let through all the states orthogonal to $|\psi_{\boldsymbol{\theta}_0}\rangle$.

IV. NOISY POSTSELECTION

We now consider the effect of noise on postselected metrology. In the following, we analyze two important models of noise. One is photon loss and the other is depolarizing noise.

Photon loss can be modeled with the following channel:

$$L[\hat{\rho}_{\boldsymbol{\theta}}] := \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{l}} = T\,\hat{\rho}_{\boldsymbol{\theta}} + (1-T)\,\hat{\rho}_{\mathrm{vac}}, \tag{21}$$

i.e., with probability T we recover the original state and with probability $(1-T)$ the original state is destroyed, leaving the vacuum state. Hence, if photon loss occurs after the postselection filter, the number of detected photons is simply reduced by a factor of T. Therefore, the information available to the experimenter also decreases by a factor of T.

Let us now consider the effect of photon loss before the postselection filter. First, we note that the vacuum state does not contribute to detector saturation. Further, the vacuum state carries no information about $\boldsymbol{\theta}$. Hence $\hat{\rho}_{\mathrm{vac}}$ can be discarded without any loss of information. Therefore, we choose our filter such that $\hat{K}|\mathrm{vac}\rangle = 0$. The postselection probability is

$$\begin{aligned} P_{\boldsymbol{\theta}}^{\mathrm{ps}} &= \mathrm{Tr}[T\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger} + (1-T)\,\hat{K}\hat{\rho}_{\mathrm{vac}}\hat{K}^{\dagger}] \\ &= \mathrm{Tr}[T\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger}] \\ &= T\;\mathrm{Tr}[\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger}]. \end{aligned}$$

The postselected density matrix is

$$\begin{aligned} \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{l,\,ps}} &= \frac{1}{P_{\boldsymbol{\theta}}^{\mathrm{ps}}}[T\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger} + (1-T)\,\hat{K}\hat{\rho}_{\mathrm{vac}}\hat{K}^{\dagger}] \\ &= \frac{\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger}}{\mathrm{Tr}[\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger}]} := \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{ps}}. \end{aligned}$$

Hence the postselected state is the same as in the case of no photon loss, but the postselection probability is decreased by a factor of T. Therefore, the rate of information arriving at the detector is also reduced by a factor of T. The optimal filter is still contained in the family of optimal noiseless filters. In particular, the filter's amplification is still unbounded. The filter is also lossless when compared to the noisy, prefiltered state: $\eta(P_{\boldsymbol{\theta}}^{\mathrm{ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{l,\,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{l}}) = 1$.

Depolarizing noise is more complicated. The action of the depolarizing channel $D[\cdot]$ on a state $\hat{\rho}_{\boldsymbol{\theta}}$ can be written as

$$D[\hat{\rho}_{\boldsymbol{\theta}}] := \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{n}} = (1-\epsilon)\hat{\rho}_{\boldsymbol{\theta}} + \frac{\epsilon}{d}\,\hat{1}, \tag{22}$$

where $0 \leqslant \epsilon \leqslant 1$ sets the strength of the noise. Below we consider two scenarios: noise acting before or after postselection. We thus consider the two states

$$\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{ps,\,n}} = \frac{1}{(1-\epsilon)P_{\boldsymbol{\theta}}^{\mathrm{ps}} + \epsilon}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger} + \frac{\epsilon}{d}\hat{1}\right], \tag{23}$$

$$\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{n,\,ps}} = \frac{1}{P_{\boldsymbol{\theta}}^{\mathrm{ps}}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^{\dagger} + \frac{\epsilon}{d}\hat{K}\hat{K}^{\dagger}\right], \tag{24}$$

in consecutive sections.

A. Noise after postselection

Let us first examine the case of noise acting after postselection, as depicted in Fig. 2. In Appendix C 1, we show that the criterion for optimality is unchanged from the noiseless case.

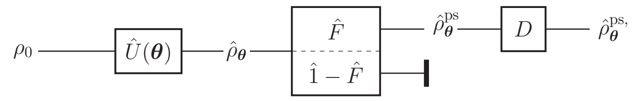

FIG. 2. Depolarizing channel D is placed after the encoded state $\hat{\rho}_{\boldsymbol{\theta}}$ is filtered by the postselective measurement $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$.

Hence the JAL filter is optimal. We calculate the information amplification for the JAL filter $\hat{F}$:

$$\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{ps,\,n}}, \hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{n}}) = \frac{1}{t^2} + O(|\boldsymbol{\delta}|^2), \tag{25}$$

where t^2 was defined in Eq. (13), and, in this case, is equal to the probability of postselection, $P_{\boldsymbol{\theta}}^{\mathrm{ps}}$. Therefore, when comparing an experiment with depolarizing noise acting after the postselection to an experiment with the same noise but no postselection, we see that the JAL filter leads to no additional loss of information. In this case, postselection is still lossless and the optimal noisy filter is unchanged from the noiseless scenario.

B. Noise before postselection

We now consider a depolarizing channel acting before the postselection filter (Fig. 3). In Appendix C 2, we calculate the QFIM of $\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{n,\,ps}}$ for a family of filters that are closely related to the optimal noiseless filters:

$$\hat{F} = (p_{\boldsymbol{\theta}} - B)|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}| + B\,\hat{\Pi}_u + D\,\hat{\Pi}_n, \tag{26}$$

where $p_{\boldsymbol{\theta}}, B, D \in [0,1]$, $\hat{\Pi}_u$ is the orthogonal projection onto $\mathcal{U}$, and $\hat{\Pi}_n$ is the orthogonal projection onto $\mathcal{U}^{\perp}$. This experimentally motivated family [12] lends itself to analytical analysis and incorporates the JAL filter. In Appendix D 1 we show that, for qubit sensors ($u = d = 2$), this family is optimal.

Intuitively, we split the action of $\hat{F}$ on the useful subspace $\hat{\Pi}_u$ (containing the state $|\psi_{\boldsymbol{\theta}}\rangle$ and its derivatives) and the orthogonal subspace $\hat{\Pi}_n$, which contains no information on $\boldsymbol{\theta}$. While in the noiseless case any choice of D is still optimal, in the noisy case, the $\hat{\Pi}_n$ subspace only contains noisy probe states. As explained later in this section, we therefore set $D = 0$.

However, in Appendix D 2, we show that the aforementioned family of filters is not always optimal. We construct a specific example where $u = 2$ and $d = 3$. In our example it is advantageous to choose $\hat{F}$ such that $|\psi_{\boldsymbol{\theta}}\rangle$ is not an eigenstate of $\hat{F}$. Then, states corresponding to noise are mixed with states corresponding to changes in $\boldsymbol{\theta}$.

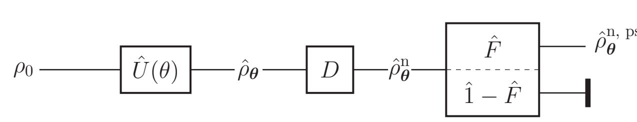

FIG. 3. Depolarizing channel D is placed before the encoded state $\hat{\rho}_{\boldsymbol{\theta}}$ is filtered by the postselective measurement $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$.

1. Analysis of information amplification and compression efficiency

In Appendix C 2 we calculate the information amplification for the family of filters in Eq (26):

$$\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}}) = \frac{\left(1-\epsilon+2\frac{\epsilon}{d}\right)\frac{p_{\boldsymbol{\theta}}}{B}}{\left(1-\epsilon+\frac{\epsilon}{d}\right)\frac{p_{\boldsymbol{\theta}}}{B}+\frac{\epsilon}{d}\left[(u-1)+\frac{D}{B}(d-u)\right]} \times \frac{1}{\left(1-\epsilon+\frac{\epsilon}{d}\right)\frac{p_{\boldsymbol{\theta}}}{B}+\frac{\epsilon}{d}}. \tag{27}$$

The compression efficiency is obtained by multiplying the above equation by $P_{\boldsymbol{\theta}}^{\text{ps}}$:

$$\eta\left(P_{\boldsymbol{\theta}}^{\text{ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}}\right) = \frac{\left(1-\epsilon+2\frac{\epsilon}{d}\right)p_{\boldsymbol{\theta}}}{\left(1-\epsilon+\frac{\epsilon}{d}\right)\frac{p_{\boldsymbol{\theta}}}{B}+\frac{\epsilon}{d}}. \tag{28}$$

The criteria for a filter to be optimal is decided by the experimental setup. Below, we consider two possible experimental regimes. First, suppose that postprocessing is the limiting factor. In this case, one wants to receive the most information from the smallest number of probe measurements, which corresponds to maximizing the information amplification [Eq. (27)]. We work in the limit in which the quantum experiment is cheap to run, so that the postselection probability can be made arbitrarily small; one simply repeats the experiment as many times as is necessary to produce enough successfully postselected states to reduce the error below a set value.

Alternatively, suppose that detector saturation is the limiting factor. In this case, one fixes a maximum postselection probability $P_{\max}$. Ideally, one would set $P_{\max}$ equal to the ratio between the maximum-probe-measurement rate and the maximum-probe-production rate. Then, one maximizes the compression efficiency [Eq. (28)] such that the postselection probability is no greater than $P_{\max}$. This maximizes the rate of information arriving at the detector, while ensuring that it will not saturate.

In the first case, $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}})$ is maximized for $D = 0$. In the second case, the compression efficiency is independent of D. Setting $D = 0$ gives the lowest postselection probability, which avoids detector saturation. Therefore, in both regimes the best choice is $D = 0$. Denoting $p_{\boldsymbol{\theta}}/B = t^2$, Eq. (27) becomes

$$\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}}) = \frac{\left(1-\epsilon+2\frac{\epsilon}{d}\right)t^2}{\left[\left(1-\epsilon+\frac{\epsilon}{d}\right)t^2+\frac{\epsilon}{d}(u-1)\right]\left[\left(1-\epsilon+\frac{\epsilon}{d}\right)t^2+\frac{\epsilon}{d}\right]}, \tag{29}$$

which is parametrized by t^2. We now proceed to maximize this expression.

2. Postprocessing is dominant

In the case where postprocessing carries the largest overhead, we are interested in maximizing the information amplification, given by Eq. (29). This expression attains a maximum at

$$t_{\text{pp}}^2 = \frac{\sqrt{u-1}\,\epsilon}{d(1-\epsilon)+\epsilon}. \tag{30}$$

In Fig. 4(a), we plot $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}})$ vs t^2 for different values of ϵ and $d = u = 2$ (noting that $t_{\text{pp}} < 1$). We see that the information amplification reaches a maximum at $t = t_{\text{pp}}$. The maximum of $\mathcal{A}$ increases and t_{pp} moves towards zero as ϵ decreases. Figure 4(b) shows the same plot, but for $d = 10$, $u = 5$. In this case, the optimal t_{pp} can be greater than 1.

Substituting t_{pp}^2 into Eq. (30), and taking the limit $d \to \infty$, with $u \sim d$, the maximum information amplification simplifies to $1/\epsilon$:

$$\lim_{d\to\infty} \max \mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}}) = \frac{1}{\epsilon}. \tag{31}$$

Any filter with $t = t_{\text{pp}}$ will achieve the maximum information compression. Additionally, it is sensible to minimize the (expected) required number of probes, which corresponds to maximizing $P_{\boldsymbol{\theta}}^{\text{ps}}$ for a fixed $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}}^{\text{n}})$. Rewriting

$$P_{\boldsymbol{\theta}}^{\text{ps}} = B\left[\left(1-\epsilon+\frac{\epsilon}{d}\right)t_{\text{pp}}^2+\frac{\epsilon}{d}(u-1)\right], \tag{32}$$

we see that maximizing $P_{\boldsymbol{\theta}}^{\text{ps}}$ corresponds to setting B to its maximum possible value. Thus we deduce the optimal form of $\hat{F}$ to mitigate postprocessing costs:

$$p_{\boldsymbol{\theta}} = \min\left[1, t_{\text{pp}}^2\right], \quad B = \min\left[1, \frac{1}{t_{\text{pp}}^2}\right]. \tag{33}$$

In Fig. 4(c), we plot the maximum information amplification ($t = t_{\text{pp}}$) against ϵ at different values of d, u.

When compared to an experiment with the same level of noise, our postselection filter can sometimes perform better for stronger depolarizing noise. As noise reduces the total available information, this does not mean that it would be beneficial to artificially increase ϵ. In Fig. 4(d) we plot the maximum information amplification $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{n,ps}}, \hat{\rho}_{\boldsymbol{\theta}})$, this time evaluated with respect to the noiseless state. As expected, overall our scheme performs worse with increasing noise.

3. Detector saturation is dominant

Here, we want to maximize the compression efficiency [Eq. (28)], subject to the constraint that the postselection probability $P_{\boldsymbol{\theta}}^{\text{ps}}$ is no greater than $P_{\max} \in [0, 1]$. We calculate this maximum in Appendix E. For notational brevity, we define

$$b = \left(1-\epsilon+\frac{\epsilon}{d}\right), \quad c = \frac{\epsilon}{d}(u-1). \tag{34}$$

Mathematically, the optimal filter is most succinctly described by the following five (counting min and max) cases.

(i) $P_{\max} \geqslant b + c$. Then the optimal filter has $p_{\boldsymbol{\theta}} = B = 1$.

(ii) $P_{\max} < b + c$. Letting $t^2 = p_{\boldsymbol{\theta}}/B$, the optimal filter has

$$p_{\boldsymbol{\theta}} = \frac{P_{\max}}{b + c/t^2}, \quad B = \frac{P_{\max}}{bt^2 + c}. \tag{35}$$

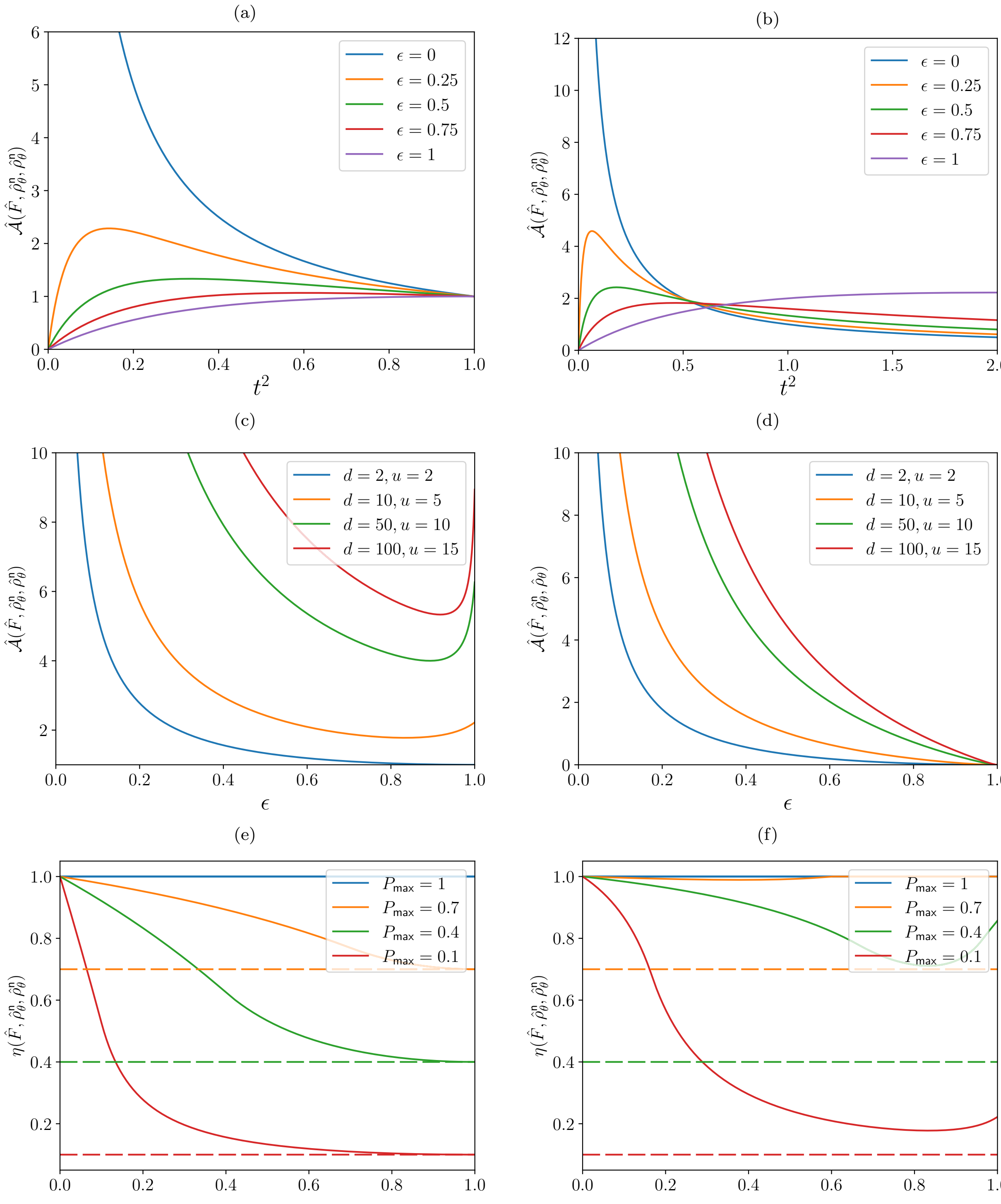

FIG. 4. (a) Information amplification vs t^2 for $d = u = 2$. (b) $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}}^{\rm n})$ vs t^2 for $d = 10$ and $u = 5$. (c) Maximum $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}}^{\rm n})$ vs ϵ for different values of d and u. $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}}^{\rm n})$ reaches a minimum at $\epsilon < 1$, then increases to d/u at $\epsilon = 1$. When compared to an experiment with the same level of noise, our filter performs better for large noise. (d) Maximum $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}})$ vs ϵ for various values of d and u. $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}})$ is monotonically decreasing with ϵ, reaching $\hat{\mathcal{A}}(\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,ps}, \hat{\rho}_{\boldsymbol{\theta}}) = 0$ at $\epsilon = 1$. As expected, information amplification (with respect to a noiseless experiment) decreases with stronger noise. (e) Dashed lines: compression efficiency for the naive classical filter $\hat{F} = P_{\max}\hat{1}$, for $d, u = 2$. Solid lines: compression efficiency for the optimal detector saturation filter. Our filter always performs better than the naive classical filter. As $\epsilon \to 0$ our filter can compress all of the available information into an arbitrarily small number of states. (f) Same as (e), but with $d = 10$ and $u = 5$. Our filter compresses all the available information for $P_{\max} > b + c$. For $P_{\max} < b + c$, even as $\epsilon = 1$, our filter performs better than the naive classical filter, by a factor of d/u.

There are two subcases depending on the value of t_{pp}.

(1) $t_{pp} \leqslant 1$. Then, the optimal filter has

$$t^2 = \max\left[t_{pp}^2, \frac{P_{max} - c}{b}\right]. \quad (36)$$

(2) $t_{pp} \geqslant 1$. Then, the optimal filter has

$$t^2 = \min\left[t_{pp}^2, \left(\frac{c}{P_{max} - b}\right)^+\right], \quad (37)$$

where $a^+ = a$ if $a \geqslant 0$ and $a^+ = \infty$ if $a < 0$.

Physically, the optimal filter can fall into one of three distinct categories depending on the size of P_{max}. The three categories arise from grouping the cases above into optimal filters with common behaviors. To describe the categories, we introduce P_* as the largest value of P_{max} for which the optimal filter has t^2 equal to t_{pp}^2. We find the following three different categories.

(i) $b + c = 1 - \epsilon + u\epsilon/d \leqslant P_{max}$, whereupon $p_{\boldsymbol{\theta}} = B = 1$. In this scenario, the filter simply blocks all states that can only arise from noise and lets through all the states that carry information about the parameters. This regime is lossless (when compared to the information carried in $\hat{\rho}_{\boldsymbol{\theta}}^{n}$).

(ii) $P_* \leqslant P_{max} < b + c = 1 - \epsilon + u\epsilon/d$. In this scenario, the filter compresses information. Since compression with noise is lossy, it does the smallest amount of compression possible, i.e., such that $P_{\boldsymbol{\theta}}^{ps} = P_{max}$. It does this by maximizing the information amplification, which corresponds to setting t as close as possible to t_{pp} while still satisfying $P_{\boldsymbol{\theta}}^{ps} = P_{max}$. In this regime, the filter incurs some loss due to noise.

(iii) $0 \leqslant P_{max} < P_*$. When P_{max} reaches P_*, it is no longer advantageous to compress information. Instead, $p_{\boldsymbol{\theta}}/B = t_{pp}^2$ stays constant while $P_{\boldsymbol{\theta}}^{ps} = P_{max}$, so that both $p_{\boldsymbol{\theta}}$ and B decrease. The optimal filter can be decomposed as a compressive filter with $t = t_{pp}$ followed by a filter that is proportional to $\hat{1}$. In this scenario, the filter compresses some of the information, but is also forced to blindly discard a constant fraction of the states.

We can compare our strategy with a naive classical strategy that would blindly discard a constant fraction of the states:

$$\hat{F}_{nav} = P_{max}\,\hat{1}. \quad (38)$$

This filter gives a compression efficiency of P_{max}. In Fig. 4(e), we plot the compression efficiency against noise for different values of P_{max}, at $d = u = 2$. For small noise ($\epsilon \to 0$), our postselection strategy can compress information into an arbitrarily small number of measurement probes, while preserving all of the available information. As expected, this advantage decreases and becomes vanishingly small as $\epsilon \to 1$. In this limit, the performance of our filter reduces to that of the naive classical strategy, which randomly keeps a fraction P_{max} of the input probes. In Fig. 4(f), we repeat the same calculation for the cases $d = 10$ and $u = 5$. This time, if $P_{max} > b + c$, it is possible to preserve all of the available information, even as $\epsilon \to 1$. For $P_{max} < b + c$, the situation is similar to the case $d = u = 2$, but our postselection strategy now always performs better than the naive classical strategy.

V. CONCLUSION

In this paper we derived the family of optimal postselection filters for noiseless multiparameter quantum metrology. In the absence of noise, these filters are lossless: they compress information equally across all parameters, while decreasing the postselection probability by the same factor. We showed that the previously proposed JAL filter [26] is contained within our family of optimal filters.

Noise in real experiments leads to natural limits on compression of information. The quantum optics experiment in Ref. [12] was affected by errors in t^2, the filter transmission probability, and the unitary operation $\hat{U}(\boldsymbol{\theta})$, which bounded the maximum possible compression of information and made the filtering procedure somewhat lossy. This motivated our analysis of the effect of noise on postselected quantum metrology.

We focused our noise analysis on the worst-case scenario of depolarizing noise of strength ϵ. We considered situations where this noise was applied either before or after the probes have been postselected. When noise acts after the postselection, we found that the JAL filter remains optimal.

We also analyzed the JAL filter's performance when noise acts before the postselection. In this case, we considered two regimes, whereby either postprocessing or detector saturation is the main concern. To make the problem tractable, we considered a subset of experimentally sensible filters [Eq. (C17)]. To alleviate postprocessing costs, one wants to receive the most information in the smallest number of probes, which corresponds to maximizing the information amplification [Eq. (27)]. To protect against detector saturation, we fixed a maximum postselection probability P_{max} and maximized the compression efficiency [Eq. (28)]. We showed that the JAL filter is not always optimal, but with slight modifications it performs well in both the postprocessing and detector saturation regimes. Thus, if one can estimate the strength of noise in an experiment, one should implement a different filter compared to the case of no noise.

We showed that for qubit sensors ($u = d = 2$) the family of filters in Eq. (26) is optimal. However, for finite $u, d \neq 2$, we show that this subset of filters is not always optimal. In fact, it can sometimes be advantageous for states corresponding to noise to be mixed with states corresponding to changes in $\boldsymbol{\theta}$.

With this work, we hope to extend the use cases of postselected metrology to nonoptical systems.

ACKNOWLEDGMENTS

The authors wish to thank N. Y. Halpern, N. Gladstein, and Y. B. Yilmaz for their useful discussions and comments. F.S. was supported by the Harding Foundation, W.S. was supported by the EPSRC (Engineering and Physical Sciences Research Council) and Hitachi, and D.R.M.A.-S. was supported by the Lars Hierta's Memorial Foundation and Girton College.

APPENDIX A: PROOF OF OPTIMAL NOISELESS FILTER

In this Appendix we prove the main Theorem from Sec. III.

From Eq. (12), we see that the QFIM only contains terms that depend on $\hat{F}_u$ and thus we deduce part 1 of the Theorem.

We now focus on the diagonal entries of the QFIM: $\mathcal{I}_{j,j}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{ps})$. Taking the derivative of $\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle \equiv 1$, we see that

$$\mathrm{Re}(\langle\partial_j\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle) = 0. \quad (A1)$$

Thus we can write

$$|\partial_j \psi_\theta\rangle = i x_j |\psi_\theta\rangle + \alpha_j |\psi_\theta{}^{\perp,j}\rangle, \tag{A2}$$

for $x_j \in \mathbb{R}$ a *real* coefficient, $\alpha_j \in \mathbb{C}$ and some normalized set of $|\psi^{\perp,j}\rangle$, orthogonal to $|\psi_\theta\rangle$ (note that they may not be orthogonal to each other). We define

$$\langle\psi_\theta|\hat{F}_u|\psi_\theta\rangle = P_\theta^{\text{ps}} \in [0,1], \tag{A3}$$

$$\langle\psi_\theta{}^{\perp,j}|\hat{F}_u|\psi_\theta{}^{\perp,j}\rangle = B_j \in [0,1], \tag{A4}$$

$$\langle\psi_\theta|\hat{F}_u|\psi_\theta{}^{\perp,j}\rangle = C_j \in \mathbb{C}. \tag{A5}$$

Suppose that $\hat{F}_u$ is optimal. Substituting these expressions into Eq. (12), we find

$$\mathcal{I}_{j,j}(\theta|\hat{\rho}_\theta^{\text{ps}}) = \frac{4}{P_\theta^{\text{ps}}}\left(x_j^2 P_\theta^{\text{ps}} + |\alpha_j|^2 B_j + i x_j(\alpha_j^* C_j^* - \alpha_j C_j)\right) - \frac{4}{(P_\theta^{\text{ps}})^2}\left|i x_j P_\theta^{\text{ps}} + \alpha_j C_j\right|^2, \tag{A6}$$

$$= 4\left[\frac{|\alpha_j|^2 B_j}{P_\theta^{\text{ps}}} - \frac{|\alpha_j C_j|^2}{(P_\theta^{\text{ps}})^2}\right]. \tag{A7}$$

Setting $\hat{F}_u = \hat{1}$, we have

$$\mathcal{I}_{j,j}(\theta|\hat{\rho}_\theta) = 4|\alpha_j|^2. \tag{A8}$$

By assumption $\mathcal{I}_{j,j}(\theta|\hat{\rho}_\theta) \neq 0$; therefore, it follows that

$$\frac{\mathcal{I}_{j,j}(\theta|\hat{\rho}_\theta^{\text{ps}})}{\mathcal{I}_{j,j}(\theta|\hat{\rho}_\theta)} = \left[\frac{B_j}{P_\theta^{\text{ps}}} - \frac{|C_j|^2}{(P_\theta^{\text{ps}})^2}\right]. \tag{A9}$$

Note that optimality of $\hat{F}$ is equivalent to $\mathcal{A}(\hat{\rho}_\theta^{\text{ps}}, \hat{\rho}_\theta) = 1/P_\theta^{\text{ps}}$. Thus Eq. (A9) implies that $B_j = 1$ and $C_j = 0$ (recalling that $B_j \in [0,1]$, since $0 \leqslant \hat{F} \leqslant \hat{1}$).

Now suppose that $\hat{F}|\psi_\theta^{\perp,j}\rangle \neq |\psi_\theta^{\perp,j}\rangle$ for some j. Since $B_j = 1$ by above, this implies that there is some state $|\alpha\rangle$ satisfying $\langle\psi_\theta^{\perp,j}|\alpha\rangle = 0$ and a non-negative $a > 0$ such that

$$\hat{F}\left|\psi_\theta^{\perp,j}\right\rangle = \left|\psi_\theta^{\perp,j}\right\rangle + a|\alpha\rangle. \tag{A10}$$

Letting $|\phi\rangle = (1/a)|\psi_\theta^{\perp,j}\rangle + (1/2)|\alpha\rangle$, we calculate

$$\langle\phi|\hat{F}|\phi\rangle = \frac{1}{a}\langle\psi_\theta^{\perp,j}|\hat{F}\left(\frac{1}{a}\left|\psi_\theta^{\perp,j}\right\rangle + \frac{1}{2}|\alpha\rangle\right) + \frac{1}{2}\langle\alpha|\hat{F}\left(\frac{1}{a}\left|\psi_\theta^{\perp,j}\right\rangle + \frac{1}{2}|\alpha\rangle\right), \tag{A11}$$

$$= \frac{1}{a^2} + 1 + \frac{1}{4}\langle\alpha|\hat{F}|\alpha\rangle, \tag{A12}$$

$$> \frac{1}{a^2} + \frac{1}{4} \tag{A13}$$

$$= \langle\phi| =\rangle\langle\phi|\hat{1}|\phi\rangle, \tag{A14}$$

which contradicts $\hat{F} \leqslant \hat{1}$. We deduce that $\hat{F}|\psi_\theta^{\perp,j}\rangle = |\psi_\theta^{\perp,j}\rangle$ for every j.

Therefore, an optimal $\hat{F}$ must have

$$\hat{F}_u = \left(P_\theta^{\text{ps}} - 1\right)|\psi_\theta\rangle\langle\psi_\theta| + \hat{\Pi}_u. \tag{A15}$$

It remains to check that such a filter is indeed optimal. Equation (A9) shows that it correctly amplifies diagonal terms of the QFIM. Next we check the off-diagonal terms.

Using Eq. (12), we have that

$$\mathcal{I}\left(\theta|\hat{\rho}_\theta^{\text{ps}}\right)_{i,j} = 4\text{Re}\left[\frac{1}{P_\theta^{\text{ps}}}\langle\partial_i\psi_\theta|\hat{F}_u|\partial_j\psi_\theta\rangle - \frac{1}{\left(P_\theta^{\text{ps}}\right)^2}\langle\partial_i\psi_\theta|\hat{F}_u|\psi_\theta\rangle\langle\psi_\theta|\hat{F}_u|\partial_j\psi_\theta\rangle\right]. \tag{A16}$$

Substituting Eq. (A2) for $|\partial_j\psi_\theta\rangle$,

$$|\partial_j \psi_\theta\rangle = i x_j |\psi_\theta\rangle + \alpha_j |\psi_\theta{}^{\perp,j}\rangle, \tag{A17}$$

one can calculate

$$\mathcal{I}_{j,k}\left(\theta|\hat{\rho}_\theta^{\text{ps}}\right) = 4\,\text{Re}\left[\frac{1}{P_\theta^{\text{ps}}}\langle\partial_j\psi_\theta|\hat{F}_u|\partial_k\psi_\theta\rangle - \frac{1}{\left(P_\theta^{\text{ps}}\right)^2}\langle\partial_j\psi_\theta|\hat{F}_u|\psi_\theta\rangle\langle\psi_\theta|\hat{F}_u|\partial_k\psi_\theta\rangle\right], \tag{A18}$$

$$= \frac{4}{P_\theta^{\text{ps}}}\text{Re}[(-i x_j\langle\psi_\theta| + \alpha_j^*\langle\psi_\theta{}^{\perp,j}|)\hat{F}_u(i x_k|\psi_\theta\rangle + \alpha_k|\psi_\theta{}^{\perp,k}\rangle)] - \frac{4}{\left(P_\theta^{\text{ps}}\right)^2}\text{Re}[(-i x_j\langle\psi_\theta| + \alpha_j\langle\psi_\theta{}^{\perp,j}|)\hat{F}_u|\psi_\theta\rangle \times \langle\psi_\theta|\hat{F}_u(i x_k|\psi_\theta\rangle + \alpha_k|\psi_\theta{}^{\perp,k}\rangle)], \tag{A19}$$

$$= \frac{4}{P_\theta^{\text{ps}}}\text{Re}\left[x_j x_k P_\theta^{\text{ps}} + \alpha_j^*\alpha_k\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta{}^{\perp,k}\rangle - i x_j\alpha_k\langle\psi_\theta|F|\psi_\theta{}^{\perp,k}\rangle + i x_k\alpha_j^*\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta\rangle\right] - \frac{4}{\left(P_\theta^{\text{ps}}\right)^2}\text{Re}\left[x_k x_k\left(P_\theta^{\text{ps}}\right)^2 - i x_j\alpha_k P_\theta^{\text{ps}}\langle\psi_\theta|F|\psi_\theta{}^{\perp,k}\rangle + i x_k\alpha_j^* P_\theta^{\text{ps}}\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta\rangle + \alpha_j^*\alpha_k\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta\rangle\langle\psi_\theta|F|\psi_\theta{}^{\perp,k}\rangle\right], \tag{A20}$$

$$= \frac{4}{P_\theta^{\text{ps}}}\text{Re}\left[\alpha_j^*\alpha_k\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta{}^{\perp,k}\rangle - \frac{\alpha_j^*\alpha_k}{P_\theta^{\text{ps}}}\langle\psi_\theta{}^{\perp,j}|F|\psi_\theta\rangle\langle\psi_\theta|F|\psi_\theta{}^{\perp,k}\rangle\right], \tag{A21}$$

$$= \frac{4\,\text{Re}\left(\alpha_j^*\alpha_k\langle\psi_\theta^{\perp,j}|\psi_\theta^{\perp,k}\rangle\right)}{P_\theta^{\text{ps}}}, \tag{A22}$$

$$= \frac{\mathcal{I}_{j,k}(\theta|\hat{\rho}_\theta)}{P_\theta^{\text{ps}}}. \tag{A23}$$

In line (A20), we used Eq. (A3) to write $\langle\psi_{\boldsymbol{\theta}}|\hat{F}_u|\psi_{\boldsymbol{\theta}}\rangle = P_{\boldsymbol{\theta}}^{\text{ps}}$ and in line (A22) we used the fact that $\hat{F}|\psi_{\boldsymbol{\theta}}^{\perp,j}\rangle = |\psi_{\boldsymbol{\theta}}^{\perp,j}\rangle$ for every j and that $\langle\psi_{\boldsymbol{\theta}}^{\perp,j}|\psi_{\boldsymbol{\theta}}\rangle = 0$. This completes the proof. ■

It may be that $\mathcal{U}$ has a dimensionality less than d, the dimensionality of the unitary $U(\boldsymbol{\theta})$. For example, this will always happen if $M+1<d$, where M is the number of parameters. In general, $\mathcal{U}$ can have dimension u, with $u \leqslant M+1$ and, if $u<d$, the optimal filter is not fully determined. In an orthonormal basis, where $|\psi_{\boldsymbol{\theta}}\rangle$ is the first basis vector, and the first u basis vectors span $\mathcal{U}$, we have that a filter $\hat{F}$ is optimal if it takes the form

$$\hat{F} = \begin{pmatrix} P_{\boldsymbol{\theta}}^{\text{ps}} & 0 & \dots & 0 & \hat{c}^\dagger \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \hat{c} & 0 & \dots & 0 & \hat{D} \end{pmatrix}, \tag{A24}$$

where $\hat{c}$ is a (complex) vector of length $d-u$ and $\hat{D}$ is a $d-u$ by $d-u$ (complex) matrix such that $\hat{0} \leqslant \hat{F} \leqslant \hat{1}$.

The simplest example of an optimal postselection filter is for $\hat{c}=0$ and $\hat{D}=\hat{1}$, in which case, setting $P_{\boldsymbol{\theta}}^{\text{ps}} = t^2$, we recover the JAL filter:

$$\hat{F} = (t^2-1)\hat{\rho}_{\boldsymbol{\theta}} + \hat{1}. \tag{A25}$$

Note that the JAL filter does not require knowledge of $\mathcal{U}$.

So far, we have assumed that our filter is constructed based on perfect knowledge of the parameters, i.e., our initial estimate $\boldsymbol{\theta}_0$ is equal to the true value of the parameters $\boldsymbol{\theta}$. In practice, however, one does not know $\boldsymbol{\theta}$ exactly. Instead, one has an estimate $\boldsymbol{\theta}_0 = \boldsymbol{\theta} - \boldsymbol{\delta}$ and implements a filter using $\boldsymbol{\theta}_0$. In Ref. [26], it was shown that if one implements the JAL filter using $\boldsymbol{\theta}_0$, then $\mathcal{A}(\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}}, \hat{\rho}_{\boldsymbol{\theta}})$ and $P_{\boldsymbol{\theta}}^{\text{ps}}$ are only perturbed at order $\boldsymbol{\delta}^2$—there is no order $\boldsymbol{\delta}$ correction.

In Appendix B, we study the order $\boldsymbol{\delta}^2$ correction when using the JAL filter, deriving the upper bound

$$\mathcal{A}\big(\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}}, \hat{\rho}_{\boldsymbol{\theta}}\big) \leqslant \frac{1}{t^2}\left[1 - \frac{1-t^2}{t^2}\,\boldsymbol{\delta}^T \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})\,\boldsymbol{\delta}\right]. \tag{A26}$$

Equation (A26) provides a way to calculate the expected decrease in the information amplification to second order in $\boldsymbol{\delta}$. It allows one to understand which parameters in $\boldsymbol{\delta}$ are responsible for the largest loss of information. In the trivial case when $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$ is diagonal, the parameters θ_i with the largest QFI will also lead to the greatest loss.

APPENDIX B: INFORMATION AMPLIFICATION SCALING FOR THE JAL FILTER

In this Appendix, we study the second-order corrections to the information amplification when $\boldsymbol{\delta} \neq \mathbf{0}$. Consider the perturbed JAL filter

$$\hat{F} = (t^2-1)\hat{\rho}_{\boldsymbol{\theta}_0} + \hat{1}. \tag{B1}$$

We expand $\hat{U}(\boldsymbol{\theta}_0)$ to second order in $\boldsymbol{\delta}$:

$$\hat{U}(\boldsymbol{\theta}_0) = \hat{U}(\boldsymbol{\theta}) + [\partial_i \hat{U}(\boldsymbol{\theta})](\boldsymbol{\theta}_0 - \boldsymbol{\theta})_i + \tfrac{1}{2}\big[\partial_{i,j}^2 \hat{U}(\boldsymbol{\theta})\big](\boldsymbol{\theta}_0-\boldsymbol{\theta})_i(\boldsymbol{\theta}_0-\boldsymbol{\theta})_j. \tag{B2}$$

The above equation can be written as

$$\hat{U}(\boldsymbol{\theta}_0) = U(\boldsymbol{\theta}) + [\nabla_{\boldsymbol{\theta}}\hat{U}(\boldsymbol{\theta})]^T(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + \tfrac{1}{2}(\boldsymbol{\theta}_0-\boldsymbol{\theta})^T H(\boldsymbol{\theta})(\boldsymbol{\theta}_0-\boldsymbol{\theta}), \tag{B3}$$

where $\hat{H}(\theta) = \partial_{i,j}^2 U(\boldsymbol{\theta})$ is the Hessian matrix. Hence

$$\hat{U}(\boldsymbol{\theta}_0) = \hat{U}(\boldsymbol{\theta}) - i\hat{U}(\boldsymbol{\theta})\hat{d}_1 + \frac{i}{2}\hat{U}(\boldsymbol{\theta})\hat{d}_2 + O(\delta^3), \tag{B4}$$

where $\hat{d}_1$ and $\hat{d}_2$ are defined as

$$\hat{d}_1 = -i\,\hat{U}^\dagger(\boldsymbol{\theta})[\nabla_{\boldsymbol{\theta}}\hat{U}(\boldsymbol{\theta})]^T\boldsymbol{\delta}, \tag{B5}$$

$$\hat{d}_2 = -i\,\hat{U}^\dagger(\boldsymbol{\theta})\,\boldsymbol{\delta}^T\hat{H}(\boldsymbol{\theta})\,\boldsymbol{\delta}. \tag{B6}$$

We want to find $\hat{d}_1^\dagger$ and $\hat{d}_2^\dagger$. Consider

$$U(\boldsymbol{\theta}_0)U(\boldsymbol{\theta}_0)^\dagger = \hat{1}. \tag{B7}$$

Inserting the Taylor expansion, we find

$$\hat{1} = \left[\hat{U}(\boldsymbol{\theta}) - i\hat{U}(\boldsymbol{\theta})\hat{d}_1 + \frac{i}{2}\hat{U}(\boldsymbol{\theta})\hat{d}_2\right] \times \left[\hat{U}(\boldsymbol{\theta}) - i\hat{U}(\boldsymbol{\theta})\hat{d}_1 + \frac{i}{2}\hat{U}(\boldsymbol{\theta})\hat{d}_2\right]^\dagger \tag{B8}$$

$$= \hat{U}(\boldsymbol{\theta})\hat{U}^\dagger(\boldsymbol{\theta}) + i\hat{U}(\boldsymbol{\theta})[\hat{d}_1 - \hat{d}_1^\dagger]\hat{U}^\dagger(\boldsymbol{\theta}) + \frac{i}{2}\hat{U}(\boldsymbol{\theta})[\hat{d}_2 - \hat{d}_2^\dagger - 2\,i\,\hat{d}_1\hat{d}_1^\dagger]\hat{U}^\dagger(\boldsymbol{\theta}). \tag{B9}$$

Looking at the term of order $\boldsymbol{\delta}$, we immediately see that $\hat{d}_1 = \hat{d}_1^\dagger$. Instead, the second-order term gives

$$\hat{d}_2^\dagger = \hat{d}_2 - 2\,i\,\hat{d}_1\hat{d}_1^\dagger. \tag{B10}$$

Now, we can write $\hat{\rho}_{\boldsymbol{\theta}_0}$ as

$$\hat{\rho}_{\boldsymbol{\theta}_0} = \hat{U}(\boldsymbol{\theta}_0)\rho_0\hat{U}(\boldsymbol{\theta}_0)^\dagger \tag{B11}$$

$$= \left[\hat{U}(\boldsymbol{\theta}) - i\hat{U}(\boldsymbol{\theta})\hat{d}_1 + \frac{i}{2}\hat{U}(\boldsymbol{\theta})\hat{d}_2\right]\rho_0 \times \left[\hat{U}(\boldsymbol{\theta})^\dagger + i\hat{d}_1^\dagger\hat{U}(\boldsymbol{\theta})^\dagger - \frac{i}{2}\hat{d}_2^\dagger\hat{U}(\boldsymbol{\theta})^\dagger\right] \tag{B12}$$

$$= \hat{U}(\boldsymbol{\theta})\left[1 - i\hat{d}_1 + \frac{i}{2}\hat{d}_2\right]\rho_0\left[1 + i\hat{d}_1 - \frac{i}{2}\hat{d}_2 - \hat{d}_1^2\right]\hat{U}(\boldsymbol{\theta})^\dagger, \tag{B13}$$

$$= \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}_1] - \frac{i}{2}[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}_2] + \hat{D}_1\hat{\rho}_{\boldsymbol{\theta}}\hat{D}_1 - \hat{\rho}_{\boldsymbol{\theta}}\hat{D}_1^2 + O(\delta^3), \tag{B14}$$

where we defined $\hat{D}_1, \hat{D}_2$:

$$\hat{D}_1 = U(\boldsymbol{\theta})\hat{d}_1\hat{U}(\boldsymbol{\theta})^\dagger, \tag{B15}$$

$$\hat{D}_2 = U(\boldsymbol{\theta})\hat{d}_2\hat{U}(\boldsymbol{\theta})^\dagger. \tag{B16}$$

We can now calculate the postselection probability $P_{\boldsymbol{\theta}}^{\text{ps}}$, correct to order $\boldsymbol{\delta}^2$:

$$P_{\boldsymbol{\theta}}^{\text{ps}} = \text{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}], \tag{B17}$$

$$= \langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle, \tag{B18}$$

$$= \langle\psi_{\boldsymbol\theta}|\Big[(t^2-1)\Big(\hat\rho_{\boldsymbol\theta} + i[\hat\rho_{\boldsymbol\theta},\hat D_1] - \frac{i}{2}[\hat\rho_{\boldsymbol\theta},\hat D_2] + \hat D_1\hat\rho_{\boldsymbol\theta}\hat D_1 - \hat\rho_{\boldsymbol\theta}\hat D_1^2\Big) + \hat{1}\Big]|\psi_{\boldsymbol\theta}\rangle, \tag{B19}$$

$$= t^2 - (1-t^2)[\langle\psi_{\boldsymbol\theta}|(\hat D_1\hat\rho_{\boldsymbol\theta}\hat D_1)|\psi_{\boldsymbol\theta}\rangle - \langle\psi_{\boldsymbol\theta}|\hat\rho_{\boldsymbol\theta}\hat D_1^2|\psi_{\boldsymbol\theta}\rangle], \tag{B20}$$

$$= t^2 + (1-t^2)[\langle\psi_{\boldsymbol\theta}|\hat D_1^2|\psi_{\boldsymbol\theta}\rangle - |\langle\psi_{\boldsymbol\theta}|\hat D_1|\psi_{\boldsymbol\theta}\rangle|^2], \tag{B21}$$

$$= t^2 + (1-t^2)[\langle\psi_0|\hat d_1^2|\psi_0\rangle - |\langle\psi_0|\hat d_1|\psi_0\rangle|^2], \tag{B22}$$

$$= t^2 + (1-t^2)\sum_{i,j}\boldsymbol\delta_i\boldsymbol\delta_j[\langle\partial_i\psi_{\boldsymbol\theta}|\partial_j\psi_{\boldsymbol\theta}\rangle - \langle\partial_i\psi_{\boldsymbol\theta}|\psi_{\boldsymbol\theta}\rangle\langle\psi_{\boldsymbol\theta}|\partial_j\psi_{\boldsymbol\theta}\rangle], \tag{B23}$$

$$= t^2 + \frac{1}{4}(1-t^2)\sum_{i,j}\boldsymbol\delta_i\boldsymbol\delta_j\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})_{i,j}, \tag{B24}$$

$$= t^2 + \frac{1}{4}(1-t^2)\,\boldsymbol\delta^T\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta. \tag{B25}$$

Hence we find

$$P_{\boldsymbol\theta}^{\rm ps} = t^2 + \frac{1}{4}(1-t^2)\,\boldsymbol\delta^T\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta + O(\boldsymbol\delta^3). \tag{B26}$$

Since the QFIM is positive semidefinite, the term of order $O(|\boldsymbol\delta|^2)$ is always positive. Therefore, for $\boldsymbol\delta$ small, the postselection probability increases with increasing $|\boldsymbol\delta|$.

We can now find an upper bound on $\mathcal{A}(\hat\rho_{\boldsymbol\theta}^{\rm ps},\hat\rho_{\boldsymbol\theta})$. From the discussion in Ref. [26], we know that

$$\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm ps}) = \frac{1}{t^2}[\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}) + \Delta], \tag{B27}$$

where Δ is a matrix of order $O(|\boldsymbol\delta|^2)$ which we wish to bound. Because $\eta(P_{\boldsymbol\theta}^{\rm ps},\hat\rho_{\boldsymbol\theta}^{\rm ps},\hat\rho_{\boldsymbol\theta}) \leqslant 1$, we can write

$$P_{\boldsymbol\theta}^{\rm ps}\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm ps}) \leqslant \mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}). \tag{B28}$$

Substituting Eqs. (B26) and (B27) into Eq. (B28), we find

$$\Big[t^2 + \frac{1}{4}(1-t^2)\,\boldsymbol\delta^T\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta\Big]\frac{1}{t^2}[\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}) + \Delta] \leqslant \mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}), \tag{B29}$$

$$\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}) + \Delta + \frac{1-t^2}{4t^2}\boldsymbol\delta^T\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}) \leqslant \mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}), \tag{B30}$$

$$\Rightarrow \Delta \leqslant -\frac{1-t^2}{4t^2}[\boldsymbol\delta^T\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta]\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}). \tag{B31}$$

Therefore, we reach the important result

$$\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm ps}) \leqslant \frac{1}{t^2}\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\Big[1 - \frac{1-t^2}{4t^2}\,\boldsymbol\delta^T\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})\,\boldsymbol\delta\Big] + O(\boldsymbol\delta^3), \tag{B32}$$

which shows that any $\boldsymbol\delta \neq 0$ decreases the information amplification.

APPENDIX C: NOISY QFIM CALCULATIONS

In this Appendix, we calculate the QFIM when noise acts before or after filtering. To calculate the QFIM, we use the explicit expression from Ref. [35]:

$$\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})_{i,j} = 2\sum_{\substack{n,m=1\\ \lambda_n+\lambda_m>0}}^{d}\frac{\langle\lambda_n|\partial_j\hat\rho_{\boldsymbol\theta}|\lambda_m\rangle\langle\lambda_m|\partial_i\hat\rho_{\boldsymbol\theta}|\lambda_n\rangle}{\lambda_n+\lambda_m}. \tag{C1}$$

1. Noise after postselection

We start by calculating the QFIM of $\hat\rho_{\boldsymbol\theta}^{\rm n}$ using Eq. (C1), ignoring postselection for the moment. We calculate the eigenvalues $\{\lambda_i\}$ and eigenvectors $\{|\lambda_i\rangle\}$ of $\hat\rho_{\boldsymbol\theta}^{\rm n}$:

$$\lambda_1 = (1-\epsilon) + \frac{\epsilon}{d},\quad \lambda_i = \frac{\epsilon}{d}\quad \text{for}\quad i\neq 1, \tag{C2}$$

$$|\lambda_1\rangle = |\psi_{\boldsymbol\theta}\rangle,\quad |\lambda_i\rangle = \big|\psi_{\boldsymbol\theta}^{\perp,i}\big\rangle\quad \text{for}\quad i\neq 1, \tag{C3}$$

where we have defined an orthonormal eigenbasis $\{|\psi_{\boldsymbol\theta}\rangle, |\psi_{\boldsymbol\theta}{}^{\perp,2}\rangle, \dots, |\psi_{\boldsymbol\theta}{}^{\perp,d}\rangle\}$ and $1 \leqslant i \leqslant d$.

Substituting $\partial_i\hat\rho_{\boldsymbol\theta}^{\rm n}$ in Eq. (C1), where

$$\partial_i\hat\rho_{\boldsymbol\theta}^{\rm n} = (1-\epsilon)(|\partial_i\psi_{\boldsymbol\theta}\rangle\langle\psi_{\boldsymbol\theta}| + |\psi_{\boldsymbol\theta}\rangle\langle\partial_i\psi_{\boldsymbol\theta}|), \tag{C4}$$

we see that the term with $n=m=1$ is zero, since $\langle\partial_i\psi_{\boldsymbol\theta}|\psi_{\boldsymbol\theta}\rangle + \langle\psi_{\boldsymbol\theta}|\partial_i\psi_{\boldsymbol\theta}\rangle = \partial_i(\langle\psi_{\boldsymbol\theta}|\psi_{\boldsymbol\theta}\rangle) = \partial_i(1) = 0$. Similarly, the terms with both $n, m \neq 1$ are also zero, since $\langle\psi_{\boldsymbol\theta}|\psi_{\boldsymbol\theta}{}^{\perp,n}\rangle = 0$. The only nonzero terms are for $n\neq 1$ and $m = 1$ (or vice versa), giving

$$\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm n})_{i,j} \tag{C5}$$

$$= 4(1-\epsilon)^2\sum_{n\neq 1}\frac{\mathrm{Re}[\langle\partial_i\psi_{\boldsymbol\theta}|\lambda_n\rangle\langle\lambda_n|\partial_j\psi_{\boldsymbol\theta}\rangle]}{\lambda_n+\lambda_1}, \tag{C6}$$

$$= 4(1-\epsilon)^2\frac{\mathrm{Re}[\langle\partial_i\psi_{\boldsymbol\theta}|(\hat{1} - |\lambda_1\rangle\langle\lambda_1|)|\partial_j\psi_{\boldsymbol\theta}\rangle]}{\lambda_2+\lambda_1}, \tag{C7}$$

$$= \frac{4(1-\epsilon)^2}{(1-\epsilon)+\frac{2\epsilon}{d}}\mathrm{Re}[\langle\partial_i\psi_{\boldsymbol\theta}|\partial_j\psi_{\boldsymbol\theta}\rangle - \langle\partial_i\psi_{\boldsymbol\theta}|\psi_{\boldsymbol\theta}\rangle\langle\psi_{\boldsymbol\theta}|\partial_j\psi_{\boldsymbol\theta}\rangle], \tag{C8}$$

$$= \frac{(1-\epsilon)^2}{(1-\epsilon)+\frac{2\epsilon}{d}}\,\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta})_{i,j}. \tag{C9}$$

Since we have not assumed any particular form for the pure state $|\psi_{\boldsymbol\theta}\rangle$, we can postselect it with a general two-outcome POVM $\{\hat F_1 = \hat F, \hat F_2 = 1 - \hat F\}$. Blocking the states corresponding to outcome $\hat F_2$ and letting through those corresponding to outcome $\hat F_1 = \hat F$, we find that the postselected state is

$$|\psi_{\boldsymbol\theta}{}^{\rm ps}\rangle = \frac{\hat K|\psi_{\boldsymbol\theta}\rangle}{\sqrt{P_{\boldsymbol\theta}^{\rm ps}}},\quad \text{where } P_{\boldsymbol\theta}^{\rm ps} = \mathrm{Tr}\,[\hat F|\psi_{\boldsymbol\theta}\rangle\langle\psi_{\boldsymbol\theta}|], \tag{C10}$$

where $\hat F = \hat K^\dagger\hat K$. Hence

$$\mathcal{I}\big(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm ps,\,n}\big) = \frac{(1-\epsilon)^2}{(1-\epsilon)+\frac{2\epsilon}{d}}\mathcal{I}\big(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm ps}\big). \tag{C11}$$

At the same time, from the result in Eq. (C9),

$$\mathcal{I}\big(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}^{\rm n}\big) = \frac{(1-\epsilon)^2}{(1-\epsilon)+\frac{2\epsilon}{d}}\mathcal{I}(\boldsymbol\theta|\hat\rho_{\boldsymbol\theta}). \tag{C12}$$

Taking the ratio of Eqs. (C11) and (C12), and using $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}}) \approx \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})/t^2 + O(|\boldsymbol{\delta}|^2)$, we find that

$$\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps, n}}) = \frac{1}{t^2}\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\text{n}}) + O(|\boldsymbol{\delta}|^2). \tag{C13}$$

Hence all of the loss of information comes from noise and filtering is still lossless.

2. Noise before postselection

When noise acts before postselection, the resulting noisy postselected state is

$$\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}} = \frac{1}{P_{\boldsymbol{\theta}}^{\text{ps}}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + \frac{\epsilon}{d}\hat{K}\hat{K}^\dagger\right], \tag{C14}$$

where the normalization constant $P_{\boldsymbol{\theta}}^{\text{ps}}$ is given by

$$P_{\boldsymbol{\theta}}^{\text{ps}} = (1-\epsilon)\,\text{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}] + \frac{\epsilon}{d}\,\text{Tr}[\hat{F}]. \tag{C15}$$

We consider filters of the form

$$\hat{F} = (p_{\boldsymbol{\theta}} - B)|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}| + B\,\hat{\Pi}_u + D\,\hat{\Pi}_n, \tag{C16}$$

where $p_{\boldsymbol{\theta}}, B, D \in [0,1]$, $\hat{\Pi}_u$ is the orthogonal projection onto $\mathcal{U}$, and $\hat{\Pi}_n$ is the orthogonal projection onto $\mathcal{U}^\perp$. For suitable bases, this filter is represented by the matrix

$$\hat{F} = \begin{pmatrix} \begin{pmatrix} p_{\boldsymbol{\theta}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & B \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} D & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D \end{pmatrix} \end{pmatrix}. \tag{C17}$$

We start by making the assumption that $\boldsymbol{\delta} = 0$, i.e., $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$. In Appendix F, we then show that the result is unchanged for small $\boldsymbol{\delta}$, up to a term of order $O(|\boldsymbol{\delta}|^2)$.

Before calculating the QFIM using Eq. (C1), we need to find the eigenvalues and eigenvectors of $\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}$. We first look at the eigenvalues and eigenvectors of $\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}$. Letting $u = \dim \mathcal{U}$, we can pick orthonormal bases $\{|\psi_{\boldsymbol{\theta}}\rangle, |\psi_{\boldsymbol{\theta}}{}^{\perp,2}\rangle, \dots, |\psi_{\boldsymbol{\theta}}{}^{\perp,u}\rangle\}$ of $\mathcal{U}$ and $\{|\psi_{\boldsymbol{\theta}}{}^{\perp,u+1}\rangle, \dots, |\psi_{\boldsymbol{\theta}}{}^{\perp,d}\rangle\}$ of $\mathcal{U}^\perp$ that satisfy

$$\hat{F}|\psi_{\boldsymbol{\theta}}\rangle = p_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle, \tag{C18}$$

$$\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle = B|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle \quad \text{for} \quad 2 \leqslant i \leqslant u, \tag{C19}$$

$$\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle = D|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle \quad \text{for} \quad u+1 \leqslant i \leqslant d. \tag{C20}$$

Now, we calculate the eigenvalue λ_1 of the eigenstate $|\lambda_1\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}\rangle/\sqrt{p_{\boldsymbol{\theta}}}$:

$$\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}|\lambda_1\rangle \tag{C21}$$

$$= \frac{1}{P_{\boldsymbol{\theta}}^{\text{ps}}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + \frac{\epsilon}{d}\hat{K}\hat{K}^\dagger\right]\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p_{\boldsymbol{\theta}}}}, \tag{C22}$$

$$= \frac{1}{P_{\boldsymbol{\theta}}^{\text{ps}}}\left[(1-\epsilon)\hat{K}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\hat{K}^\dagger\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p_{\boldsymbol{\theta}}}} + \frac{\epsilon}{d}\hat{K}\frac{\hat{F}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p_{\boldsymbol{\theta}}}}\right], \tag{C23}$$

$$= \frac{\left(1-\epsilon+\frac{\epsilon}{d}\right)p_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p_{\boldsymbol{\theta}}}} = \lambda_1|\lambda_1\rangle. \tag{C24}$$

Similarly, the eigenvalues $\lambda_2, \dots, \lambda_u$ of the eigenstates $|\lambda_i\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle/\sqrt{B}$, for $2 \leqslant i \leqslant u$, are

$$\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}|\lambda_i\rangle \tag{C25}$$

$$= \frac{1}{P_{\boldsymbol{\theta}}^{\text{ps}}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + \frac{\epsilon}{d}\hat{K}\hat{K}^\dagger\right]\frac{\hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle}{\sqrt{B}}, \tag{C26}$$

$$= \frac{\epsilon}{d\,P_{\boldsymbol{\theta}}^{\text{ps}}}\hat{K}\frac{\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle}{\sqrt{B}}, \tag{C27}$$

$$= \frac{\epsilon B}{d\,P_{\boldsymbol{\theta}}^{\text{ps}}}\frac{\hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle}{\sqrt{B}} = \lambda_i|\lambda_i\rangle. \tag{C28}$$

Doing the same calculations for the eigenvalues $\lambda_{u+1}, \dots, \lambda_d$ of the eigenstates $|\lambda_i\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,i}\rangle/\sqrt{D}$, we see that B in Eq. (C28) is replaced by D. Therefore,

$$\lambda_1 = \frac{(1-\epsilon+\epsilon/d)\,p_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}, \tag{C29}$$

$$\lambda_i = \frac{\epsilon\,B}{d\,P_{\boldsymbol{\theta}}^{\text{ps}}} \quad \text{for} \quad 2 \leqslant i \leqslant u, \tag{C30}$$

$$\lambda_i = \frac{\epsilon\,D}{d\,P_{\boldsymbol{\theta}}^{\text{ps}}} \quad \text{for} \quad u+1 \leqslant i \leqslant d. \tag{C31}$$

Now that we have found the eigenvalues and eigenvectors of $\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}$, we can evaluate the QFIM using Eq. (C1). For simplicity, in the following calculation set $a = (1-\epsilon)$ and $b = \epsilon/d$. We need to calculate $\partial_i\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}$:

$$\partial_i\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}} = \frac{a}{P_{\boldsymbol{\theta}}^{\text{ps}}}\hat{K}\partial_i\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger - a\frac{\partial_i p_{\boldsymbol{\theta}}}{(P_{\boldsymbol{\theta}}^{\text{ps}})^2}[a\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + b\hat{K}\hat{K}^\dagger], \tag{C32}$$

$$= \frac{a\,p_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}(|\partial_i\lambda_1\rangle\langle\lambda_1| + |\lambda_1\rangle\langle\partial_i\lambda_1|) + \frac{a\partial_i p_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}|\lambda_1\rangle\langle\lambda_1| - a^2 p_{\boldsymbol{\theta}}\frac{\partial_i p_{\boldsymbol{\theta}}}{(P_{\boldsymbol{\theta}}^{\text{ps}})^2}|\lambda_1\rangle\langle\lambda_1|, \tag{C33}$$

$$= \frac{a\,p_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}\partial_i\hat{\rho}_1 + ab\frac{(\partial_i p_{\boldsymbol{\theta}})}{(P_{\boldsymbol{\theta}}^{\text{ps}})^2}[p_{\boldsymbol{\theta}} + B(u-1) + D\,u_\perp]\hat{\rho}_1, \tag{C34}$$

where we set $u_\perp = d - u$ and introduced

$$\hat{\rho}_1 = |\lambda_1\rangle\langle\lambda_1|. \tag{C35}$$

In Eq. (C34) we also made use of the fact that

$$|\partial_i\lambda_1\rangle = \left(\frac{\hat{K}|\partial_i\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p_{\boldsymbol{\theta}}}} - \frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{2\sqrt{p_{\boldsymbol{\theta}}^3}}\partial_i p_{\boldsymbol{\theta}}\right), \tag{C36}$$

so that

$$|\partial_i\lambda_1\rangle\langle\partial_i\lambda_1|\lambda_1 + |\lambda_1\rangle\langle\lambda_1|\partial_i\lambda_1 = \frac{1}{p_{\boldsymbol{\theta}}}\hat{K}\partial_i\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger - \frac{\partial_i p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}}|\lambda_1\rangle\langle\lambda_1|. \tag{C37}$$

Consider a term of the form $\langle\lambda_n|\partial_i\hat{\rho}_{\boldsymbol{\theta}}|\lambda_m\rangle$, as appearing in Eq. (C1). If $m = n$ or both $m, n > 1$, then $\langle\lambda_n|\partial_i\hat{\rho}_1|\lambda_m\rangle = 0$ (taking the derivative of $\langle\lambda_1| \equiv\rangle 1$ implies that $\langle\lambda_1|\partial_i\hat{\rho}_1|\lambda_1\rangle = 0$). On the other hand, if $m \neq n$, then $\langle\lambda_1|\hat{\rho}_1|\lambda_1\rangle = 0$. Further, note that, because the terms of the form $|\partial_i\lambda_1\rangle$ appearing in the QFIM are contained in the subspace $\mathcal{U}$, the only nonzero contribution comes from $m, n \leqslant u$. Thus Eq. (C1) breaks up into two sums: one with $n = 1$ and $m \neq 1$ and one with $m = n = 1$. The first of these sums is given by

$$4\frac{a^2 p_{\theta}^2}{\left(P_{\theta}^{\mathrm{ps}}\right)^2}\sum_{n=2}^{u}\mathrm{Re}\left(\frac{\langle\partial_j\lambda_1|\lambda_n\rangle\langle\lambda_n|\partial_i\lambda_1\rangle}{\lambda_n+\lambda_1}\right), \tag{C38}$$

$$= 4\frac{a^2 p_{\theta}^2}{P_{\theta}^{\mathrm{ps}}}\frac{\mathrm{Re}(\langle\partial_j\lambda_1|(\hat{1}-|\lambda_1\rangle\langle\lambda_1|)|\partial_i\lambda_1\rangle)}{(a+b)\,p_{\theta}+bB}, \tag{C39}$$

$$= 4\frac{a^2 p_{\theta}^2}{P_{\theta}^{\mathrm{ps}}}\frac{\mathrm{Re}(\langle\partial_j\lambda_1|\partial_i\lambda_1\rangle-\langle\partial_j\lambda_1|\lambda_1\rangle\langle\lambda_1|\partial_i\lambda_1\rangle)}{(a+b)\,p_{\theta}+bB}, \tag{C40}$$

$$= \frac{a^2 p_{\theta}^2}{P_{\theta}^{\mathrm{ps}}[(a+b)\,p_{\theta}+bB]}\mathcal{I}\left(\theta|\hat{\rho}_{\theta}^{\mathrm{ps}}\right)_{i,j}. \tag{C41}$$

The second term is $m = n = 1$:

$$2a^2b^2\frac{(\partial_i p_{\theta})(\partial_j p_{\theta})}{\left(P_{\theta}^{\mathrm{ps}}\right)^4}[p_{\theta}+B(u-1)+D\,u_{\perp}]^2 \times\frac{|\langle\lambda_1|\hat{\rho}_1|\lambda_1\rangle|^2}{2\lambda_1}, \tag{C42}$$

$$= \frac{a^2b^2\,(\partial_i p_{\theta})(\partial_j p_{\theta})}{\left(P_{\theta}^{\mathrm{ps}}\right)^4}\frac{[p_{\theta}+B(u-1)+D\,u_{\perp}]^2}{\lambda_1}, \tag{C43}$$

$$= \frac{a^2b^2\,(\partial_i p_{\theta})(\partial_j p_{\theta})}{(a+b)\left(P_{\theta}^{\mathrm{ps}}\right)^3\,p_{\theta}}[p_{\theta}+B(u-1)]^2. \tag{C44}$$

Adding the two terms together, we find

$$\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\mathrm{n,\,ps}})_{i,j} = \frac{1}{P_{\theta}^{\mathrm{ps}}}\frac{(1-\epsilon)^2 p_{\theta}^2}{\left(1-\epsilon+\frac{\epsilon}{d}\right)p_{\theta}+\frac{\epsilon}{d}B}\mathcal{I}\left(\boldsymbol{\theta}|\hat{\rho}_{\theta}^{\mathrm{ps}}\right)_{i,j} + \frac{\epsilon}{d}\frac{(1-\epsilon)^2\,(\partial_i p_{\theta})(\partial_j p_{\theta})}{\left(1-\epsilon+\frac{\epsilon}{d}\right)\left(P_{\theta}^{\mathrm{ps}}\right)^3\,p_{\theta}}[p_{\theta}+B(u-1)+D\,u_{\perp}]^2. \tag{C45}$$

Finally, note that

$$\partial_i p_{\theta} = (1-\epsilon)\mathrm{Tr}[\left(|\partial_i\psi_{\theta}\rangle\langle\partial_i\psi_{\theta}|\psi_{\theta}+|\psi_{\theta}\rangle\langle\psi_{\theta}|\partial_i\psi_{\theta}\right)\hat{K}^{\dagger}\hat{K}], \tag{C46}$$

$$= (1-\epsilon)t^2(\langle\psi_{\theta}|\partial_i\psi_{\theta}\rangle+\langle\partial_i\psi_{\theta}|\psi_{\theta}\rangle), \tag{C47}$$

$$= (1-\epsilon)t^2\partial_i(\langle\psi_{\theta}|\psi_{\theta}\rangle) = 0, \tag{C48}$$

and thus the second term in Eq. (C45) is zero. In general, for $\boldsymbol{\delta}\neq 0$, $p_{\theta} = t^2 + O(\boldsymbol{\delta}^2)$; hence $(\partial_i p_{\theta}) = O(\boldsymbol{\delta})$ and the lowest-order correction is of order $\boldsymbol{\delta}^2$. Therefore, we can write

$$\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\theta}^{\mathrm{n,\,ps}})_{i,j} = \frac{1}{P_{\theta}^{\mathrm{ps}}}\frac{(1-\epsilon)^2 p_{\theta}^2}{\left(1-\epsilon+\frac{\epsilon}{d}\right)p_{\theta}+\frac{\epsilon}{d}B}\mathcal{I}\left(\boldsymbol{\theta}|\hat{\rho}_{\theta}^{\mathrm{ps}}\right)_{i,j}. \tag{C49}$$

As in the noiseless case, the family of filters in Eq. (26) scale the entries of the QFIM by the same factor.

APPENDIX D: FILTER OPTIMIZATION

In this Appendix, we treat noisy postselection explicitly, for some small values of u and d. We fully optimize the filter for $u = d = 2$ and then show that nondiagonal filters are advantageous for $u = 2, d = 3$.

Since most of our states are non-normalized, it is first useful to explicitly deal with normalization. Consider the case of a single parameter:

$$\hat{\rho}_{\theta} = \frac{\hat{\rho}_{\theta}'}{P_{\theta}^{\mathrm{ps}}}, \tag{D1}$$

where $P_{\theta}^{\mathrm{ps}} = \mathrm{Tr}(\hat{\rho}_{\theta}')$ normalizes the state (and in our case will be the probability of passing the filter). We can then calculate

$$\partial\hat{\rho}_{\theta} = \frac{\partial\hat{\rho}_{\theta}'}{P_{\theta}^{\mathrm{ps}}} - \frac{\partial P_{\theta}^{\mathrm{ps}}}{\left(P_{\theta}^{\mathrm{ps}}\right)^2}\hat{\rho}_{\theta}. \tag{D2}$$

Decomposing the symmetric logarithmic derivative $\hat{\Lambda}$ as

$$\hat{\Lambda} = \hat{\Lambda}' - \frac{\partial P_{\theta}^{\mathrm{ps}}}{P_{\theta}^{\mathrm{ps}}}\hat{1}, \tag{D3}$$

it is then easy to see that $\hat{\Lambda}'$ satisfies the "reduced" equation

$$\frac{1}{2}\{\hat{\Lambda}', \hat{\rho}_{\theta}'\} = \partial\hat{\rho}_{\theta}'. \tag{D4}$$

We can calculate the quantum Fisher information:

$$\mathcal{I}(\theta|\hat{\rho}_{\theta}) = \mathrm{Tr}(\partial\hat{\rho}_{\theta}\hat{\Lambda}), \tag{D5}$$

$$= \mathrm{Tr}\left[\left(\frac{\partial\hat{\rho}_{\theta}'}{P_{\theta}^{\mathrm{ps}}} - \frac{\partial P_{\theta}^{\mathrm{ps}}}{P_{\theta}^{\mathrm{ps}}}\hat{\rho}_{\theta}\right)\left(\hat{\Lambda}' - \frac{\partial P_{\theta}^{\mathrm{ps}}}{P_{\theta}^{\mathrm{ps}}}\hat{1}\right)\right], \tag{D6}$$

$$= \frac{1}{P_{\theta}^{\mathrm{ps}}}\left[\mathcal{I}(\theta|\hat{\rho}_{\theta}') - \frac{\partial P_{\theta}^{\mathrm{ps}}}{P_{\theta}^{\mathrm{ps}}}\mathrm{Tr}(\hat{\rho}_{\theta}'\hat{\Lambda}') - \frac{\partial P_{\theta}^{\mathrm{ps}}}{P_{\theta}^{\mathrm{ps}}}\mathrm{Tr}(\partial\hat{\rho}_{\theta}') + \frac{(\partial P_{\theta}^{\mathrm{ps}})^2}{P_{\theta}^{\mathrm{ps}}}\right], \tag{D7}$$

$$= \frac{1}{P_{\theta}^{\mathrm{ps}}}\left[\mathcal{I}(\theta|\hat{\rho}_{\theta}') - \frac{(\mathrm{Tr}[\partial\hat{\rho}_{\theta}'])^2}{P_{\theta}^{\mathrm{ps}}}\right]. \tag{D8}$$

1. Full filter optimization in $d = u = 2$ (qubit sensor)

We fix some point $\boldsymbol{\theta}$ in the parameter space. For the moment, we consider variations of a single parameter and take $\partial = \partial_1$. We expand

$$|\partial\psi_{\theta}\rangle = ix|\psi_{\theta}\rangle + \alpha|\psi_{\perp}\rangle, \tag{D9}$$

where $|\psi_{\theta}\rangle, |\psi_{\perp}\rangle$ are an orthonormal basis of our Hilbert space [45]. Choosing a filter corresponds to picking a 2×2 matrix (with respect to the aforementioned basis) $\hat{K}$ such that

$$0 \leqslant \hat{K}^{\dagger}\hat{K} \leqslant \hat{1}. \tag{D10}$$

If one attempts to optimize directly over 2×2 matrices, constraint (D10) is highly nontrivial, so we first parametrize $\hat{K}$ to simplify (D10).

Suppose that we have some choice of filter $\hat{K}$; we can write it in its singular value decomposition $\hat{K} = \hat{V}\hat{D}\hat{W}$, where $\hat{D}$ is diagonal with non-negative real eigenvalues, $\hat{W}$ is an SU(2) matrix, and $\hat{V}$ is a U(2) matrix. Let $\hat{K}' = \hat{V}^{\dagger}\hat{K}$; then the postselected state is changed by a unitary and thus the QFI and postselection probabilities do not change. Therefore, $\hat{K}'$ has exactly the same performance as $\hat{K}$ and it is sufficient to consider filters of the form $\hat{D}\hat{W}$. Note that $\hat{K}^\dagger\hat{K} = \hat{W}^\dagger\hat{D}^2\hat{W}$ and constraint (D10) is equivalent to $0 \leqslant \hat{D} \leqslant \hat{1}$. Using the general form of an SU(2) unitary, our filter is thus parametrized by

$$\hat{D} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad \text{where} \quad 0 \leqslant a, b \leqslant 1, \tag{D11}$$

$$\hat{W} = \begin{pmatrix} \gamma & -\beta^* \\ \beta & \gamma^* \end{pmatrix}, \quad \text{where} \quad \gamma, \beta \in \mathbb{C},\ |\gamma|^2 + |\beta|^2 = 1, \tag{D12}$$

and $\hat{K} = \hat{D}\hat{W}$. The postselected state is then given by

$$\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}} = \frac{(1-\epsilon)\hat{D}\hat{W}|\psi_\theta\rangle\langle\psi_\theta|\hat{W}^\dagger\hat{D} + \frac{\epsilon}{2}\hat{D}^2}{P_{\boldsymbol{\theta}}^{\text{ps}}} := \frac{\hat{\rho}'_{\boldsymbol{\theta}}}{P_{\boldsymbol{\theta}}^{\text{ps}}}, \tag{D13}$$

where $P_{\boldsymbol{\theta}}^{\text{ps}} = \text{Tr}(\hat{\rho}'_{\boldsymbol{\theta}})$ is the probability of postselection. One can calculate

$$\hat{\rho}'_{\boldsymbol{\theta}} = \frac{1}{2}\begin{pmatrix} a^2\epsilon + 2(1-\epsilon)|\gamma|^2 & 2ab(1-\epsilon)\gamma\beta^* \\ 2ab(1-\epsilon)\gamma^*\beta & b^2\epsilon + 2(1-\epsilon)|\beta|^2 \end{pmatrix}, \tag{D14}$$

$$\partial\hat{\rho}'_{\boldsymbol{\theta}} = (1-\epsilon) \times \begin{pmatrix} -a^2(\alpha\gamma^*\beta^* + \alpha^*\gamma\beta) & ab[\alpha^*\gamma^2 - \alpha(\beta^*)^2] \\ ab[\alpha(\gamma^*)^2 - \alpha^*\beta^2] & b^2(\alpha\gamma^*\beta^* + \alpha^*\gamma\beta) \end{pmatrix}. \tag{D15}$$

We thus find

$$P_{\boldsymbol{\theta}}^{\text{ps}} = \text{Tr}(\hat{\rho}'_{\boldsymbol{\theta}}) = \frac{1}{2}[(a^2+b^2)\epsilon + 2(1-\epsilon)(a^2|\gamma|^2 + b^2|\beta|^2)]. \tag{D16}$$

By finding the symmetric logarithmic derivative, we use Eq. (D8) to calculate

$$\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}}) = \frac{4a^2b^2(1-\epsilon)^2|\alpha|^2}{\left(P_{\boldsymbol{\theta}}^{\text{ps}}\right)^2}. \tag{D17}$$

Recall that we wish to maximize $\mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}^{\text{n, ps}})$, while minimizing $P_{\boldsymbol{\theta}}^{\text{ps}}$. Consider the optimal values of γ, β for fixed values of a, b. Since γ, β only appear in $P_{\boldsymbol{\theta}}^{\text{ps}}$ both of our goals are achieved by minimizing $P_{\boldsymbol{\theta}}^{\text{ps}}$ subject to $|\gamma|^2 + |\beta|^2 = 1$. This is equivalent to minimizing

$$f_{a,b}(\gamma) = a^2|\gamma|^2 + b^2(1-|\gamma|^2). \tag{D18}$$

If $a \geqslant b$ this is clearly minimized for $|\gamma| = 0$; otherwise, if $a < b$ it is clearly minimized for $|\gamma| = 1$. Let

$$\hat{\Omega} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{D19}$$

Then we can see that the best possible filter performance is attained by a filter with

$$\hat{W} = \begin{cases} \hat{1}_2, & \text{for} \quad a < b, \\ \hat{\Omega}, & \text{for} \quad a \geqslant b. \end{cases} \tag{D20}$$

If $W = \hat{1}_2$, then $\hat{K}$ (and hence $\hat{F}$) are both diagonal. Otherwise, note that $\hat{\Omega}^T\hat{K}$ is a diagonal filter with the same filter performance. Hence in either case the best filter performance is attained by a diagonal filter.

Now consider the full multiparameter QFIM. Since $d = 2$, we can express any state as a linear combination of $\{|\psi_{\boldsymbol{\theta}}\rangle, |\psi_\perp\rangle\}$ and thus the expansion in Eq. (D9) will hold for any parameter derivative (for different values of x and α). The optimum scaling for the diagonal terms of the QFIM is then achieved by a diagonal filter (as seen above) and thus if this filter also scales the off diagonal terms then it will be optimal. Indeed in Appendix C 2 we show such a diagonal filter does scale the off diagonal terms and thus the optimal filter in the multiparameter case is also diagonal.

2. Nondiagonal filters are better in $u = 2, d = 3$

For notational brevity, we consider the case of a single parameter θ. We have some filter (with Kraus operator) $\hat{K}$, a pure state $|\psi_\theta\rangle$, and a depolarizing noise rate ϵ. For brevity, we will set $\epsilon = 1/2$ so that it can be absorbed into the normalization constant. We then find

$$\hat{\rho}'_\theta = \hat{K}|\psi_\theta\rangle\langle\psi_\theta|\hat{K}^\dagger + \hat{K}\hat{K}^\dagger/d. \tag{D21}$$

Fixing some θ, then we can decompose

$$|\partial\psi_\theta\rangle = ix|\psi_\theta\rangle + \alpha|\psi_\perp\rangle, \tag{D22}$$

where $|\psi_\theta\rangle, |\psi_\perp\rangle$ are orthonormal. Assume our Hilbert space $\mathcal{H}$ is three dimensional, which is the minimal dimension such that $\mathcal{H}$ is not spanned by $|\psi_\theta\rangle, |\psi_\perp\rangle$. Thus we can find a third normalized state $|n\rangle$ orthogonal to $|\psi_\theta\rangle$ and $|\psi_\perp\rangle$. Consider the representation of $\hat{K}$ in this basis. Take the canonical choice $(t-1)|\psi_\theta\rangle\langle\psi_\theta| + \hat{\Pi}_u$ from the noiseless case, with the addition of a single off-diagonal term $b \in \mathbb{R}$

$$\hat{K} = \begin{pmatrix} t & 0 & b \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{D23}$$

We can calculate

$$\hat{F} = \hat{K}^\dagger\hat{K} = \begin{pmatrix} t^2 & 0 & tb \\ 0 & 1 & 0 \\ tb & 0 & b^2 \end{pmatrix}. \tag{D24}$$

Note $\hat{F}$ has eigenvalues $0, 1, t^2 + b^2$. Thus as long as $0 \leqslant b \leqslant \sqrt{1-t^2}$, we have $0 \leqslant \hat{F} \leqslant \hat{1}$ as required. We then find

$$\hat{\rho}'_\theta = \begin{pmatrix} (4/3)t^2 & 0 & tb/3 \\ 0 & 1/3 & 0 \\ tb/3 & 0 & b^2/3 \end{pmatrix}, \quad \partial\hat{\rho}'_\theta = \begin{pmatrix} 0 & \alpha^* t & 0 \\ \alpha t & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{D25}$$

One can explicitly solve for $\hat{\Lambda}'$:

$$\hat{\Lambda}' = \begin{pmatrix} 0 & \hat{\Lambda}'^*_{12} & 0 \\ \hat{\Lambda}'_{21} & 0 & \hat{\Lambda}'^*_{23} \\ 0 & \hat{\Lambda}'_{32} & 0 \end{pmatrix}, \tag{D26}$$

where

$$\hat{\Lambda}'^*_{12} = \hat{\Lambda}'_{21} = \frac{6(1+b^2)t\alpha}{1+b^2+(4+3b^2)t^2}, \tag{D27}$$

$$\hat{\Lambda}'^*_{23} = \hat{\Lambda}'_{32} = -\frac{6bt^2\alpha}{1+b^2+(4+3b^2)t^2}. \tag{D28}$$

We give quantum Fisher information of

$$\mathcal{I}(\theta|\hat{\rho}'_\theta) = \frac{12|\alpha|^2(1+b^2)t^2}{1+b^2+(4+3b^2)t^2}, \tag{D29}$$

$$= \frac{12|\alpha|^2 t^2}{1+\left(3+\frac{1}{1+b^2}\right)t^2}, \tag{D30}$$

$$= P^{\text{ps}}_{\boldsymbol{\theta}}\,\mathcal{I}(\theta|\hat{\rho}_\theta), \tag{D31}$$

using Eq. (D8) and noting $\partial\hat{\rho}'_\theta$ is traceless. The information rate is manifestly increasing in $|b|$ and thus it is maximized by setting $b=\sqrt{1-t^2}$, i.e., maximal mixing.

We can compare this with the canonical choice in the noiseless case: $(t-1)|\psi_\theta\rangle\langle\psi_\theta| + \hat{1}$, which is diagonal in our basis. Consider a slightly more general filter $\hat{K}'$, which we take to be diagonal:

$$\hat{K}' = \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & r \end{pmatrix}, \tag{D32}$$

where $r \in [0, 1]$ controls the probability of transmitting an $|n\rangle$ state (which can only come from noise).

In this case one can similarly calculate the information rate

$$P^{\text{ps}}_{\boldsymbol{\theta}}\,\mathcal{I}(\theta|\hat{\rho}_\theta) = \frac{12|\alpha|^2t^2}{1+4t^2}, \tag{D33}$$

which is found to be independent of r. Comparing Eqs. (D29) and (D33), we see that any nonzero value of b increases the information rate. Thus mixing is beneficial and the optimal filter in the noiseless case is not optimal in the noisy case.

APPENDIX E: OPTIMAL FILTER FOR DETECTOR SATURATION

In this Appendix we find the optimal values of B and $p_{\boldsymbol{\theta}}$ for when detector saturation is the limiting factor. We recall from the main text [Eq. (28)] that we are attempting to solve the following optimization problem:

$$\begin{aligned} &\text{maximize} \quad \eta(p_{\boldsymbol{\theta}}, B) = \frac{\left(1-\epsilon+2\frac{\epsilon}{d}\right)p_{\boldsymbol{\theta}}}{\left(1-\epsilon+\frac{\epsilon}{d}\right)\frac{p_{\boldsymbol{\theta}}}{B}+\frac{\epsilon}{d}}, \\ &\text{subject to} \quad 0 \leqslant p_{\boldsymbol{\theta}}, B \leqslant 1, \\ &\qquad \left(1-\epsilon+\frac{\epsilon}{d}\right)p_{\boldsymbol{\theta}} + \frac{\epsilon}{d}(u-1)B \leqslant P_{\max}. \end{aligned} \tag{E1}$$

For notational brevity, we define $x = p_{\boldsymbol{\theta}},\ y = B,\ b = (1-\epsilon+\frac{\epsilon}{d}),\ c = \frac{\epsilon}{d}(u-1)$, and $g = \frac{\epsilon}{d}$. In this notation, we see that the optimization problem (E1) is equivalent to

$$\begin{aligned} &\text{maximize} \qquad f(x,y) = \frac{(b+g)x}{b\frac{x}{y}+g}, \\ &\text{subject to} \qquad 0 \leqslant x, y \leqslant 1, \\ &\qquad bx + cy \leqslant P_{\max}. \end{aligned} \tag{E2}$$

We note that $f(x, y)$ is a strictly increasing function of x and y (for $x, y \geqslant 0$) and thus the optimum point will be on the boundary of the feasible set. We deduce that if (1,1) is in the feasible set it will be the optimum point. This occurs iff $b + c \leqslant P_{\max}$.

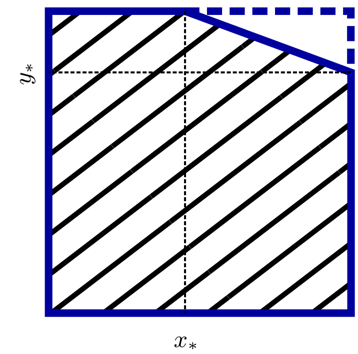

FIG. 5. Feasible region (shaded) of the optimization problem (E2) as a subset of the unit square. The linear constraint intersects the unit square at $(1, y_*)$ and $(x_*, 1)$.

Hence we consider $b + c > P_{\max}$, in which case the feasible set is the intersection of the unit square with a line of negative gradient, as depicted in Fig. 5.

Since f is strictly increasing in the x and y directions, its maximum will be obtained on the linear constraint, i.e., it will satisfy $bx + cy = P_{\max}$. As before, we introduce $t^2 = x/y$, so that

$$x(t) = \frac{P_{\max}}{b + c/t^2}, \quad y(t) = \frac{P_{\max}}{bt^2 + c}. \tag{E3}$$

We note that $f(x(t), y(t)) = P_{\max}\,\mathcal{A}(t)$, where $\mathcal{A}(t)$ is the information amplification given by Eq. (29):

$$\mathcal{A}(t) = \frac{(b+g)t^2}{(bt^2+c)(bt^2+g)}. \tag{E4}$$

We thus wish to maximize $\mathcal{A}(t)$ over the possible values of t. Let x_* be the smallest value of x attained on the line $bx + cy = P_{\max}$, while $0 \leqslant x, y \leqslant 1$. If the line intersects the unit square on the boundary $y = 1$, then $x_* > 0$. Otherwise, the line intersects the square on the boundary $x = 0$ and $x_* = 0$. We define y_* similarly. A graphical example is given in Fig. 5. We thus see that the allowed range of t is given by

$$x_* \leqslant t^2 \leqslant 1/y_*. \tag{E5}$$

Since $x_*, y_* \leqslant 1$, by definition $t = 1$ is always allowed.

We find that $\mathcal{A}(t) \to 0$ as $t \to 0, \infty$ and $\mathcal{A}(t)$ has a single stationary point, which therefore must be a maximum. As before, the maximum occurs at $t = t_{\text{pp}}$, which can be bigger or smaller than 1, depending on u, d, and ϵ. As $\mathcal{A}(t)$ is monotonic on the ranges $[0, t_{\text{pp}}]$, $[t_{\text{pp}}, \infty)$ we see that $\mathcal{A}(t)$ will be maximized by the point in the interval $[\sqrt{x_*}, \sqrt{1/y_*}]$ that is closest to t_{pp}. Explicitly, there are two cases, depending on whether t_{pp} is bigger or smaller than 1.

(i) $t_{\text{pp}} \leqslant 1$. We have

$$t^2 = \max\left[t_{\text{pp}}^2, x_*\right] = \max\left[t_{\text{pp}}^2, \frac{P_{\max}-c}{b}\right]. \tag{E6}$$

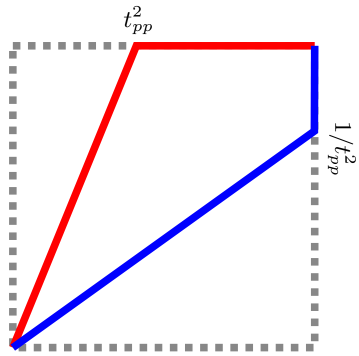

FIG. 6. Path traced out by the optimum point as $P_{\max}$ varies in the cases $t_{\rm pp} \leqslant 1$ (red, upper line) and $t_{\rm pp} \geqslant 1$ (blue, lower line). For a fixed value of $P_{\max}$ the optimum value is given by the intersection of the line $bx + cy = P_{\max}$ and the path of the optimum point if such an intersection exists; otherwise, it is given by (1,1).

(ii) $t_{\rm pp} \geqslant 1$. We have

$$t^2 = \min\left[t_{\rm pp}^2, 1/y_*\right] = \min\left[t_{\rm pp}^2, \left(\frac{c}{P_{\max} - b}\right)^+\right], \tag{E7}$$

where $a^+ = a$ if $a \geqslant 0$ and $a^+ = \infty$ if $a < 0$.

This solution is best described geometrically. Consider starting $P_{\max}$ at 1 and then decreasing it to zero. The maximum point of f then traces out a path $[x(P_{\max}), y(P_{\max})]$. If $t_{\rm pp} \leqslant 1$ this path moves along the boundary $y = 1$ until it reaches the point $(t_{\rm pp}^2, 1)$. At this point, it moves towards the origin, along the line $x/y = t_{\rm pp}^2$. If $t_{\rm pp} \geqslant 1$ the behavior is mirrored, with the optimum point now moving along the y axis to $(1, 1/t_{\rm pp}^2)$ before moving along the line $x/y = t_{\rm pp}^2$ to the origin. The two types of path are sketched in Fig. 6.

APPENDIX F: NOISY INFORMATION AMPLIFICATION SCALING FOR THE JAL FILTER

In this Appendix we show that, in the case of noisy postselection, the information amplification for the JAL filter is unchanged to first order in $\boldsymbol{\delta}$.

Referring back to the Taylor expansion in Appendix B, we now keep only the terms to first order in $\boldsymbol{\delta}$:

$$\hat{\rho}_{\boldsymbol{\theta}_0} = \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}] + O(|\boldsymbol{\delta}|^2), \tag{F1}$$

where $\hat{D} = -i[\nabla_{\boldsymbol{\theta}}\hat{U}(\boldsymbol{\theta})]^{\rm T}\boldsymbol{\delta}U(\boldsymbol{\theta})^\dagger$. Up to a term of order δ^2, the filter $\hat{F}$ can then be written as

$$\hat{F} = (t^2 - 1)\hat{\rho}_{\boldsymbol{\theta}_0} + \hat{1} = (t^2 - 1)\hat{\rho}_{\boldsymbol{\theta}} + \hat{1} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}](t^2 - 1). \tag{F2}$$

Denote the unperturbed eigenvectors (when $\boldsymbol{\delta} = 0$) by $|\lambda_1\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}\rangle/t$ and $|\lambda_n\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle$ (for $2 \leqslant n \leqslant d$), with corresponding eigenvalues

$$\lambda_1 = \frac{t^2}{P_{\boldsymbol{\theta}}^{\rm ps}}\left(1 - \epsilon + \frac{\epsilon}{d}\right), \quad \lambda_n = \frac{\epsilon}{dP_{\boldsymbol{\theta}}^{\rm ps}} \quad \text{for} \quad n \geqslant 2. \tag{F3}$$

We denote the true, perturbed eigenvectors and eigenvalues by $|\lambda_n'\rangle$, λ_n'. In the following derivation, we will find the leading order corrections in $\boldsymbol{\delta}$ of the perturbed eigenvectors and eigenvalues. Using these corrections, we will then calculate the perturbed QFIM and show it is unchanged to order $\boldsymbol{\delta}$.

Looking at the action of $\hat{\rho}_{\boldsymbol{\theta}}^{n,\rm ps}$ on $|\lambda_1\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}\rangle/t$, we see that

$$\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,\,ps}|\lambda_1\rangle \tag{F4}$$
$$= \frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + \frac{\epsilon}{d}\hat{F}^\dagger\right]\hat{K}\frac{|\psi_{\boldsymbol{\theta}}\rangle}{t}, \tag{F5}$$
$$= \frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\left[(1-\epsilon)\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{t}\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle + \frac{\epsilon}{d}\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{t}\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle + \frac{\epsilon}{d}\sum_n\frac{\hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle}{t}\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle\right], \tag{F6}$$
$$= \frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\left[(1-\epsilon)t^2|\lambda_1\rangle + \frac{\epsilon}{d}t^2|\lambda_1\rangle + i\sum_n\frac{\epsilon}{d\,t}(t^2-1)\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}]|\psi_{\boldsymbol{\theta}}\rangle|\lambda_n\rangle\right], \tag{F7}$$
$$= \frac{t^2\left(1-\epsilon+\frac{\epsilon}{d}\right)}{P_{\boldsymbol{\theta}}^{\rm ps}}|\lambda_1\rangle - i\frac{\epsilon(t^2-1)}{dt\,P_{\boldsymbol{\theta}}^{\rm ps}} \times \sum_n\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|\hat{D}|\psi_{\boldsymbol{\theta}}\rangle|\lambda_n\rangle. \tag{F8}$$

Carrying out the same calculation for $|\lambda_n\rangle = \hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle$

$$\hat{\rho}_{\boldsymbol{\theta}}^{\rm n,\,ps}|\lambda_n\rangle \tag{F9}$$
$$= \frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\left[(1-\epsilon)\hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger + \frac{\epsilon}{d}\hat{F}^\dagger\right]\hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle, \tag{F10}$$
$$= \frac{1}{P_{\boldsymbol{\theta}}^{\rm ps}}\left[(1-\epsilon)\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{t}t\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle + \frac{\epsilon}{d}\sum_m\hat{K}|\psi_{\boldsymbol{\theta}}{}^{\perp,m}\rangle\langle\psi_{\boldsymbol{\theta}}{}^{\perp,m}|\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle + \frac{\epsilon}{d}\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{t}t\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle\right], \tag{F11}$$
$$= i\frac{t\left(1-\epsilon+\frac{\epsilon}{d}\right)(t^2-1)}{P_{\boldsymbol{\theta}}^{\rm ps}}\langle\psi_{\boldsymbol{\theta}}|\hat{D}|\psi_{\boldsymbol{\theta}}{}^{\perp,n}\rangle|\lambda_1\rangle + \frac{\epsilon|\lambda_n\rangle}{dP_{\boldsymbol{\theta}}^{\rm ps}}. \tag{F12}$$

In the above calculations, we used the following identities:

$$\langle\psi_{\boldsymbol{\theta}}|[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}]|\psi_{\boldsymbol{\theta}}\rangle = \langle\psi_{\boldsymbol{\theta}}|\hat{\rho}_{\boldsymbol{\theta}}\hat{D} - \hat{D}\hat{\rho}_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle = \langle\psi_{\boldsymbol{\theta}}|\hat{D}|\psi_{\boldsymbol{\theta}}\rangle - i\langle\psi_{\boldsymbol{\theta}}|\hat{D}|\psi_{\boldsymbol{\theta}}\rangle = 0, \tag{F13}$$

$$\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|[\hat{\rho}_{\boldsymbol{\theta}}, D]|\psi_{\boldsymbol{\theta}}\rangle = \langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|\hat{\rho}_{\boldsymbol{\theta}}\hat{D} - \hat{D}\hat{\rho}_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle = -\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|\hat{D}|\psi_{\boldsymbol{\theta}}\rangle, \tag{F14}$$

$$\langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}]|\psi_{\boldsymbol{\theta}}{}^{\perp,m}\rangle = \langle\psi_{\boldsymbol{\theta}}{}^{\perp,n}|\hat{\rho}_{\boldsymbol{\theta}}\hat{D} - \hat{D}\hat{\rho}_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}{}^{\perp,m}\rangle = 0. \tag{F15}$$

Therefore, $\hat{\rho}_{\theta}^{\text{n, ps}}$ can be written in the $\{|\lambda_1\rangle, \ldots, |\lambda_d\rangle\}$ basis as

$$\hat{\rho}_{\theta}^{\text{n, ps}} = \begin{pmatrix} \lambda_1 & \lambda_{1,\bar{1}} & \ldots & \lambda_{1,\bar{1}} \\ \lambda_{\bar{1},1} & \lambda_{\bar{1}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\bar{1},1} & 0 & \ldots & \lambda_{\bar{1}} \end{pmatrix}, \tag{F16}$$

where

$$\lambda_1 = \frac{t^2}{P_{\theta}^{\text{ps}}}\left(1 - \epsilon + \frac{\epsilon}{d}\right), \tag{F17}$$

$$\lambda_{\bar{1}} = \frac{\epsilon}{d P_{\theta}^{\text{ps}}}, \tag{F18}$$

$$\lambda_{1,\bar{1}} = i\frac{t}{P_{\theta}^{\text{ps}}}\left(1 - \epsilon + \frac{\epsilon}{d}\right)(t^2 - 1)\langle\psi_{\theta}|\hat{D}|\psi_{\theta}{}^{\perp,n}\rangle, \tag{F19}$$

$$\lambda_{\bar{1},1} = -i\frac{\epsilon(t^2 - 1)}{dt\, P_{\theta}^{\text{ps}}}\langle\psi_{\theta}{}^{\perp,n}|\hat{D}|\psi_{\theta}\rangle. \tag{F20}$$

By solving $\det(\hat{\rho}_{\theta}^{\text{n, ps}} - \lambda\,\hat{1}) = 0$, we find the perturbed eigenvalues. This is equivalent to

$$[\lambda_1 - \lambda] \cdot [\lambda_{\bar{1}} - \lambda]^{d-1} = \lambda_{\bar{1},1}\,\lambda_{1,\bar{1}}\,[\lambda_{\bar{1}} - \lambda]^{d-2}(d-1). \tag{F21}$$

Hence we have

$$[\lambda_{\bar{1}} - \lambda]^{d-2}\{[\lambda_1 - \lambda][\lambda_{\bar{1}} - \lambda] - \lambda_{\bar{1},1}\,\lambda_{1,\bar{1}}\,(d-1)\} = 0. \tag{F22}$$

Because $\lambda_{\bar{1},1}\,\lambda_{1,\bar{1}}$ is order δ^2, it follows that the eigenvalues are the same up to a term of order δ^2, i.e., $\lambda_n' = \lambda_n + O(\delta^2)$.

From the matrix form of $\hat{\rho}_{\theta}^{\text{n, ps}}$, we see that the perturbed eigenvectors will have corrections of order δ

$$|\lambda_1'\rangle = |\lambda_1\rangle + \sum_{n\neq 1}\alpha_{1,n}|\lambda_n\rangle + O(\delta^2), \tag{F23}$$

$$|\lambda_n'\rangle = |\lambda_n\rangle + \alpha_n|\lambda_1\rangle + O(\delta^2) \quad \text{for} \quad n > 2, \tag{F24}$$

where α_i are terms of order δ.

We can now evaluate the first-order correction in δ of the QFIM. We use the formula

$$2\frac{a^2 p_{\theta}^2}{\left(P_{\theta}^{\text{ps}}\right)^2}\sum_{n,m}\frac{\langle\lambda_n'|(\partial_i\hat{\rho}_1)|\lambda_m'\rangle\langle\lambda_m'|(\partial_j\hat{\rho}_1)|\lambda_n'\rangle}{\lambda_n' + \lambda_m'}, \tag{F25}$$

where $\hat{\rho}_1 = |\lambda_1\rangle\langle\lambda_1|$ and λ_n', $|\lambda_n'\rangle$ are now the perturbed eigenvalues and eigenvectors. To evaluate $\hat{\rho}_{\theta}^{\text{n, ps}}$ we used Eq. (C34) and neglected the term proportional to $\partial_i p_{\theta}$, because it is second order in δ. We also used the fact that $p_{\theta} = t^2 + O(\delta^2)$. Note that the eigenvalues λ_n' are unchanged to order δ. Let us examine the following cases.

(i) $n = m = 1$: the first factor in the numerator is proportional to $\sum_{l\neq 1}\alpha_{1,l}^*\langle\lambda_l|\partial_i\lambda_1\rangle + \sum_{l\neq 1}\alpha_{1,l}\langle\partial_i\lambda_1|\lambda_l\rangle$, which is order δ. Indeed, notice that the first-order term cancels out: $\langle\lambda_1|\partial_i\lambda_1\rangle + \langle\partial_i\lambda_1|\lambda_1\rangle = \partial_i\langle\lambda_1|\lambda_1\rangle = 0$. When multiplied by the second factor (the j derivative), this gives a second-order contribution in δ.

(ii) $n, m \neq 1$: the first term in the numerator is proportional to $\alpha_m^*\langle\lambda_n|\partial_i\lambda_1\rangle + \alpha_n\langle\partial_i\lambda_1|\lambda_m\rangle$, which is first order in δ. Similarly, the second term is also first order in δ. The contribution from these terms is second order in δ.

(iii) $n \neq 1, m = 1$ (or vice versa): to leading order in δ, the first factor in the numerator is given by

$$\begin{aligned}&\langle\lambda_n|\partial_i\lambda_1\rangle + \alpha_n^*\langle\lambda_1|\partial_i\lambda_1\rangle + \alpha_n^*\langle\partial_i\lambda_1|\lambda_1\rangle \\ &= \langle\lambda_n|\partial_i\lambda_1\rangle,\end{aligned} \tag{F26}$$

where we used $\langle\lambda_1|\partial_i\lambda_1\rangle + \langle\partial_i\lambda_1|\lambda_1\rangle = \partial_i\langle\lambda_1|\lambda_1\rangle = 0$. Similarly, to order δ^2, the second factor in the numerator is $\langle\partial_j\lambda_1|\lambda_n\rangle$. When this term is multiplied by the first term in Eq. (F26), we recover the formula for the unperturbed QFIM calculated in the main text, to order δ^2.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, Phys. Rev. Lett. **96**, 010401 (2006).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222 (2011).
[3] H. Grote, K. Danzmann, K. L. Dooley, R. Schnabel, J. Slutsky, and H. Vahlbruch, First long-term application of squeezed states of light in a gravitational-wave observatory, Phys. Rev. Lett. **110**, 181101 (2013).
[4] M. Tse, Quantum-enhanced advanced LIGO detectors in the era of gravitational-wave astronomy, Phys. Rev. Lett. **123**, 231107 (2019).
[5] F. Acernese, Increasing the astrophysical reach of the advanced virgo detector via the application of squeezed vacuum states of light, Phys. Rev. Lett. **123**, 231108 (2019).
[6] H. Yonezawa, Quantum-enhanced optical-phase tracking, Science **337**, 1514 (2012).
[7] F. Wolfgramm, A. Cerè, F. A. Beduini, A. Predojević, M. Koschorreck, and M. W. Mitchell, Squeezed-light optical magnetometry, Phys. Rev. Lett. **105**, 053601 (2010).
[8] B.-B. Li, J. Bílek, U. B. Hoff, L. S. Madsen, S. Forstner, V. Prakash, C. Schäfermeier, T. Gehring, W. P. Bowen, and U. L. Andersen, Quantum enhanced optomechanical magnetometry, Optica **5**, 850 (2018).
[9] J. Borregaard and A. S. Sørensen, Near-Heisenberg-limited atomic clocks in the presence of decoherence, Phys. Rev. Lett. **111**, 090801 (2013).
[10] L. Pezzè and A. Smerzi, Heisenberg-limited noisy atomic clock using a hybrid coherent and squeezed state protocol, Phys. Rev. Lett. **125**, 210503 (2020).
[11] M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, Biological measurement beyond the quantum limit, Nat. Photon. **7**, 229 (2013).
[12] N. Lupu-Gladstein, Y. B. Yilmaz, D. R. M. Arvidsson-Shukur, A. Brodutch, A. O. T. Pang, A. M. Steinberg, and N. Y. Halpern, Negative quasiprobabilities enhance phase estimation in quantum-optics experiment, Phys. Rev. Lett. **128**, 220504 (2022).
[13] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Quantum advantage in postselected metrology, Nat. Commun. **11**, 3775 (2020).
[14] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100, Phys. Rev. Lett. **60**, 1351 (1988).

[15] O. Hosten and P. Kwiat, Observation of the spin Hall effect of light via weak measurements, Science **319**, 787 (2008).

[16] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, Ultrasensitive beam deflection measurement via interferometric weak value amplification, Phys. Rev. Lett. **102**, 173601 (2009).

[17] M. D. Turner, C. A. Hagedorn, S. Schlamminger, and J. H. Gundlach, Picoradian deflection measurement with an interferometric quasi-autocollimator using weak value amplification, Opt. Lett. **36**, 1479 (2011).

[18] M. Pfeifer and P. Fischer, Weak value amplified optical activity measurements, Opt. Express **19**, 16508 (2011).

[19] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Precision frequency measurements with interferometric weak values, Phys. Rev. A **82**, 063822 (2010).

[20] D. J. Starling, P. B. Dixon, N. S. Williams, A. N. Jordan, and J. C. Howell, Continuous phase amplification with a Sagnac interferometer, Phys. Rev. A **82**, 011802(R) (2010).

[21] X.-Y. Xu, Y. Kedem, K. Sun, L. Vaidman, C.-F. Li, and G.-C. Guo, Phase estimation with weak measurement using a white light source, Phys. Rev. Lett. **111**, 033604 (2013).

[22] O. S. Magaña-Loaiza, M. Mirhosseini, B. Rodenburg, and R. W. Boyd, Amplification of angular rotations using weak measurements, Phys. Rev. Lett. **112**, 200401 (2014).

[23] G. Strübi and C. Bruder, Measuring ultrasmall time delays of light by joint weak measurements, Phys. Rev. Lett. **110**, 083605 (2013).

[24] G. I. Viza, J. Martínez-Rincón, G. A. Howland, H. Frostig, I. Shomroni, B. Dayan, and J. C. Howell, Weak-values technique for velocity measurements, Opt. Lett. **38**, 2949 (2013).

[25] P. Egan and J. A. Stone, Weak-value thermostat with 0.2 mK precision, Opt. Lett. **37**, 4991 (2012).

[26] J. H. Jenne and D. R. M. Arvidsson-Shukur, Unbounded and lossless compression of multiparameter quantum information, Phys. Rev. A **106**, 042404 (2022).

[27] M. G. A. Paris, Quantum estimation for quantum technology, Int. J. Quantum Inf. **07**, 125 (2009).

[28] J. G. Smith, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Iterative quantum-phase-estimation protocol for shallow circuits, Phys. Rev. A **106**, 062615 (2022).

[29] W. Salmon, S. Strelchuk, and D. Arvidsson-Shukur, Only classical parameterised states have optimal measurements under least squares loss, Quantum **7**, 998 (2023).

[30] W. Salmon, S. Strelchuk, and D. Arvidsson-Shukur, James-Stein estimation in quantum Gaussian sensing, arXiv:2404.02203.

[31] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1999).

[32] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, in *Breakthroughs in Statistics: Foundations and Basic Theory*, edited by S. Kotz and N. L. Johnson (Springer, New York, 1992), pp. 235–247.

[33] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994).

[34] A. Fujiwara and H. Nagaoka, Quantum Fisher metric and estimation for pure state models, Phys. Lett. A **201**, 119 (1995).

[35] J. Liu, H. Yuan, X. Lu, and X. Wang, Quantum Fisher information matrix and multiparameter estimation, J. Phys. A Math. **53**, 023001 (2019).

[36] C. W. Helstrom, Quantum detection and estimation theory, J. Stat. Phys. **1**, 231 (1969).

[37] F. Albarelli, M. Barbieri, M. Genoni, and I. Gianani, A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging, Phys. Lett. A **384**, 126311 (2020).

[38] For a pure-state or a full rank-density matrix, getting within a factor of 2 requires a collective, entangled, measurement on many copies of the system, which may not be physically tractable [37]. Moreover, for general mixed states, there can be further complications. Nonetheless, the SLD QFI is a well established metric for the information content of a state.

[39] A. Fujiwara, Quantum channel identification problem, Phys. Rev. A **63**, 042304 (2001).

[40] S. Pang, J. Dressel, and T. A. Brun, Entanglement-assisted weak value amplification, Phys. Rev. Lett. **113**, 030401 (2014).

[41] S. Pang and T. A. Brun, Improving the precision of weak measurements by postselection measurement, Phys. Rev. Lett. **115**, 120401 (2015).

[42] N. J. Beaudry and R. Renner, An intuitive proof of the data processing inequality, Quant. Inf. Comput. **12**, 0432 (2012).

[43] S. Das, S. Modak, and M. N. Bera, Saturating quantum advantages in postselected metrology with the positive Kirkwood-Dirac distribution, Phys. Rev. A **107**, 042413 (2023).

[44] $\mathcal{U}$ is the space of states that can arise due to variations in the parameters—the "useful" subspace. When $\dim\mathcal{U} < d$, $\mathcal{U}^{\perp}$ carries no information about the parameters.

[45] We have that $\alpha \in \mathbb{C}$ may be a general complex number, but since $\langle\psi_{\boldsymbol{\theta}}| \equiv\rangle 1$, x must be purely real.