Minimum-entanglement protocols for function estimation

Adam Ehrenberg,* Jacob Bringewatt,* and Alexey V. Gorshkov
Joint Center for Quantum Information and Computer Science, NIST and University of Maryland College Park, Maryland 20742, USA and Joint Quantum Institute, NIST and University of Maryland College Park, Maryland 20742, USA

(Received 3 September 2022; revised 22 August 2023; accepted 6 September 2023; published 29 September 2023)

We derive a family of optimal protocols, in the sense of saturating the quantum Cramér-Rao bound, for measuring a linear combination of d field amplitudes with quantum sensor networks, a key subprotocol of general quantum sensor network applications. We demonstrate how to select different protocols from this family under various constraints. Focusing primarily on entanglement-based constraints, we prove the surprising result that highly entangled states are not necessary to achieve optimality in many cases. Specifically, we prove necessary and sufficient conditions for the existence of optimal protocols using at most k-partite entanglement. We prove that the protocols which satisfy these conditions use the minimum amount of entanglement possible, even when given access to arbitrary controls and ancillas. Our protocols require some amount of time-dependent control, and we show that a related class of time-independent protocols fail to achieve optimal scaling for generic functions.

DOI: 10.1103/PhysRevResearch.5.033228

I. INTRODUCTION

Entanglement is a hallmark of quantum theory and plays an essential role in many quantum technologies. Consider single-parameter metrology, where one seeks to determine an unknown phase θ that is independently and identically coupled to d sensors via a linear Hamiltonian $\hat{H}$. Given a probe state $\hat{\rho}$, evolution under $\hat{H}$ encodes θ into $\hat{\rho}$ where it can then be measured. If the sensors are classically correlated the ultimate attainable uncertainty is the so-called standard quantum limit $\Delta\theta \sim 1/\sqrt{d}$ [1], which can be surpassed only if the states are prepared in an entangled state [2,3]; if $O(d)$-partite entanglement is used, the Heisenberg limit $\Delta\theta \sim 1/d$ can be achieved [4–6]. The necessity of entanglement for optimal measurement has also been explored in numerous other contexts [7,8], for instance, in sequential measurement schemes (where one may apply the encoding unitary multiple times) [9,10], in the presence of decoherence [11–14], when the coupling Hamiltonian is nonlinear [15–17], or in reference to resource theories for metrology [18–21].

In this paper, we consider the amount of entanglement required to saturate the quantum Cramér-Rao bound, which lower-bounds the variance of measuring an unknown quantity [22–25], in the prototypical multiparameter setting of a quantum sensor network, where d independent, unknown parameters $\boldsymbol{\theta}$ (boldface denotes vectors) are each coupled to a unique quantum sensor. Specifically, we revisit the problem of optimally measuring a single linear function $q(\boldsymbol{\theta})$ [26–35], which is a crucial element of optimal protocols for more general quantum sensor network problems (the case of measuring one or multiple analytic functions [36,37] and the case where the parameters $\boldsymbol{\theta}$ are not independent [38] reduce asymptotically to the linear problem considered here). Therefore, we focus on measuring a single linear function of independent parameters for ease of presentation while emphasizing that our results generalize.

Given the similarity of measuring a single linear function to the single-parameter case and the fact that such functions of local parameters are global properties of the system, one might expect (provided all the local parameters nontrivially appear in q) that d-partite entanglement is necessary. This intuition is reinforced by the fact that all existing optimal protocols for this problem do, in fact, make use of d-partite entanglement [26,27,32].

We show that such intuition is faulty and only holds in the case where q is approximately an average of the unknown parameters. In particular, we derive a family of protocols that saturate necessary and sufficient algebraic conditions to achieve optimal performance in this setting, and we prove necessary and sufficient conditions on q for the existence of optimal protocols using at most $(k < d)$-partite entanglement. The more uniformly distributed q is among the unknown parameters, the more entanglement is required. We also consider other resources of interest, such as the average entanglement used over the course of the protocol, as well as the number of entangling gates needed to perform these protocols, and discuss optimizing them within our scheme.

Finally, we address the impracticality of certain assumptions that have typically been made in the more theoretically focused literature on function estimation protocols. Specifically, we show that so-called probabilistic protocols fail to achieve the Heisenberg limit except for a narrow class of functions.

*These authors contributed equally.

II. PROBLEM SETUP

We first briefly review the problem of measuring a linear function of unknown parameters in a quantum sensor network [26,27,29–32]. Consider a network of d qubit quantum sensors coupled to d independent, unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ via a Hamiltonian of the form

$$\hat{H}(s) = \sum_{i=1}^{d} \frac{1}{2}\theta_i \hat{\sigma}_i^z + \hat{H}_c(s), \tag{1}$$

where $\hat{\sigma}_i^{x,y,z}$ are the Pauli operators acting on qubit i and $\hat{H}_c(s)$ for $s \in [0, t]$ is any choice of time-dependent, $\boldsymbol{\theta}$-independent control Hamiltonian, potentially including coupling to an arbitrary number of ancillas. That is, $\hat{H}_c(s)$ accounts for any possible parameter-independent contributions to the Hamiltonian, including those acting on any extended Hilbert space with a (finite) dimension larger than that of the network of d qubit sensors directly coupled to the unknown parameters [39]. We encode the parameters $\boldsymbol{\theta}$ into a quantum state $\hat{\rho}$ via the unitary evolution generated by a Hamiltonian of this form for a time t. Given some choices of initial probe state, control $\hat{H}_c(s)$, final measurement, and classical postprocessing, we seek to construct an estimator for a linear combination $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ of the unknown parameters, where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is a set of known coefficients. Throughout this paper, we assume without loss of generality that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Reference [27] established that the fundamental limit for the mean-square error $\mathcal{M}$ of an estimator for q is

$$\mathcal{M} \geqslant \frac{\|\boldsymbol{\alpha}\|_\infty^2}{t^2}, \tag{2}$$

where t is the total evolution time.

Equation (2) is derived via the single-parameter quantum Cramér-Rao bound [15,22–25]. This is somewhat surprising: while we seek to measure only a single quantity $q(\boldsymbol{\theta})$, d parameters control the evolution under Eq. (1), so we do not *a priori* satisfy the conditions for the use of the single-parameter quantum Cramér-Rao bound. However, we can justify its validity for our system: consider an infinite set of imaginary scenarios, each corresponding to a choice of artificially fixing $d-1$ degrees of freedom and leaving only $q(\boldsymbol{\theta})$ free to vary. Under any such choice, our final quantum state depends on a single parameter q, and we can apply the single-parameter quantum Cramér-Rao bound. While this requires giving ourselves information that we do not have, additional information can only reduce $\mathcal{M}$, and, therefore, any such choice provides a lower bound on $\mathcal{M}$ when we do not have such information. To obtain the tightest possible bound there must be some choice(s) of artificially fixing $d-1$ degrees of freedom that gives us no (useful) information about $q(\boldsymbol{\theta})$. We will derive algebraic conditions that characterize such choices.

Thus, we may apply the single-parameter quantum Cramér-Rao bound

$$\mathcal{M} \geqslant \frac{1}{\mathcal{F}(q)} \geqslant \frac{1}{t^2 \|\hat{g}_q\|_s^2}, \tag{3}$$

where $\mathcal{F}$ is the quantum Fisher information, $\hat{g}_q = \partial\hat{H}/\partial q$ (the partial derivative fixes the other $d-1$ degrees of freedom), and the seminorm $\|\hat{g}_q\|_s$ is the difference of the largest and smallest eigenvalues of $\hat{g}_q$ [15]. For our problem, the best choice of fixing extra degrees of freedom—in the sense of yielding the tightest bound via Eq. (3)—gives $\|\hat{g}_q\|_s^2 = 1/\|\boldsymbol{\alpha}\|_\infty^2$, yielding Eq. (2) [27]. The proof of this fact is provided in Appendix F for completeness.

III. CONDITIONS FOR SATURABLE BOUNDS

While the argument above justifies applying the single-parameter bound in our multiparameter scenario, it offers no roadmap for constructing optimal protocols. The quantum Fisher information matrix $\mathcal{F}(\boldsymbol{\theta})$ provides an information-theoretic solution to this issue. When calculating $\mathcal{F}(\boldsymbol{\theta})$ we restrict to pure probe states, as the convexity of the quantum Fisher information matrix implies mixed states fail to produce optimal protocols [40,41]. For pure probe states and unitary evolution for time t under the Hamiltonian in Eq. (1), it has matrix elements [41]

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4\left[\tfrac{1}{2}\langle\{\hat{\mathcal{H}}_i(t), \hat{\mathcal{H}}_j(t)\}\rangle - \langle\hat{\mathcal{H}}_i(t)\rangle\langle\hat{\mathcal{H}}_j(t)\rangle\right], \tag{4}$$

where $\{\cdot,\cdot\}$ denotes the anticommutator and

$$\hat{\mathcal{H}}_i(t) = -\int_0^t ds \hat{U}^\dagger(s)\hat{g}_i\hat{U}(s), \tag{5}$$

with $\hat{g}_i = \partial\hat{H}/\partial\theta_i = \hat{\sigma}_j^z/2$ and $\hat{U}$ the time-ordered exponential of $\hat{H}$. The expectation values in Eq. (4) are taken with respect to the initial probe state.

Choosing $d-1$ degrees of freedom to fix in hopes of using the single-parameter bound then corresponds to a basis transformation $\boldsymbol{\theta} \to \boldsymbol{q}$, where we take $q_1 = q$ to be our quantity of interest, and the other arbitrary $q_{j>1}$ are the extra degrees of freedom. This basis transformation has a corresponding Jacobian J such that $\mathcal{F}(\boldsymbol{q}) = J^\top\mathcal{F}(\boldsymbol{\theta})J$. To obtain the bound in Eq. (2) and have no information about $q(\boldsymbol{\theta})$ from the extra degrees of freedom $q_{j>1}$, $\mathcal{F}(\boldsymbol{q})$ must have the following properties:

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{6}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1) \tag{7}$$

(recall $|\alpha_1| = \|\boldsymbol{\alpha}\|_\infty$ without loss of generality). Via the inverse basis transformation $\boldsymbol{q} \to \boldsymbol{\theta}$, we find Eqs. (6) and (7) are satisfied if and only if

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = \mathcal{F}(\boldsymbol{\theta})_{j1} = \frac{\alpha_j}{\alpha_1}t^2, \tag{8}$$

where we assume here and for the rest of the main text that $|\alpha_1| > |\alpha_j|\ \forall j > 1$ for ease of presentation. Our main result (see Theorem 1) is unchanged by this assumption, although its proof and that of several other results becomes more tedious. The explicit derivation of Eq. (8), along with the generalization of our results beyond this assumption, is provided in Appendix F.

Finally, we remark that the problem of function estimation is mathematically equivalent to the concept of nuisance parameters in the literature on classical (cf. Ref. [42]) and quantum estimation theory [43–45]. One finds similarly derived bounds in these contexts [46]. However, the protocols

we now describe, and especially their entanglement features, are new to this work.

IV. A FAMILY OF OPTIMAL PROTOCOLS

We now derive a family of protocols that achieve Eq. (8). A particular protocol consists of preparing a pure initial state $\hat{\rho}_0 = |\psi(0)\rangle\langle\psi(0)|$, evolving $\hat{\rho}_0$ under the unitary generated by $\hat{H}(s)$ for time t, performing some positive operator-valued measurement, and computing an estimator for q from the measurement outcomes. Given $\hat{\rho}_0$ and $\hat{H}(s)$, $\mathcal{F}(\boldsymbol{\theta})$ can be computed via Eq. (4).

The protocols we propose will use $\hat{H}_c(s)$ to coherently switch between probe states with different sensitivities to the unknown parameters $\boldsymbol{\theta}$, thereby accumulating an overall sensitivity to the unknown function of interest, q. In particular, we consider the following set $\mathcal{T}$ of $N = 3^{d-1}$ one-parameter families of catlike states:

$$|\psi(\boldsymbol{\tau};\varphi)\rangle = \frac{1}{\sqrt{2}}(|\boldsymbol{\tau}\rangle + e^{i\varphi}|-\boldsymbol{\tau}\rangle), \tag{9}$$

where each family of states is labeled by a vector $\boldsymbol{\tau} \in \{0, \pm 1\}^d$ such that

$$|\boldsymbol{\tau}\rangle = \bigotimes_{j=1}^{d} \begin{cases} |0\rangle, & \tau_j \neq -1 \\ |1\rangle, & \tau_j = -1, \end{cases} \tag{10}$$

and $\varphi \in \mathbb{R}$ parametrizes individual states in the family. We require that $\tau_1 = 1$, as any optimal protocol must always be sensitive to this most important parameter (see Lemma 2 in Appendix A). Each of the probe states described in Eqs. (9) and (10) is a superposition of exactly two states in the $\hat{\sigma}^z$ basis (which we call "branches"). Note that these states use no ancilla.

Our protocols proceed in three main stages: a state initialization stage, a parameter encoding stage, and, finally, a measurement stage. In the state initialization stage, we prepare the probe state $|\psi(\boldsymbol{\tau};0)\rangle$ that is then coupled to the parameters in the parameter encoding stage via a Hamiltonian of the form of Eq. (1). During this parameter encoding stage, we use the control Hamiltonian to coherently switch between families of probe states at particular (optimized) times, such that the relative phase between the branches is preserved during the switches [that is, $\hat{H}_c(s)$ changes $\boldsymbol{\tau}$, but not φ]. This can be done using finitely many controlled-NOT (CNOT) and $\hat{\sigma}^x$ gates. We stay in the family of states $|\psi(\boldsymbol{\tau}^{(n)};\varphi)\rangle$ for time $p_n t$, where $p_n \in [0, 1]$ such that $\sum_n p_n = 1$. Here n indexes some enumeration of the families of states in $\mathcal{T}$. There are three possibilities for the relative phase that qubit j induces between the two branches due to the time spent in family n. If $\tau_j^{(n)} = 0$, then no relative phase is accrued because qubit j is disentangled. If $\tau_j^{(n)} = 1$, the relative phase imprinted by $\hat{\sigma}_j^z/2$ is $p_n\theta_j t$, while if $\tau_j^{(n)} = -1$, the relative phase is $-p_n\theta_j t$. Thus, the jth qubit always induces a relative phase of $p_n\tau_j^{(n)}\theta_j t$. Accounting for all qubits, being in family n for time $p_n t$ induces a relative phase

$$\phi_n = \sum_j p_n t \tau_j^{(n)} \theta_j. \tag{11}$$

Given some time-dependent probe $|\psi(t)\rangle$ which is in each family $|\psi(\boldsymbol{\tau}^{(n)};\varphi)\rangle$ for time $p_n t$, the total phase ϕ accumulated between the branches over the course of the entire parameter encoding stage of the protocol is

$$\phi = \sum_n \phi_n = \sum_n \sum_j p_n t \tau_j^{(n)} \theta_j = \sum_j (T\boldsymbol{p})_j \theta_j t, \tag{12}$$

where we implicitly defined $\boldsymbol{p} = (p_1, \ldots, p_N)^\top$ and the $d \times N$ matrix T with matrix elements $T_{mn} = \tau_m^{(n)}$. If $\boldsymbol{p}$ is chosen such that $T\boldsymbol{p} \propto \boldsymbol{\alpha}$ this total phase is $\propto qt$. More formally, choosing $\boldsymbol{p}$ such that

$$T\boldsymbol{p} = \frac{\boldsymbol{\alpha}}{\alpha_1} \tag{13}$$

achieves the saturability condition in Eq. (8). Algebraic details of this calculation are provided in Appendix B.

Any non-negative solution (in the sense that $p_n \geqslant 0\ \forall n$) to Eq. (13) specifies a valid set of states and evolution times satisfying Eq. (8). Because the system in Eq. (13) is highly underconstrained, such protocols do not necessarily use all 3^{d-1} families of states in $\mathcal{T}$. As an illustrative example, consider the solutions to Eq. (13) for two qubits. The available families of states are described by

$$T = (\boldsymbol{\tau}^{(1)} \quad \boldsymbol{\tau}^{(2)} \quad \boldsymbol{\tau}^{(3)}) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}. \tag{14}$$

By Eq. (13), the fraction of time spent in each family of states must satisfy

$$p_1 + p_2 + p_3 = 1, \tag{15}$$

$$p_1 - p_2 = \frac{\alpha_2}{\alpha_1}. \tag{16}$$

Solving in terms of p_1 leads to the one-parameter family of solutions $p_2 = p_1 - \frac{\alpha_2}{\alpha_1}$ and $p_3 = 1 + \frac{\alpha_2}{\alpha_1} - 2p_1$, where $p_n \in [0, 1]$ for all n. Without loss of generality, assume $\alpha_1 = 1$. Then non-negativity is achieved by

$$p_1 \in \begin{cases} \left[\alpha_2, \frac{1+\alpha_2}{2}\right], & \alpha_2 \geqslant 0 \\ \left[0, \frac{1+\alpha_2}{2}\right], & \alpha_2 < 0. \end{cases} \tag{17}$$

There are many solutions satisfying these constraints. Of particular note, there is a two-family protocol that does not require using exclusively maximally entangled states: for $\alpha_2 > 0$, let $p_1 = \alpha_2$ so that $p_2 = 0$ and $p_3 = 1 - \alpha_2$; for $\alpha_2 < 0$, let $p_1 = 0$ so that $p_2 = -\alpha_2$ and $p_3 = 1 + \alpha_2$.

We refer to protocols achieving Eq. (13) [or, equivalently, Eq. (8)] as optimal. Note, however, that achieving these conditions is a property of the probe state(s) used and does not *a priori* guarantee the existence of measurements to extract q. Therefore, we now move on to describing the third main stage of our protocols, which is the explicit measurement scheme: apply a sequence of $\hat{\sigma}_i^x$ and CNOT gates to the final state of a protocol to transform it into $1/\sqrt{2}(|0\rangle + e^{iqt/\alpha_1}|1\rangle)(|0\cdots 0\rangle)$. Then perform single-qubit phase estimation to measure q [47].

Such phase estimation is not as simple as it might appear, however. Because we are interested in how our error scales in the $t \to \infty$ limit, a naive approach loses track of which 2π interval the phase is in [48–50]. We could assume that

this information is known *a priori* [27], but this is unjustified in practice as the required knowledge is of precision $\sim|\alpha_1|/t$; i.e., it is already within the Heisenberg limit. More realistically, starting with any t-independent prior knowledge of the unknown phase, we use the so-called robust phase estimation protocols from Refs. [51–53] to saturate Eq. (2) up to a modest constant factor. Such protocols work by optimally dividing the total time t into K stages with stage k using a time $2\nu_k t_k$ such that $2\sum_{k=1}^{K}\nu_k t_k = t$. In each stage, one encodes the parameters into the state for a time t_k and then makes a ($\hat{\sigma}^x$ or $\hat{\sigma}^y$) measurement. This is repeated $2\nu_k$ times in order to obtain an estimate of q, which in each stage becomes a more and more precise estimate. Provided the time of the final stage scales linearly with the total time, i.e., $t_K \sim t$, Heisenberg scaling in time is still achieved and we can estimate q with a mean-square error achieving the bound in Eq. (2) up to a constant factor. For completeness, we review this measurement scheme in more detail in Appendix C.

To summarize, a full optimal protocol is as follows:

(1) Using any relevant experimental desiderata and optimization algorithm, find a non-negative solution $\boldsymbol{p}$ to Eq. (13).

(2) Restrict $\boldsymbol{p}$ to its $\overline{N}$ nonzero elements, and restrict T to the corresponding columns. If desired, reorder the elements of $\boldsymbol{p}$ and the columns of T. The $\overline{N}$ $\boldsymbol{\tau}$ corresponding to the columns of T will be the families of states used in the protocol.

(3) Initialize a quantum state on the d qubits to $|0\rangle^{\otimes d}$.

(4) Using CNOT and $\hat{\sigma}^x$ gates, prepare $|\psi(\boldsymbol{\tau}^{(1)};0)\rangle$, the first state of the protocol. Couple the state to the Hamiltonian $\hat{H}$ and remain in this family for time $p_1 t_k$, leading to state $|\psi(\boldsymbol{\tau}^{(1)};\phi_1)\rangle$, where $\phi_1 = \sum_j p_1 t_k \tau_j^{(1)}\theta_j$. Here, t_k is the time required by the current step of the robust phase estimation protocol.

(5) Using CNOT and $\hat{\sigma}^x$ gates, coherently switch to $|\psi(\boldsymbol{\tau}^{(2)};\phi_1)\rangle$ from $|\psi(\boldsymbol{\tau}^{(1)};\phi_1)\rangle$. Remain in this family for time $p_2 t_k$, leading to state $|\psi(\boldsymbol{\tau}^{(2)};\phi_1+\phi_2)\rangle$, with $\phi_2 = \sum_j p_2 t_k \tau_j^{(2)}\theta_j$.

(6) Repeat this process for all states in the restricted T, staying in the family parametrized by $\boldsymbol{\tau}^{(n)}$ for time $p_n t_k$, leading to a final state $|\psi(\boldsymbol{\tau}^{(\overline{N})};qt_k)\rangle$.

(7) Using CNOT and $\hat{\sigma}^x$ gates, convert this final state to $1/\sqrt{2}(|0\rangle + e^{iqt_k}|1\rangle)|0\rangle^{\otimes d-1}$.

(8) Make a measurement on the first qubit of the final state (see Appendix C for more details) and repeat starting from step 3. After $2\nu_k$ repetitions, move to the next stage of the robust phase estimation protocol, and use an updated evolution time t_k. After a number of stages, K, as prescribed by the robust phase estimation protocol, extract a final estimate of q with a mean-square error achieving the bound in Eq. (2) up to a constant factor.

Having described the full details of the protocol, including the subtleties involved in subdividing the total time t into different stages in order to implement robust phase estimation, in the rest of the paper, for simplicity of presentation, we will simply consider the total encoding time t and act as if the parameters can be encoded into the state in one step, using evolution for this full time. This should be viewed as a notational shorthand such that t can be replaced with the relevant t_k at any given stage when implementing the full protocol.

V. MINIMUM-ENTANGLEMENT SOLUTIONS

We now focus on solutions from our family of protocols that require the minimum amount of entanglement. Specifically, we prove necessary and sufficient conditions on $\boldsymbol{\alpha}$ for the existence of a protocol that uses at most k-partite entanglement. This is the primary technical result of our paper. We emphasize that, while the protocols in the previous section use a particular choice of controls that does not include ancilla qubits, Theorem 1 applies to any protocol making use of a Hamiltonian described via Eq. (1).

Theorem 1 (Main result). Let $q(\boldsymbol{\theta}) = \boldsymbol{\alpha}\cdot\boldsymbol{\theta}$. Without loss of generality, let $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $k \in \mathbb{Z}^+$ so that

$$k - 1 < \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} \leqslant k. \tag{18}$$

An optimal protocol to estimate $q(\boldsymbol{\theta})$, where the parameters $\boldsymbol{\theta}$ are encoded into the probe state via unitary evolution under the Hamiltonian in Eq. (1), requires at least, but no more than, k-partite entanglement.

Theorem 1 justifies our claim that d-partite entanglement is not necessary unless $\|\boldsymbol{\alpha}\|_1$ is large enough, i.e., in the case of measuring an average ($\alpha_i = \frac{1}{d}\ \forall\, i$). We now sketch the proof, providing full details in Appendix D. The proof comes in two parts. First, using k-partite-entangled states from the set of catlike states considered above, we show the existence of an optimal protocol, subject to the upper bound of Eq. (18). Second, we show that, subject to the conditions in the theorem statement, there exists no optimal protocol using at most $(k-1)$-partite entanglement, proving the lower bound of Eq. (18).

Part 1. Define $T^{(k)}$ to be the submatrix of T with all columns n such that $\sum_m |T_{mn}| > k$ are eliminated, which enforces that any protocol derived from $T^{(k)}$ uses only states that are at most k-partite entangled. Define system $A(k)$ as

$$T^{(k)}\boldsymbol{p}^{(k)} = \boldsymbol{\alpha}/\alpha_1, \tag{19}$$

$$\boldsymbol{p}^{(k)} \geqslant 0. \tag{20}$$

Let $\boldsymbol{\alpha}' = \boldsymbol{\alpha}/\alpha_1$ and define system $B(k)$ as

$$(T^{(k)})^\top \boldsymbol{y} \geqslant 0, \tag{21}$$

$$\langle \boldsymbol{\alpha}', \boldsymbol{y}\rangle < 0. \tag{22}$$

By the Farkas-Minkowski lemma [54,55], system $A(k)$ has a solution if and only if system $B(k)$ does not, so it is sufficient to show that system $B(k)$ does not have a solution if $\sum_{j>1}|\alpha'_j| \leqslant k-1$, where we used that $\alpha'_1 = 1$. This can be shown by contradiction.

Part 2. The probe state must always be maximally sensitive to the first sensor qubit (see Lemma 2 in Appendix A), so $\mathcal{F}(\boldsymbol{\theta})_{1j}$ only accumulates in magnitude when qubit j is entangled with the first qubit [intuitively, Eq. (4) is similar to a connected correlator]. Using this, we show that satisfying the condition in Eq. (8) requires $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty > k-1$.

Theorem 1 provides conditions for the existence of solutions to Eq. (13) with limited entanglement, but it is not

constructive. To obtain an explicit protocol, simply solve the system of linear equations $T^{(k)}\boldsymbol{p} = \boldsymbol{\alpha}$.

Of course, instantaneous entanglement is not the only resource that one might want to minimize. For instance, one might also be interested in minimizing average entanglement over the entire protocol. This possibility is considered in Sec. VI. Other, more general, resource restrictions can be handled by setting up a constrained optimization problem that picks out certain solutions to the system of linear equations $T^{(k)}\boldsymbol{p} = \boldsymbol{\alpha}$ subject to a cost function $\mathcal{E}(\boldsymbol{p})$. For example, if certain pairs of sensors are easier to entangle than others, due to, for instance, their relative spatial location in the network, that could be encoded into $\mathcal{E}(\boldsymbol{p})$. More complicated optimizations could also take into consideration the ordering of the states used in the protocols. For example, because our protocols require coherently applying CNOT gates to move between different families of entangled states, and these gates may be costly or error-prone resources, one might wish to find protocols that minimize the usage of these gates. We discuss this possibility and the potential tradeoff between minimizing entanglement and CNOT gates in Sec. VII.

VI. AVERAGE ENTANGLEMENT

As mentioned above, one might also wish to minimize not just the size of the most-entangled family of states, but also the average entanglement used (given by weighting the size of each entangled family by the proportion of time that the family is used in the protocol). In this section (with some details deferred to Appendix E), we show that there exists a class of optimal protocols, ones that we name "non-echoed," that minimize this average entanglement. The formal definition is as follows:

Definition 2 (Non-echoed protocols). Consider some $\boldsymbol{\alpha} \in \mathbb{R}^d$ encoding a linear function of interest. Let T be the matrix which describes our families of catlike probe states, and let $\boldsymbol{p}$ specify a valid protocol such that $\boldsymbol{p} \geqslant 0$ and $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. We say that the protocol defined by $\boldsymbol{p}$ is "non-echoed" if, $\forall i$ such that p_i is strictly greater than 0, $\mathrm{sgn}(T_{ij}) \in \{0, \mathrm{sgn}(\alpha_j)\}$.

At any stage of a non-echoed protocol, letting the portion of the relative phase accumulated between the two branches of the probe state associated to the parameter θ_i be given by $c_i\theta_i$, two conditions must hold: (1) $|c_i| < |\alpha_i|$ and (2) $\mathrm{sgn}(c_i) = \mathrm{sgn}(\alpha_i)$. More intuitively, sensitivity to each parameter is accumulated "in the correct direction" at all times, meaning one does not use any sort of spin echo to produce a sensitivity to the function of interest, hence the name "non-echoed."

We now prove two useful statements about non-echoed protocols.

Lemma 1. Non-echoed protocols use minimum average entanglement.

Proof. We start with $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. Then

$$\begin{aligned}\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty &= \mathrm{sgn}(\boldsymbol{\alpha})^\top(T\boldsymbol{p}) \\ &= (\mathrm{sgn}(\boldsymbol{\alpha})^\top T)\boldsymbol{p} = \boldsymbol{w}^\top\boldsymbol{p}, \end{aligned} \tag{23}$$

where we have defined $w_j = \sum_i |T_{ij}|$ to be the sum of the absolute value of the elements of the jth column of T. That is, w_j represents how entangled the corresponding catlike family of states is. But, then, clearly $\boldsymbol{w}^\top\boldsymbol{p}$ is the average entanglement of the entire protocol. Furthermore, the second half of the proof of Theorem 1, given in Appendix D, shows that the minimum average entanglement of any optimal protocol is given by $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty$ (see the discussion after the completion of the proof). ■

The intuition behind this lemma is that if one always accumulates phase in the "correct direction," then the total amount of entanglement used over the course of the protocol must be minimized, as any extra entanglement would lead to becoming overly sensitive to some parameter, which would require some sort of echo to correct.

We further have the following theorem, which can be viewed as an extension of Theorem 1.

Theorem 3. For any $\boldsymbol{\alpha} \in \mathbb{R}^d$, there exists an optimal non-echoed protocol with minimum instantaneous entanglement for measuring $q = \boldsymbol{\alpha}\cdot\boldsymbol{\theta}$.

The proof of this theorem is given in Appendix E, and it proceeds in a very similar way to the proof of Theorem 1. The main difference is that one also restricts the allowed state families to be those with the correct sign so as to be non-echoed. And, analogously to how one can find a protocol with minimum entanglement, one can also obtain a solution that minimizes average entanglement by restricting T to only include columns such that $\mathrm{sgn}(T_{ij}) = \mathrm{sgn}(\alpha_i)$ for all i, j and then solving the corresponding system of linear equations.

VII. CNOT COSTS OF MINIMUM-ENTANGLEMENT PROTOCOLS

We now address another resource of potential interest: how many entangling (CNOT) gates are required to perform our protocols with a focus on the minimum-entanglement protocols.

We will again assume, for simplicity, that $\|\boldsymbol{\alpha}\|_\infty = \alpha_1 = 1 > |\alpha_2| \geqslant |\alpha_3| \geqslant \cdots \geqslant |\alpha_d|$. Furthermore, without loss of generality, we will adopt the convention that an optimal protocol specified by a $\boldsymbol{p} \geqslant 0$ such that $T\boldsymbol{p} = \boldsymbol{\alpha}$ begins by preparing a state in the family described by the first column of T and evolving for time $p_1 t$, and then proceeds to the appropriate state (i.e., the one with phase $p_1 t$) in the family described by the second column, then evolving for time $p_2 t$, and so on, until eventually moving to the measurement state. If $p_i = 0$, the corresponding state family is skipped and not prepared. By construction, the number of CNOT gates needed to perform this protocol is the number of gates required to generate the first state, plus the number needed to convert from the first state to the second state, and so on. Finally, one should add the number of gates needed to prepare the measurement state, which disentangles all qubits, from the final probe state [56]. The number of gates required to move from state i to state $i+1$ corresponds to the number of elements of $\boldsymbol{\tau}^{(i)}$ that are ± 1 but 0 in $\boldsymbol{\tau}^{(i+1)}$ and vice versa. In what follows, we will often consider only the gates that are used to convert between probe states (i.e., we will not consider the initial state preparation or final measurement preparation). This is physically motivated by the fact that these intermediate gates may be more difficult to perform or may be more susceptible to noise. Furthermore, assuming one is interested in the value of q at some particular moment (and not, say, continuously), one might be free to prepare and purify the initial probe state in advance of the

actual sensing task, which also justifies ignoring the initial CNOT cost.

Assume that $\overline{N}$ states are used in the protocol, i.e., $\boldsymbol{p}$ is such that it contains at most $\overline{N}$ nonzero elements. It is clear that at most $\mathcal{O}(\overline{N}^2)$ CNOT gates are needed. However, this is not necessarily optimal. In fact, Ref. [27] provides a protocol that uses d states and only $(d-1)=\mathcal{O}(d)$ intermediate CNOT gates. This "disentangling protocol" consists of using a maximally entangled Greenberger-Horne-Zeilinger state (up to $\hat{\sigma}^x$ gates) for a time $|\alpha_d|t$, then disentangling the last qubit and using the $(d-1)$-entangled state for time $(|\alpha_{d-1}|-|\alpha_d|)t$ before disentangling the next-to-last qubit and so on until reaching the final state corresponding to $\boldsymbol{\tau}=(1,0,\ldots,0)^\top$. This final state is used for time $(|\alpha_1|-|\alpha_2|)t=(1-|\alpha_2|)t$. The disentangling protocol does not minimize the instantaneous entanglement, but it does minimize average entanglement (as it is a non-echoed protocol—see Sec. VI).

Even more interestingly, Ref. [27] also provides a protocol, which we refer to as the "echoing" protocol, that uses *zero* intermediate CNOT gates. It proceeds by using d exclusively maximally entangled states (thereby minimizing neither average nor, in most cases, instantaneous entanglement), but judiciously echoing away the extra sensitivity that this extra entanglement induces.

To illustrate these protocols in the language of the current paper, we provide T and $\boldsymbol{p}$ (where, for simplicity of notation, we restrict T and $\boldsymbol{p}$ to the states that are used for a nonzero fraction of time) for the case $d=8$ and $\alpha_i>0$:

$$T^{\text{disentangling}}=\begin{pmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&0\\1&1&1&1&1&1&0&0\\1&1&1&1&1&0&0&0\\1&1&1&1&0&0&0&0\\1&1&1&0&0&0&0&0\\1&1&0&0&0&0&0&0\\1&0&0&0&0&0&0&0\end{pmatrix},\qquad \boldsymbol{p}^{\text{disentangling}}=\begin{pmatrix}\alpha_8\\\alpha_7-\alpha_8\\\alpha_6-\alpha_7\\\alpha_5-\alpha_6\\\alpha_4-\alpha_5\\\alpha_3-\alpha_4\\\alpha_2-\alpha_3\\\alpha_1-\alpha_2\end{pmatrix}, \tag{24}$$

and

$$T^{\text{echoing}}=\begin{pmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&-1\\1&1&1&1&1&1&-1&-1\\1&1&1&1&1&-1&-1&-1\\1&1&1&1&-1&-1&-1&-1\\1&1&1&-1&-1&-1&-1&-1\\1&1&-1&-1&-1&-1&-1&-1\\1&-1&-1&-1&-1&-1&-1&-1\end{pmatrix},\qquad \boldsymbol{p}^{\text{echoing}}=\begin{pmatrix}\frac{1+\alpha_8}{2}\\\frac{\alpha_7-\alpha_8}{2}\\\frac{\alpha_6-\alpha_7}{2}\\\frac{\alpha_5-\alpha_6}{2}\\\frac{\alpha_4-\alpha_5}{2}\\\frac{\alpha_3-\alpha_4}{2}\\\frac{\alpha_2-\alpha_3}{2}\\\frac{\alpha_1-\alpha_2}{2}\end{pmatrix}. \tag{25}$$

In the case of the disentangling protocol, the number of CNOT gates needed is heavily dependent on the ordering of the states. For example, consider, instead, ordering the states in the following way:

$$T^{\text{disentangling}}=\begin{pmatrix}1&1&1&1&1&1&1&1\\1&0&1&1&1&1&1&1\\1&0&1&0&1&1&1&1\\1&0&1&0&1&0&1&1\\1&0&1&0&1&0&1&0\\1&0&1&0&1&0&0&0\\1&0&1&0&0&0&0&0\\1&0&0&0&0&0&0&0\end{pmatrix}. \tag{26}$$

Here, the number of CNOT gates required is now $(d-1)+(d-2)+\cdots+1=\Theta(d^2)$. Thus, it is not only the choice of states that affects the CNOT cost of a protocol, but also their ordering. Naively, finding an optimal set of states and their optimal ordering is a difficult problem, as if one finds a protocol using $\overline{N}$ states, there are $\overline{N}!$ orders to check.

While we were unable to find a general solution to this optimization problem, numerics allow us to provide a pragmatic analysis of the cost. To begin, we considered the naive approach of finding a random (non-echoed) minimum entanglement solution using d states for random problem instances and, then, using this solution set, we brute-force searched over all column orderings of T restricted to families of states specified by this solution to find an optimal ordering in terms of CNOT cost. This was done for $d\in[3,10]$ sensors with 20 random instances each. Without loss of generality, the random problem instances were taken to have all positive coefficients. We observe a CNOT cost scaling $\sim d^2$, indicating that a random minimum-entanglement solution, even with optimal ordering, does not have the optimal linear-in-d scaling. See Fig. 1.

Consequently, more nuanced algorithms for finding a minimum-entanglement solution with better CNOT costs are desirable. To this end, we considered a greedy algorithm that yields a $\Theta(d)$ CNOT cost whenever it does not fail. The algorithm works by building up the full sensitivity to one parameter before switching coherently to a new state family (in this way, it is non-echoed—see Sec. VI). Consequently, each time we switch to a new state, one sensor qubit can be disentangled and never reentangled. In particular, we seek to build up sensitivity to the parameters according to their weight in q; i.e., we build up sensitivity to parameters going from the smallest corresponding $|\alpha_j|$ to the largest. The full algorithm is completed in at most d steps [57].

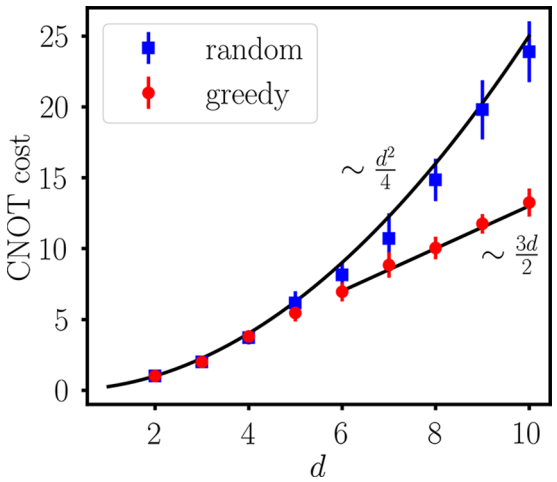

FIG. 1. CNOT costs versus number of sensors d for minimum-entanglement protocols. We consider twenty randomly chosen function instances that do not fail to yield a valid protocol via the greedy algorithm and compare the CNOT costs of the greedy algorithm to a random solution whose states are then ordered optimally. When it returns a valid protocol, the greedy algorithm recovers optimal linear scaling with d for the CNOT cost, whereas randomly chosen states have quadratic scaling, even with optimal state ordering.

However, this greedy algorithm can fail to produce a valid protocol, as it does not enforce the condition that $\|\boldsymbol{p}\|_1 = 1$. This condition will be violated for some functions—typically those with many coefficients with approximately equal magnitude. Still, when it works, this algorithm succeeds in producing CNOT-efficient minimum-entanglement protocols, as shown in Fig. 1. Finding more general algorithms that always succeed for this task remains an interesting open problem.

Independent of the algorithm used to minimize the CNOT count of an optimal protocol, the takeaway message is the same: there is an apparent tradeoff between entanglement- and gate-based resources. The disentangling protocol minimizes average entanglement, but not necessarily instantaneous entanglement, and requires only $\mathcal{O}(d)$ intermediate entangling gates; the echoing protocol uses maximal entanglement, but requires only single-particle intermediate gates. Protocols that minimize instantaneous entanglement do so at the cost of more intermediate entangling gates. Depending on the primary sources of error or the physical constraints on any given quantum sensor network implementation, one of these resources might be more important to minimize than the other. In general, determining the optimal CNOT scaling for protocols that minimize instantaneous and/or average entanglement is a crucial open question for future work.

VIII. TIME-INDEPENDENT PROTOCOLS

Another approach to constructing protocols is to use so-called probabilistic protocols. These protocols eschew control and instead exploit the convexity of the quantum Fisher information by staying in one family throughout any given run of the protocol, but by letting this family vary over different runs. Intuitively, each family is sensitive to a different function q_n such that $q = \sum_{n=1}^{\overline{N}} p_n q_n$, where $\overline{N}$ is the number of families from $\mathcal{T}$ used in the protocol, and p_n is the frequency that family n is used. In this way, one can create an estimator for q using those for q_n. In order to generate a Fisher information matrix satisfying Eq. (8) [27,32], the p_n should come from a solution to Eq. (13). These protocols have the advantage of requiring no control, but, unfortunately, suffer worse scaling with d than ours for generic functions when the available resources are comparable.

In particular, to fairly account for resources, we must fix a total time t to perform *all* stages of our protocol. Therefore, when considering a probabilistic protocol that uses multiple families from $\mathcal{T}$, but does not switch coherently between them, we must assign a time t_n to family n such that

$$\sum_{j=1}^{\overline{N}} t_n = t. \tag{27}$$

Note, we have used the fact that no stages of a probabilistic protocol with the families in $\mathcal{T}$ can be performed simultaneously. One could imagine protocols that parallelize the measurement of some q_j that involve disjoint sets of sensors. However, such protocols are necessarily nonoptimal given Lemma 2 in Appendix A, which says that any optimal protocol requires entanglement with the first qubit at all times.

We can bound the maximum of the Fisher information matrix element $\mathcal{F}(\boldsymbol{\theta})_{11}$ obtainable via such a probabilistic protocol as

$$\max \mathcal{F}(\boldsymbol{\theta})_{11} \leqslant \max_{p_n,t_n} \sum_{n=1}^{\overline{N}} p_n t_n^2, \quad \text{subject to} \sum_{n=1}^{\overline{N}} t_n = t, \quad \sum_{n=1}^{\overline{N}} p_n = 1. \tag{28}$$

where we used that $\tau_1^{(n)} = 1$ for all n. The inequality arises due to the fact that the maximization problem on the right-hand side of the inequality does not enforce that $T\boldsymbol{p} = \boldsymbol{\alpha}/\alpha_1$. We could add this as an additional constraint, but it will not be necessary.

To perform the necessary optimization, consider the Lagrangian

$$\mathcal{L} = \sum_{n=1}^{\overline{N}} p_n t_n^2 + \gamma_1 \left(t - \sum_{n=1}^{\overline{N}} t_n \right) + \gamma_2 \left(1 - \sum_{n=1}^{\overline{N}} p_n \right), \tag{29}$$

where γ_1, γ_2 are Lagrange multipliers. Therefore, we obtain the system of equations

$$\begin{aligned} 2p_n t_n - \gamma_1 &= 0 \quad (\forall n), \\ t_n^2 - \gamma_2 &= 0 \quad (\forall n), \\ \sum_{n=1}^{\overline{N}} t_n &= t, \\ \sum_{n=1}^{\overline{N}} p_n &= 1, \end{aligned} \tag{30}$$

which can be solved to yield the solution

$$\max_{p_n,t_n} \sum_{n=1}^{\overline{N}} p_n t_n^2 = \frac{t^2}{\overline{N}^2}, \tag{31}$$

for $p_n = 1/\overline{N}$ and $t_n = t/\overline{N}$ for all n. Therefore,

$$\mathcal{F}(\boldsymbol{\theta})_{1j} \leqslant \frac{t^2}{\overline{N}^2} \quad (\forall j), \tag{32}$$

which clearly fails to achieve the saturability condition for $j = 1$, unless $\overline{N} = 1$, which is only possible for a very small set of functions (generic functions require $\overline{N}$ that scale nontrivially with d). Therefore, provided one considers cases where each q_n must be learned sequentially (which is a requirement for any possibly optimal protocol via Lemma 2), we fail to achieve saturability even up to a d-independent constant for generic functions via time-independent protocols.

Note that we have, for simplicity, again restricted ourselves to the case where $\boldsymbol{\alpha}$ has a single maximal magnitude element. The more general proof follows almost identically, with some notational overhead, when generalizing beyond this condition.

IX. CONCLUSION AND OUTLOOK

We have proven that maximally entangled states are not necessary for the optimal measurement of a linear function with a quantum sensor network unless the function is sufficiently uniformly supported on the unknown parameters. While the uniformly distributed case has been considered extensively in the literature, as it provides the largest possible separation in performance between entangled and separable protocols, there is no *a priori* reason why one should be interested in only these sorts of quantities. Our results demonstrate that while the precision gains to be had are less away from the uniformly distributed regime, the required resources are also less. This result is of particular relevance to the development of near-term quantum sensor networks, where creating large-scale entangled states may not be practical. Furthermore, while algebraic approaches like the one we consider here have been used before to generate bounds for the function estimation problem [27,38], leveraging this approach to derive protocols that achieve these bounds subject to various experimental constraints is a new and widely applicable technique. We emphasize again that these results are also useful in more general settings, such as the measurement of analytic functions, as these measurements reduce to the case studied here [36–38].

To the best of the authors' knowledge, all information-theoretically optimal protocols for the estimation of a single linear function that are currently in the literature are subsumed by the framework that we develop in this work. What protocol one chooses to use will depend heavily on the experimental context; if decoherence is more problematic than the number of entangling gates that one must perform, then minimum-entanglement protocols will be preferred to the conventional protocols. However, if decoherence is mild, but two-qubit gates introduce significant errors, then a protocol such as the echoing protocol presented in Ref. [27] will be preferred. Consequently, the extent to which minimum entanglement protocols are more or less valuable than their more highly entangled counterparts depends on the details of the physical implementation of a quantum sensor network. Either way, the development of a framework to address these questions is, in and of itself, an important contribution of this work.

We also briefly point out one more resource-related constraint of protocols that rely on time-dependent control (whether in the form of $\hat{\sigma}^x$ gates, CNOT gates, or others): these protocols require precise timing of the gate applications. Uncertainty in the timing leads directly to a systematic error in the function being measured. Importantly, however, this timing issue is a limitation of all known optimal protocols for the linear function estimation task (see, e.g., Ref. [27]). We therefore view these limitations as more pertinent to experimental implementation than the theory of resource tradeoffs that we are considering here.

So far, we have not discussed the situation where we are constrained to k-partite entanglement, but k is not sufficient to achieve optimality (for any protocol) via Theorem 1. We propose the following protocol for such a scenario: Let R be a partition of the sensors into independent sets where we do not allow entanglement between sets and allow, at most, k-partite entanglement within each $r \in R$. Let $\boldsymbol{\alpha}^{(r)}$ denote $\boldsymbol{\alpha}$ restricted to r. Pick the optimal R such that the condition of Theorem 1 is satisfied for all r; that is, we ensure that *within* each independent set we obtain the optimal variance for the linear function restricted to that set. The result is a variance

$$\mathcal{M} = \frac{1}{t^2} \sum_{r \in R} \|\boldsymbol{\alpha}^{(r)}\|_\infty^2. \tag{33}$$

The optimal R is a partition of the sensors into contiguous sets (assuming for simplicity that $|\alpha_i| \geqslant |\alpha_j|$ for $i < j$) such that for all $r \in R$, $\sum_{i \in r} |\alpha_i| / \max_{i \in r} |\alpha_i| \leqslant k$, satisfying Theorem 1. We conjecture that this protocol is optimal, and it is clearly so if partitioning the problem into independent sets is optimal. However, one could imagine protocols that use different partitions for some fraction of the time. Intuitively, this should not improve the performance, but we leave analyzing this as an open question.

Finally, no optimal time-independent protocols for arbitrary linear functions exist in the literature. Finding such protocols (or proving their nonexistence) remains an open problem of interest.

ACKNOWLEDGMENTS

We thank M. Foss-Feig, Z. Eldredge, T. Goel, P. Niroula, L. P. García-Pintos, and M. Gullans for helpful discussions. We thank the anonymous referee who pointed out that the CNOT cost of our minimum-entanglement protocols deserved a more detailed analysis. This work supported in part by AFOSR MURI, AFOSR, DARPA SAVaNT ADVENT, NSF PFCQC program, ARO MURI, DoE ASCR Accelerated Research in Quantum Computing program (Award No. DE-SC0020312), NSF QLCI (Award No. OMA-2120757), U.S. Department of Energy Award No. DE-SC0019449, and the DoE ASCR Quantum Testbed Pathfinder program (Award No. DE-SC00119040). Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. This research was supported in part by

the Heising-Simons Foundation, the Simons Foundation, and National Science Foundation Grant No. NSF PHY-1748958. J.B. acknowledges support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship (Award No. DE-SC0019323).

APPENDIX A: A USEFUL LEMMA REGARDING OPTIMAL PROBE STATES

In this Appendix, we prove a useful lemma restricting the structure of the probe state for an optimal protocol.

Lemma 2. Any optimal protocol, independent of the choice of control, requires that $\langle\hat{\mathcal{H}}_1(t)\rangle = 0$, where $\mathcal{H}_1(t)$ is the time-evolved generator of the first parameter and the expectation value is taken with respect to the initial probe state. Furthermore, the probe state must be of the form

$$|\psi\rangle = \frac{|0\rangle|\chi_0\rangle + e^{i\phi}|1\rangle|\chi_1\rangle}{\sqrt{2}}, \tag{A1}$$

for all times $s \in [0, t]$, where $|\chi_0\rangle, |\chi_1\rangle$ are arbitrary states on the $d-1$ remaining sensor qubits plus, potentially, the arbitrary number of ancilla, and ϕ is an arbitrary phase in $\mathbb{R}$—they can be s dependent.

Proof. Consider the expression for the matrix elements of the quantum Fisher information matrix at time t [Eq. (4) of the main text]:

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4\big[\tfrac{1}{2}\langle\{\hat{\mathcal{H}}_i(t), \hat{\mathcal{H}}_j(t)\}\rangle - \langle\hat{\mathcal{H}}_i(t)\rangle\langle\hat{\mathcal{H}}_j(t)\rangle\big], \tag{A2}$$

where the expectation values are taken with respect to the initial probe state $|\psi(0)\rangle$. Using the integral form of $\hat{\mathcal{H}}_j(t)$ [Eq. (5) of the main text], we can write

$$\mathcal{F}(\boldsymbol{\theta})_{11} = 4\mathrm{Var}\big[\hat{\mathcal{H}}_1(t)\big] \tag{A3}$$

$$= 4\left[\int_0^t ds \int_0^t ds' \langle\psi(0)|\hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)\hat{U}^\dagger(s')\hat{g}_1\hat{U}(s')|\psi(0)\rangle\right] - 4\left[\int_0^t ds\langle\psi(0)|\hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)|\psi(0)\rangle\right]^2 \tag{A4}$$

$$= 4\int_0^t ds \int_0^t ds' \mathrm{Cov}_{|\psi(0)\rangle}[\hat{g}_1(s), \hat{g}_1(s')], \tag{A5}$$

where we recall

$$\hat{g}_1(s) := \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s), \tag{A6}$$

and $\hat{g}_1 = \partial\hat{H}/\partial\theta_1$ is the initial generator with respect to the first parameter. Once again, the covariance is with respect to the initial probe state $|\psi(0)\rangle$. We can then upper-bound this as

$$\mathcal{F}(\boldsymbol{\theta})_{11}(t) \leqslant 4\int_0^t ds \int_0^t ds' \sqrt{\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s)]\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s')]} \tag{A7}$$

$$= 4\left[\int_0^t ds \sqrt{\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s)]}\right]^2 \tag{A8}$$

$$\leqslant \left[\int_0^t ds \|\hat{g}_1\|_s\right]^2 \tag{A9}$$

$$= t^2\|\hat{g}_1\|_s^2 \tag{A10}$$

$$= t^2, \tag{A11}$$

where the first inequality bounds the covariance as the square root of the product of the variances, the second inequality bounds the standard deviation of an operator by half the seminorm [15], and the final equality uses the fact that $\hat{g}_1 = \hat{\sigma}_1^z/2$ has seminorm 1 [58].

Via Eq. (8) of the main text (rigorously derived in Appendix F) we know that an optimal protocol must have $\mathcal{F}_{11}(\boldsymbol{\theta})(t) = t^2$. Therefore, an optimal protocol must saturate the inequalities in Eq. (A7) and Eq. (A9). Equation (A9) is saturated when $\mathrm{Var}[\hat{g}_1(s)] = \|\hat{g}_1(s)\|_s = \|\hat{g}_1\|_s$ for all s. This holds if and only if $|\psi(0)\rangle = \frac{1}{\sqrt{2}}(|\lambda_{\min}\rangle + e^{i\phi}|\lambda_{\max}\rangle)$, where $|\lambda_{\min}\rangle$ and $|\lambda_{\max}\rangle$ are the eigenstates corresponding to the minimum and maximum eigenvalues of $\hat{g}_1(s)$ for all $s \in [0, t]$ and ϕ is an arbitrary phase. Given this condition, $\hat{g}_1(s)$ and $\hat{g}_1(s')$ act identically on the state $|\psi(0)\rangle$ and consequently are fully correlated when one considers the covariance of these operators with respect to the state. The Cauchy-Schwarz inequality in Eq. (A7) is immediately saturated as well.

Importantly, under this condition on the probe state, any operator in the one-parameter family $\hat{g}_1(s) = \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)$ acts identically on $|\psi(0)\rangle$ (the unitary does not change the eigenvalues, and the eigenstates are shared by all $\hat{g}_1(s)$, as argued above). Thus, one can freely substitute any operator in the one-parameter family $\hat{g}_1(s) = \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)$ for another. Therefore, for such an optimal probe state,

$$\langle\mathcal{H}_1(t)\rangle = -\int_0^t ds\langle\psi(0)|\hat{g}_1(s)|\psi(0)\rangle = t\langle\hat{g}_1\rangle = 0 \tag{A12}$$

because $\hat{g}_1 \propto \hat{\sigma}_1^z$ and, consequently, by the argument that we can replace $\hat{g}_1$ by $\hat{g}_1(s)$ when acting on the probe state,

$$\langle\psi(s)|\hat{g}_1|\psi(s)\rangle = 0 \quad (\forall s). \tag{A13}$$

The statement of the lemma immediately follows. ∎

Note that Lemma 2 holds for any optimal protocol, not just those using our catlike states. However, it also justifies our choice of probe states and why we specifically set $\tau_1 = 1$ for all $\boldsymbol{\tau}$ (i.e., to maintain an equal superposition between $|0\rangle$ and $|1\rangle$ on the first qubit).

APPENDIX B: PROOF OF THE OPTIMALITY OF CAT-STATE PROTOCOLS

In this Appendix, we will rigorously prove the optimality of the time-dependent protocols considered in the main text.

In particular, we show that the Fisher information matrix condition for saturability in Eq. (8) of the main text is satisfied by solutions to Eq. (13) of the main text when we consider protocols that use $\hat{\sigma}^x$ and CNOT controls to switch between families of catlike states in $\mathcal{T}$. That is, we show the following mapping between saturability conditions:

$$T\boldsymbol{p} = \frac{\boldsymbol{\alpha}}{\alpha_1} \Rightarrow \mathcal{F}(\boldsymbol{\theta})_{1j} = \frac{\boldsymbol{\alpha}}{\alpha_1} t^2, \tag{B1}$$

where we recall that we have assumed that $|\alpha_1| = \|\boldsymbol{\alpha}\|_\infty > |\alpha_j|$ for all $j > 1$ (in Appendix F, we will generalize beyond the assumption of a single maximum magnitude α_j at the cost of some notational inconvenience).

Using Lemma 2, we can show that for *any* optimal protocol (i.e., not just those using our catlike states)

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = 2\langle\{\hat{\mathcal{H}}_1, \hat{\mathcal{H}}_j\}\rangle \tag{B2}$$

$$= 2\int_0^t ds \int_0^t ds' \langle\psi(0)|\{\hat{g}_1(s), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{B3}$$

$$= 2\int_0^t ds \int_0^t ds' \langle\psi(0)|\{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{B4}$$

$$= 2t\int_0^t ds' \langle\psi(0)|\{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{B5}$$

$$= 2t\int_0^t ds' \langle\psi(0)|\{\hat{g}_1(s'), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{B6}$$

$$= 4t\int_0^t ds' \langle\psi(s')|\hat{g}_1\hat{g}_j|\psi(s')\rangle \tag{B7}$$

$$= t\int_0^t ds' \langle\psi(s')\big|\hat{\sigma}_1^z\hat{\sigma}_j^z\big|\psi(s')\rangle. \tag{B8}$$

The third and fifth equalities come from the argument in the proof of Lemma 2 that we may replace $\hat{g}_1(s)$ with $\hat{g}_1$ (and vice versa) when acting on optimal probe states. The penultimate equality is just a consequence of the commutativity of the initial generators.

We now apply these general results to our specific protocols. Saturating the initial Fisher information conditions in Eq. (B1) implies that we must show

$$\int_0^t ds' \langle\psi(s')|\hat{\sigma}_1^z\hat{\sigma}_j^z|\psi(s')\rangle = \frac{\alpha_j}{\alpha_1} t. \tag{B9}$$

Let the gates in our protocols be labeled as $\hat{G}_i$ where $\hat{G}_i$ is either a CNOT or $\hat{\sigma}^x$ gate. The gate $\hat{G}_i$ is applied at a time $s = t_i^*$. Then, for $s \in (t_k^*, t_{k+1}^*)$, we can write the time-dependent state as

$$|\psi(s)\rangle = |\psi(\boldsymbol{\tau}^{(k)}; \varphi)\rangle \equiv \prod_{i=0}^{k} \hat{G}_i |\psi(\boldsymbol{\tau}^{(0)}; \varphi)\rangle, \tag{B10}$$

where $|\psi(\boldsymbol{\tau}^{(0)}; 0)\rangle$ is the initial state of the protocol, φ is the relative phase between the two branches of the state that has accumulated up to time s, and, therefore, $|\psi(\boldsymbol{\tau}^{(k)}; \varphi)\rangle$ is the state produced after applying the first k gates. Because our protocols explicitly use only $\hat{\sigma}^x$ and CNOT gates to move between families in $\mathcal{T}$, we have that $|\psi(\boldsymbol{\tau}^{(k)}; \varphi)\rangle = (|0\rangle|\chi_0^{(k)}\rangle + e^{i\varphi}|1\rangle|\chi_1^{(k)}\rangle)/\sqrt{2}$, and

$$\int_0^t ds' \langle\psi(s')|\hat{\sigma}_1^z\hat{\sigma}_j^z|\psi(s')\rangle = \sum_{i=0}^{n} (t_{i+1}^* - t_i^*)\tau_j^{(i)}, \tag{B11}$$

where we implicitly define $t_0^* = 0$ and $t_{n+1}^* = t$ as the initial and final times of the protocol and $|\chi_0^{(k)}\rangle$ and $|\chi_1^{(k)}\rangle$ are some states defined on the Hilbert space which excludes the first qubit sensor. The time $t_{i+1}^* - t_i^*$ corresponds to the time we are in the probe family $|\psi(\boldsymbol{\tau}^{(i)}; \varphi)\rangle$, which in our protocols is $p_i t$. Thus, to satisfy the Fisher information conditions, we need

$$\sum_i p_i \tau_j^{(i)} = \frac{\alpha_j}{\alpha_1} \Rightarrow (T\boldsymbol{p})_j = \frac{\alpha_j}{\alpha_1}. \tag{B12}$$

This formally proves optimality of our time-dependent protocols that satisfy $T\boldsymbol{p} = \boldsymbol{\alpha}/\alpha_1$.

APPENDIX C: REVIEW OF ROBUST PHASE ESTIMATION

In this Appendix, we review, for completeness, the phase estimation protocols of Refs. [51–53] described in the main text as a method to extract the quantity of interest, q, from the state

$$1/\sqrt{2}(|0\rangle + e^{iqt/\alpha_1}|1\rangle)(|0\cdots 0\rangle), \tag{C1}$$

which is the final state obtained from our family of optimal protocols.

Again, when we refer to our protocols as optimal, we mean this in the sense that our protocols achieve the conditions on the quantum Fisher information matrix that allow the maximum possible quantum Fisher information with respect to the parameter q to be obtained. However, to completely specify the procedure by which one obtains the quantity q, an explicit phase estimation protocol is needed. As explained in the main text, such a task is complicated by the fact that for large times and/or small $\alpha_1 = \|\boldsymbol{\alpha}\|_\infty$, it is unclear what 2π interval the relative phase between the branches of Eq. (C1) is in [48,49]. The phase estimation protocols of Refs. [51–53] demonstrate how to optimize resources to deal with this issue, while still saturating the single-shot bound in Eq. (2) of the main text up to a small d- and t-independent constant. In particular, such protocols allow us to reach a mean-square error of

$$\mathcal{M} = \frac{c^2\|\boldsymbol{\alpha}\|_\infty^2}{t^2}, \tag{C2}$$

for some small (explicitly known) constant c. Reference [50] proves that this constant factor c^2 in Eq. (2) can be reduced to, at best, π^2.

While reviewing such phase estimation protocols, we follow the presentation of Ref. [53], which corrects a few minor errors in Ref. [51], as noted in the corresponding erratum [52]. We refer the reader to Ref. [53] for further details. Conveniently, by putting the final state into the form of Eq. (C1), we have reduced this problem completely to the single-qubit, multipass version of the problem described in that reference. Consequently, everything follows practically identically to their presentation.

Consider dividing the total time t, which is the relevant resource in our problem, into K stages where we evolve for a

time $M_j\delta t$ in the jth stage (δt is some small basic unit of time and $M_j \in \mathbb{N}$). We assume that we have (d,t)-independent, prior knowledge of q such that we can set δt to satisfy

$$\frac{\delta t q}{\|\boldsymbol{\alpha}\|_\infty} \in [0, 2\pi). \tag{C3}$$

In the jth stage, using one of our protocols for a time $M_j\delta t$, we prepare $2\nu_j$ independent copies of the state

$$|\psi_j\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{iqM_j\delta t/\|\boldsymbol{\alpha}\|_\infty}|1\rangle)|0\cdots 0\rangle. \tag{C4}$$

From now on we will drop the $d-1$ qubit sensors in the state $|0\cdots 0\rangle$, as they are irrelevant; however, it is worth noting that it is not necessary to put the state in this form before performing measurements. We do so to make the comparison to Ref. [53] particularly transparent. We then perform a single-qubit measurement on the first qubit sensor of each of these state copies, yielding $2\nu_j$ measurement outcomes, which we can use to estimate q. The total time of this K stage protocol is consequently given by

$$t = 2\sum_{j=1}^{K} \nu_j M_j \delta t. \tag{C5}$$

Given this setup, we choose single-qubit measurements and optimize the choice of ν_j, M_j per stage so that we can learn q bit by bit, stage by stage, in such a way that optimal scaling in d, t is still obtained [Eq. (C2)]. In particular, consider making two measurements, each ν_j times per stage (thus explaining the factor of 2 we introduced earlier): (i) a $\hat{\sigma}^x$ measurement and (ii) a $\hat{\sigma}^y$ measurement. These measurements each give us outcomes that are Bernoulli variables (i.e., with values $\in \{0, 1\}$) with outcome probabilities

$$\begin{aligned} p^{(x)}(0) &= \frac{1+\cos(M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty)}{2}, \\ p^{(x)}(1) &= 1 - p^{(x)}(0), \\ p^{(y)}(0) &= \frac{1+\sin(M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty)}{2}, \\ p^{(y)}(1) &= 1 - p^{(y)}(0), \end{aligned} \tag{C6}$$

where the first two probabilities are for the $\hat{\sigma}^x$ measurement and the latter two are for the $\hat{\sigma}^y$ measurement. Using both of these measurements allows us to resolve the twofold degeneracy in the phase $qM_j\delta t/\|\boldsymbol{\alpha}\|_\infty$ within a given $[0, 2\pi)$ interval that would arise from, e.g., a $\hat{\sigma}^x$ measurement alone. The observed probabilities of obtaining 0 for the $\hat{\sigma}^x$ and $\hat{\sigma}^y$ are independent random variables that converge in probability to their associated expectation values for $\nu_j \to \infty$. Let these observed probabilities be labeled $f_0^{(x)}$ and $f_0^{(y)}$, respectively. These measurements are nonadaptive, which makes this particular phase estimation protocol especially appealing.

At each stage, we extract an estimator $\tilde{\phi}$ of $\phi := M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty$ as

$$\tilde{\phi} := \mathrm{atan2}\big(2f_0^{(y)} - 1, 2f_0^{(x)} - 1\big) \in [0, 2\pi), \tag{C7}$$

where atan2 is the two-argument arctangent with range $[0, 2\pi)$. In the limit $\nu_j \to \infty$, this estimator indeed converges to ϕ, but the "magic" of this phase estimation scheme lies in the correct reprocessing of data stage-by-stage so that ν_j can be kept (d,t) independent. Reference [53] demonstrates rigorously that picking $M_j = 2^{j-1}$ for $j \in \{1, \ldots, K\}$ and optimizing over ν_j one can, at each stage, estimate $q/\|\boldsymbol{\alpha}\|_\infty$ with a confidence interval of size $2\pi/(3\times 2^{j-1})$ so that in each stage we learn another bit of this quantity. The results of this optimization are ν_j that decrease linearly with the step j so that as the time spent in a stage grows, the statistics we employ shrink. Importantly, it so happens that we can scale $K \to \infty$ (i.e., take an asymptotic-in-t limit) while maintaining ν_K constant. The net result is a mean-square error given by Eq. (C2) with $c = 24.26\pi$, which is a factor of 24.26 greater than the theoretical optimal value [50], but with the convenient feature that the protocol uses nonadaptive measurements. We refer the interested reader to Ref. [53] for detailed derivation of the results sketched here.

It is also worth noting that other protocols are possible. For instance, in Ref. [45], a similar two-step method is described for the estimation of global parameters (i.e., where the parameter is not restricted to a local neighborhood of parameter space). This protocol provides an explicit method to use some (ultimately negligible) fraction of the sensing time available to narrow down the location of the parameter q in parameter space, followed by an optimal local estimation. We emphasize that the explicit estimation scheme we propose (i.e., the one in Refs. [51–53]) does not require adaptive measurements, which is one of its key advantages.

APPENDIX D: FULL PROOF OF THE MAIN THEOREM

In this Appendix, we expand on the proof sketch of Theorem 1 in the main text to fully prove the result. For reference, this theorem is restated here.

Theorem 1. Let $q(\boldsymbol{\theta}) = \boldsymbol{\alpha}\cdot\boldsymbol{\theta}$. Without loss of generality, let $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $k \in \mathbb{Z}^+$ so that

$$k - 1 < \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} \leqslant k. \tag{D1}$$

An optimal protocol to estimate $q(\boldsymbol{\theta})$, where the parameters $\boldsymbol{\theta}$ are encoded into the probe state via unitary evolution under the Hamiltonian in Eq. (1) of the main text, requires at least, but no more than, k-partite entanglement.

Proof. We divide our proof into two parts. First, using k-partite-entangled states from the set of catlike states considered in the main text, we show the existence of an optimal protocol, subject to the upper bound of Eq. (D1). Second, we show that there exists no optimal protocol using at most $(k-1)$-partite entanglement, proving the lower bound of Eq. (D1).

Part 1. Define $T^{(k)}$ to be the submatrix of T with all columns n such that $\sum_m |T_{mn}| > k$ are eliminated, which enforces that any protocol derived from $T^{(k)}$ uses only states that are at most k-partite entangled. Define system $A(k)$ as

$$T^{(k)}\boldsymbol{p}^{(k)} = \boldsymbol{\alpha}/\alpha_1, \tag{D2}$$

$$\boldsymbol{p}^{(k)} \geqslant 0. \tag{D3}$$

Let $\boldsymbol{\alpha}' = \boldsymbol{\alpha}/\alpha_1$ and define system $B(k)$ as

$$(T^{(k)})^\top \boldsymbol{y} \geqslant 0, \tag{D4}$$

$$\langle \boldsymbol{\alpha}', \boldsymbol{y} \rangle < 0. \tag{D5}$$

By the Farkas-Minkowski lemma [54,55], system $A(k)$ has a solution if and only if system $B(k)$ does not. In particular, this lemma, which, geometrically, is an application of the hyperplane separation theorem [59], is as follows:

Lemma 3 (Farkas-Minkowski). Consider the system

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{D6}$$

$$\boldsymbol{x} \geqslant 0, \tag{D7}$$

with $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{b} \in \mathbb{R}^m$. The above system has a solution if and only if there is no solution $\boldsymbol{y}$ to

$$A^\top \boldsymbol{y} \geqslant 0, \tag{D8}$$

$$\langle \boldsymbol{b}, \boldsymbol{y} \rangle < 0. \tag{D9}$$

Therefore, to prove the result it is sufficient to show that system $B(k)$ does not have a solution if $\sum_{j>1} |\alpha'_j| \leqslant k-1$, where we used that $\alpha'_1 = 1$. We assume that a solution $\boldsymbol{y}$ exists and will arrive at a contradiction. Without loss of generality, we assume that $|y_j| \geqslant |y_{j+1}|$ for all $1 < j < d$. Equation (D5) implies $\sum_{j>1} \alpha'_j y_j < -y_1$. $(T^{(k)})^\top$ has a row n^* given by $\boldsymbol{\tau}^{(n^*)} = (1, 0, \ldots, 0)$, so by Eq. (D4) any solution $\boldsymbol{y}$ to system B has $y_1 \geqslant 0$. Therefore, $|\sum_{j>1} \alpha'_j y_j| > y_1$, which, by the triangle inequality, implies

$$\sum_{j>1} |\alpha'_j||y_j| > y_1. \tag{D10}$$

Because $|\alpha'_j| \leqslant 1$ for all j, because $\sum_{j>1} |\alpha'_j| \leqslant k-1$, and because $|y_j|$ for $j > 1$ are ordered in descending order, the largest the left-hand side of Eq. (D10) can be is $\sum_{j=2}^{k} |y_j|$, leading to

$$\sum_{j=2}^{k} |y_j| > y_1. \tag{D11}$$

This directly contradicts Eq. (D4) for the column of $T^{(k)}$ given by $\boldsymbol{\tau} = (1, -\mathrm{sgn}(y_2), \ldots, -\mathrm{sgn}(y_k), 0, 0, \ldots)$.

Part 2. Using Eq. (B8), we have that, for any optimal protocol,

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = t \int_0^t ds' \langle \psi(s')|\hat{\sigma}_1^z \hat{\sigma}_j^z |\psi(s')\rangle, \tag{D12}$$

where we recall that $|\psi(s)\rangle = U(s)|\psi(0)\rangle$. Because $\langle \psi(s')|\hat{\sigma}_1^z|\psi(s')\rangle = 0$ for all s' [see Eq. (A13)], the integrand is nonzero if and only if $|\psi(s')\rangle$ is such that the first qubit is entangled with the jth. Define the indicator variable

$$E_j(s') = \begin{cases} 1 & |\psi(s)\rangle \text{ entangles qubit } j \text{ and } 1 \\ 0 & \text{else,} \end{cases} \tag{D13}$$

for all j, including any possible ancilla qubits. Here, we define $E_1 = 1$ even though the first qubit is not "entangled" with itself. Further define

$$E(s') = \sum_j E_j(s') \leqslant (k-1), \tag{D14}$$

where $E(s')$ is the total number of sensor qubits entangled with the first qubit at time s' and the upper bound comes from our assumption on the partiteness of our probe states. We then have that

$$\mathcal{F}(\boldsymbol{\theta})_{1j} \leqslant t \int_0^t ds' E_j(s'). \tag{D15}$$

Furthermore, for any optimal protocol using at most $(k-1)$-partite entanglement, we require that

$$\begin{aligned} \sum_j \left| \frac{\alpha_j}{\alpha_1} t^2 \right| &= \sum_j |\mathcal{F}(\boldsymbol{\theta})_{j1}| \leqslant t \sum_j \int_0^t ds' E_j(s') \\ &= t \int_0^t ds' \sum_j E_j(s) \leqslant t \int_0^t ds'(k-1) \\ &= (k-1)t^2. \end{aligned} \tag{D16}$$

We now have a contradiction, however, as the theorem statement assumed that

$$\sum_j \left| \frac{\alpha_j}{\alpha_1} t^2 \right| = \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} t^2 > (k-1)t^2. \tag{D17}$$

This concludes the proof that $(k-1)$-partite entanglement in any form (i.e., not just from catlike probe states) is insufficient to generate an optimal protocol. ■

We also observe that the lower bound on the size of the least-entangled state used in an optimal protocol is really, at its core, a lower bound on the *average* entanglement required to saturate the conditions on the quantum Fisher information matrix. Here, average entanglement refers to weighting the size of the entangled state by the proportion of time it is used in the protocol. This lower bound is simply $\|\boldsymbol{\alpha}\|_1/\boldsymbol{\alpha}_\infty$. The lower bound on the size of the most-entangled state, or the bound on *instantaneous* entanglement, comes from ensuring that this lower bound on average entanglement is achievable (that is, if the instantaneous entanglement is too small at each stage, then the average entanglement required cannot be reached).

APPENDIX E: MINIMUM-ENTANGLEMENT NON-ECHOED PROTOCOLS

In this Appendix, we prove that there exist protocols that minimize both instantaneous and average entanglement. We recall from Sec. VI the definition of the non-echoed protocols that minimize average entanglement.

Definition 2 (Non-echoed protocols). Consider some $\boldsymbol{\alpha} \in \mathbb{R}^d$ encoding a linear function of interest. Let T be the matrix which describes our families of catlike probe states, and let $\boldsymbol{p}$ specify a valid protocol such that $\boldsymbol{p} \geqslant 0$ and $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. We say that the protocol defined by $\boldsymbol{p}$ is "non-echoed" if, $\forall i$ such that p_i is strictly greater than 0, $\mathrm{sgn}(T_{ij}) \in \{0, \mathrm{sgn}(\alpha_j)\}$.

We now prove Theorem 3 from the main text, which we again repeat for simplicity.

Theorem 3. For any function encoding $\boldsymbol{\alpha}$, there exists a non-echoed optimal protocol with minimum instantaneous entanglement.

Proof. We proceed with a relatively simple tweak of the proof of the main theorem. As in that theorem, we assume without loss of generality that $\alpha_1 = \|\boldsymbol{\alpha}\|_\infty = 1$. Also assume,

for computational simplicity, that $\alpha_{i>1} < 1$ (i.e., there is only a single maximal-magnitude element of $\boldsymbol{\alpha}$) and that $\alpha_i > 0\ \forall i$. These latter assumptions can easily be lifted, as we describe at the end of the proof.

We will again use the Farkas-Minkowski lemma [54,55] to show that no vector $\boldsymbol{y}$ exists such that

$$(T_+^{(k)})^\top \boldsymbol{y} \geqslant 0, \tag{E1}$$

$$\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle < 0, \tag{E2}$$

proving the existence of a non-echoed protocol. Here, $T_+^{(k)}$ is T restricted to non-echoed vectors [i.e., $(T_+^{(k)})_{ij} \in \{0, 1\}$] with weight at most k, where $k = \lceil \|\boldsymbol{\alpha}\|_1 \rceil$. Assume a solution $\boldsymbol{y}$ exists. Noting that $(T_+^{(k)})^\top$ has a row given by $(1, 0, \ldots, 0)$, it must be that $y_1 \geqslant 0$. Furthermore, for $\boldsymbol{y}$ to be a valid solution, we must have

$$\begin{aligned}\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle &= \alpha_1 y_1 + \sum_{i|i\neq 1, y_i \geqslant 0} \alpha_i y_i + \sum_{i|y_i<0} \alpha_i y_i \\ &= y_1 + \sum_{i|i\neq 1, y_i \geqslant 0} \alpha_i y_i + \sum_{i|y_i<0} \alpha_i y_i \leqslant 0. \end{aligned} \tag{E3}$$

We proceed with two cases. Suppose that at most $k-1$ elements of $\boldsymbol{y}$ are negative. Consider the row of $(T_+^{(k)})^\top$ that has a 1 in the first index and exactly on the indices where $y_i < 0$ (which exists because we have sufficiently restricted the number of negative elements of $\boldsymbol{y}$). Then $(T_+^{(k)})^\top \boldsymbol{y} \geqslant 0$ implies that

$$y_1 + \sum_{i|y_i \leqslant 0} y_i \geqslant 0. \tag{E4}$$

But because $\alpha_i < 1$, this immediately implies that

$$y_1 + \sum_{i|y_i \leqslant 0} \alpha_i y_i \geqslant 0, \tag{E5}$$

which means that Eq. (E3) cannot be true, yielding a contradiction.

Now suppose that there are at least k elements of $\boldsymbol{y}$ that are negative. Let S be the set of indices corresponding to the $k-1$ largest, in magnitude, y_i. Then the row of $(T_+^{(k)})^\top$ with a 1 in the first index and precisely on the indices in S leads to the condition that

$$y_1 + \sum_{i\in S} y_i \geqslant 0. \tag{E6}$$

However, given the constraint that $\alpha_{i>1} < 1$, we find that

$$y_1 + \sum_{i|i\neq 1, y_i \geqslant 0} \alpha_i y_i + \sum_{i|y_i<0} \alpha_i y_i \geqslant y_1 + \sum_{i\in S} y_i \geqslant 0, \tag{E7}$$

which is again a contradiction.

We briefly comment on how to lift the two assumptions we mentioned earlier. First, in the case where there exist multiple maximal elements, the same argument that generalizes the main theorem will also generalize this argument—see Appendix F. Second, if we allow $\alpha_i < 0$, it is simple to see that a protocol still exists; simply replace $(T_+^{(k)})_{ij} = 1$ with $\mathrm{sgn}(\alpha_i)$ (and leave 0s untouched). ■

Thus, Lemma 1 and Theorem 3 prove there exist protocols that can minimize both instantaneous entanglement (i.e., the maximum size of a catlike state used in the protocol) and the average entanglement over the course of the entire protocol.

APPENDIX F: RELAXING THE ASSUMPTION ON A SINGLE MAXIMUM ELEMENT

In this Appendix, we will generalize beyond the assumption in the main text that $|\alpha_1| > |\alpha_j|$ for all $j > 1$. Conceptually, nothing is changed by relaxing the assumption, but the algebra becomes somewhat more tedious. In the process, we rigorously derive Eq. (2) and Eq. (8) of the main text.

1. Generalizing Equation (8) of the main text

We start with specifically generalizing Eq. (8). To begin, define

$$L := \{i \,|\, |\alpha_i| = |\alpha_1|\}. \tag{F1}$$

The assumption $|\alpha_1| > |\alpha_j|$ for all $j > 1$, stated in the main text, is equivalent to assuming $|L| = 1$. For arbitrary size L, we have the following set of conditions for the single-parameter bound on $q(\boldsymbol{\theta})$ to be saturable [Eqs. (6) and (7) of the main text]:

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{F2}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1). \tag{F3}$$

Recall that $\mathcal{F}(\boldsymbol{q}) = J^\top \mathcal{F}(\boldsymbol{\theta}) J$, where J is the Jacobian for the basis transformation from $\boldsymbol{\theta}$ to $\boldsymbol{q}$, $q_1 = q$ is the linear function we wish to measure, and the other q_j are some other degrees of freedom we fix. We will show that Eqs. (F2) and (F3) are satisfied if and only if

$$\sum_{i\in L} \frac{\mathrm{sgn}(\alpha_1)}{\mathrm{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji} \lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \tag{F4}$$

where $\lambda_i \geqslant 0$ such that $\sum_i \lambda_i = 1$. If $|L| = 1$, this reduces to Eq. (8) of the main text.

It will be important to briefly recount how we obtain the single-parameter bound we are trying to saturate [27,38]. In particular, referring to Eq. (3) of the main text, we seek a choice of basis that minimizes $\|\hat{g}_q\|_s^2$, which will yield the tightest possible bound on $\mathcal{M}$, the mean-square error of q. Let us formally define our basis for $\mathbb{R}^d$ as $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(d)}\}$, where $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}$. We then have that J^{-1} has rows given by these vectors. Let $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(d)}\}$ be the basis dual to this one. That is, these vectors form the columns of J and satisfy $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$. We can then write

$$\boldsymbol{\theta}^\top = (JJ^{-1}\boldsymbol{\theta})^\top = (J^{-1}\boldsymbol{\theta})^\top J^\top, \tag{F5}$$

which allows us to rewrite our Hamiltonian in the convenient form

$$\hat{H} = \frac{1}{2}\boldsymbol{\theta}^\top \hat{\boldsymbol{\sigma}} + \hat{H}_c(s) = \frac{1}{2}\sum_{i=1}^{d} (\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\theta}) \boldsymbol{\beta}^{(i)} \cdot \hat{\boldsymbol{\sigma}} + \hat{H}_c(s), \tag{F6}$$

where $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1^z, \ldots, \hat{\sigma}_d^z)^\top$. Then

$$\hat{g}_q(0) = \frac{\partial \hat{H}}{\partial q} = \frac{\partial \hat{H}}{\partial(\boldsymbol{\alpha}^{(1)} \cdot \boldsymbol{\theta})} = \frac{\boldsymbol{\beta} \cdot \hat{\boldsymbol{\sigma}}}{2}, \tag{F7}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$. Because the seminorm is time independent (see Ref. [15]), we immediately have that

$$\left\| \hat{g}_q \right\|_s = \|\boldsymbol{\beta}\|_1, \tag{F8}$$

and our tightest bound is given by

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1,$$
$$\text{such that } \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \tag{F9}$$

Note that

$$1 = \sum_i \alpha_i \beta_i \leqslant \sum_i |\alpha_i||\beta_i| \leqslant |\alpha_1| \sum_i |\beta_i| = |\alpha_1| \|\boldsymbol{\beta}\|_1. \tag{F10}$$

The first inequality is tight if either $\text{sgn}(\beta_i) = \text{sgn}(\alpha_i)$ or $\beta_i = 0$ for all i. The second is slightly more complicated to saturate. Recall $L = \{i \,|\, |\alpha_i| = |\alpha_1|\}$. Then the second inequality is tight if and only if

$$\beta_i = 0 \text{ for } i \notin L, \tag{F11}$$

$$\sum_{i \in L} |\beta_i| = \frac{1}{|\alpha_1|}. \tag{F12}$$

Any solution $\boldsymbol{\beta}$ specifies the first column of the Jacobian J and allows us to rewrite the conditions in Eqs. (F2) and (F3) as

$$\mathcal{F}(\boldsymbol{q})_{11} = \boldsymbol{\beta}^\top \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = \frac{t^2}{\alpha_1^2}, \tag{F13}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = (\boldsymbol{\beta}^{(i)})^\top \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = 0 \quad (\forall\, i \neq 1). \tag{F14}$$

As $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$, Eq. (F14) immediately implies that the vector $\mathcal{F}(\boldsymbol{\theta})\boldsymbol{\beta}$ must be proportional to $\boldsymbol{\alpha}$ and Eq. (F13) specifies the constant of proportionality. In particular, we require

$$\mathcal{F}(\boldsymbol{\theta})\boldsymbol{\beta} = \frac{t^2}{\alpha_1^2} \boldsymbol{\alpha}. \tag{F15}$$

Invoking Eqs. (F11) and (F12) and the condition that $\text{sgn}(\beta_i) = \text{sgn}(\alpha_i)$ for $\beta_i \neq 0$, we write $\beta_i = \lambda_i \text{sgn}(\alpha_i)/|\alpha_1|$, where $\lambda_i \geqslant 0$ for $i \in L$ and $\lambda_i = 0$ for $i \notin L$ such that $\sum_i \lambda_i = 1$. The individual components of Eq. (F15) imply

$$\sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ij} \text{sgn}(\alpha_i) \lambda_i = \sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ji} \text{sgn}(\alpha_i) \lambda_i = \frac{t^2}{|\alpha_1|} \alpha_j, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geqslant 0, \tag{F16}$$

which, using $|\alpha_1| = \text{sgn}(\alpha_1)\alpha_1$ and that $\text{sgn}(\alpha_1)\text{sgn}(\alpha_i) = \text{sgn}(\alpha_1)/\text{sgn}(\alpha_i)$ for $i \in L$, yields

$$\sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ij} \lambda_i = \sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji} \lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geqslant 0, \tag{F17}$$

which reduces to Eq. (8) of the main text, when $|L| = 1$, as desired.

2. Generalizing the derivation of Equation (13) of the main text

At this point, we can generalize the derivation of Eq. (13) of the main text to this setting of more than one maximum element of $\boldsymbol{\alpha}$. In particular, Lemma 2 can be immediately extended to the following:

Lemma 4. Any optimal protocol, independent of the choice of control, requires that $\langle \hat{\mathcal{H}}_j(t) \rangle = 0$ for all $j \in L$ and that the probe state be of the form

$$|\psi\rangle = \frac{(\bigotimes_{j \in L} |b_j\rangle)|\chi_0\rangle + e^{i\phi} (\bigotimes_{j \in L} |b_j + 1\rangle)|\chi_1\rangle}{\sqrt{2}}, \tag{F18}$$

for all times $s \in [0, t]$, where

$$b_j = \begin{cases} 0 & \text{if } \text{sgn}(\alpha_j) = 1 \\ 1 & \text{if } \text{sgn}(\alpha_j) = -1, \end{cases} \tag{F19}$$

and $\phi, |\chi_0\rangle, |\chi_1\rangle$ can be arbitrary and s dependent. The addition inside the second ket of Eq. (F18) is mod 2.

Proof. We have the following two facts: (1) $\sum_{i \in L} \lambda_i (\text{sgn}(\alpha_j)/\text{sgn}(\alpha_i)) \mathcal{F}(\boldsymbol{\theta})_{ij} = t^2$ for all $j \in L$ [by Eq. (F17)] and (2) $|\mathcal{F}(\boldsymbol{\theta})_{ij}| \leqslant \mathcal{F}(\boldsymbol{\theta})_{jj}$ for all i (by the fact that the Fisher information matrix is positive semidefinite). These facts imply that an optimal protocol must have $\mathcal{F}(\boldsymbol{\theta})_{jj} = t^2$ for all $j \in L$. The fact that $\langle \hat{\mathcal{H}}_j(t) \rangle = 0$ for all $j \in L$ and the fact that all sensors in L must be in a catlike state over computational basis states follows immediately via an identical calculation to the proof of Lemma 2 for each $j \in L$. From Eq. (B8) it follows directly that these catlike states over the qubit sensors in L must take the form in the theorem statement in order to achieve the correct sign on the components of $\mathcal{F}(\boldsymbol{\theta})$. ■

Using Lemma 4, it is clear that we should restrict the set $\mathcal{T}$ of states such that $\tau_j^{(n)} = \text{sgn}(\alpha_j)/\text{sgn}(\alpha_1)$ for all $j \in L$ and all $\boldsymbol{\tau}^{(n)}$. This is the generalization of the fact that, when $|L| = 1$, we require $\tau_1^{(n)} = 1$ for all $\boldsymbol{\tau}^{(n)}$.

In addition, given the required form of the optimal states, it is easy to generalize Eq. (B9) to the condition that

$$\sum_{i \in L} \left[\lambda_i \int_0^t ds' \langle \psi(s') | \hat{\sigma}_i^z \hat{\sigma}_j^z | \psi(s') \rangle \right] = \frac{\alpha_j}{\alpha_1} t, \tag{F20}$$

which implies that, for protocols switching between states in the modified $\mathcal{T}$,

$$\sum_{i \in L} \left[\lambda_i \sum_{l=0}^{n} (t_{l+1}^* - t_l^*) \tau_j^{(l)} \right] = \frac{\alpha_j}{\alpha_1} t, \tag{F21}$$

where we assume that we switch to the state labeled by $\boldsymbol{\tau}^{(l)}$ at time t_l^*. As before, in our protocols $t_{l+1}^* - t_l^* = p_l t$. In addition, $\sum_i \lambda_i = 1$. So an optimal protocol requires

$$t \sum_{l=0}^{n} p_l \tau_j^{(l)} = \frac{\alpha_j}{\alpha_1} t \Rightarrow T\boldsymbol{p} = \boldsymbol{\alpha}, \tag{F22}$$

recovering Eq. (13) of the main text for general L, with the addition that we fix $T_{jn} = \tau_j^{(n)} = \text{sgn}(\alpha_j)/\text{sgn}(\alpha_1)$ for all $j \in L$ and all n.

3. Generalizing the proof of Theorem 1 of the main text

Recall, we divided the proof into two parts. First, we showed the existence of an optimal protocol using

k-partite-entangled catlike states, subject to the upper bound of the theorem statement. Second, we showed that, subject to the lower bound of the theorem statement, there exists no optimal protocol using only $(k-1)$-partite entanglement.

Let us begin by addressing how the first part changes upon relaxing the assumption that $|\alpha_1| > |\alpha_j|$ for all $j > 1$. Note that, given our choice that $\tau_j^{(n)} = \mathrm{sgn}(\alpha_j)/\mathrm{sgn}(\alpha_1)$ for all $j \in L$ and all $\boldsymbol{\tau}^{(n)}$, the first $|L|$ rows of $T^{(k)}$ yield redundant equations in Eq. (19) of the main text. Therefore, we can define $\tilde{T}^{(k)}$ as $T^{(k)}$ with all rows $j \in L \setminus \{1\}$ eliminated. Similarly, $\tilde{\boldsymbol{\alpha}}$ is $\boldsymbol{\alpha}$ with elements $j \in L \setminus \{1\}$ eliminated. Furthermore, define the new system of equations, which we call system $\tilde{A}$:

$$\tilde{T}^{(k)}\tilde{\boldsymbol{p}}^{(k)} = \tilde{\boldsymbol{\alpha}}/\alpha_1, \tag{F23}$$

$$\tilde{\boldsymbol{p}}^{(k)} \geqslant 0. \tag{F24}$$

System A has a solution if and only if system $\tilde{A}$ does. We can proceed as in the proof in Appendix D to show via the Farkas-Minkowski lemma that system $\tilde{A}$ has a solution if $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty \leqslant k \Rightarrow \|\tilde{\boldsymbol{\alpha}}\|_1/\|\tilde{\boldsymbol{\alpha}}\|_\infty \leqslant k - |L| + 1$. The details of the proof of this part are completely identical with this substitution.

The second part of the proof can similarly be adjusted straightforwardly. In particular, to satisfy the condition of Eq. (F17), which is the generalization of Eq. (8) in the main text, for $j \in L$ we require

$$\frac{\alpha_j}{\alpha_1}t^2 = \frac{\mathrm{sgn}(\alpha_j)}{\mathrm{sgn}(\alpha_1)}t^2 = \sum_{i\in L}\frac{\mathrm{sgn}(\alpha_1)}{\mathrm{sgn}(\alpha_i)}\mathcal{F}(\boldsymbol{\theta})_{ij}\lambda_i, \tag{F25}$$

which implies

$$t^2 = \sum_{i\in L}\frac{\mathrm{sgn}(\alpha_i)}{\mathrm{sgn}(\alpha_j)}\mathcal{F}(\boldsymbol{\theta})_{ij}\lambda_i. \tag{F26}$$

This in turn implies that for $i, j \in L$

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = \frac{\mathrm{sgn}(\alpha_i)}{\mathrm{sgn}(\alpha_j)}t^2. \tag{F27}$$

Therefore, for all $i \in L$ we require $\mathcal{F}(\boldsymbol{\theta})_{ii} = t^2$. From here, arguments identical to those in Appendix D apply to all $i \in L$, not just $i = 1$. That is, all the probe states must always be fully entangled on the qubits in L and matrix elements $\mathcal{F}(\boldsymbol{\theta})_{ij}$ for $i \in L$, $j \notin L$ can only accumulate magnitude if sensor j is also entangled with the qubits in L. Assuming the existence of an optimal protocol using $(k-1)$-partite entanglement, a contradiction arises in an identical way.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, Science **306**, 1330 (2004).

[2] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Metrology, Phys. Rev. Lett. **96**, 010401 (2006).

[3] L. Pezzé and A. Smerzi, Entanglement, Nonlinear Dynamics, and the Heisenberg Limit, Phys. Rev. Lett. **102**, 100401 (2009).

[4] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, Fisher information and multiparticle entanglement, Phys. Rev. A **85**, 022321 (2012).

[5] G. Tóth, Multipartite entanglement and high-precision metrology, Phys. Rev. A **85**, 022322 (2012).

[6] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acín, and M. Lewenstein, Asymptotic role of entanglement in quantum metrology, Phys. Rev. A **94**, 012339 (2016).

[7] G. Tóth and I. Apellaniz, Quantum metrology from a quantum information science perspective, J. Phys. A: Math. Theor. **47**, 424006 (2014).

[8] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, Quantum-enhanced measurements without entanglement, Rev. Mod. Phys. **90**, 035006 (2018).

[9] A. Luis, Phase-shift amplification for precision measurements without nonclassical states, Phys. Rev. A **65**, 025802 (2002).

[10] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free Heisenberg-limited phase estimation, Nature (London) **450**, 393 (2007).

[11] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Improvement of Frequency Standards with Quantum Entanglement, Phys. Rev. Lett. **79**, 3865 (1997).

[12] B. Escher, R. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, Nat. Phys. **7**, 406 (2011).

[13] S. Boixo and C. Heunen, Entangled and Sequential Quantum Protocols with Dephasing, Phys. Rev. Lett. **108**, 120402 (2012).

[14] R. Demkowicz-Dobrzański and L. Maccone, Using Entanglement Against Noise in Quantum Metrology, Phys. Rev. Lett. **113**, 250801 (2014).

[15] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Generalized Limits for Single-Parameter Quantum Estimation, Phys. Rev. Lett. **98**, 090401 (2007).

[16] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves, Quantum-limited metrology with product states, Phys. Rev. A **77**, 012317 (2008).

[17] T. Tilma, S. Hamaji, W. J. Munro, and K. Nemoto, Entanglement is not a critical resource for quantum metrology, Phys. Rev. A **81**, 022108 (2010).

[18] G. Gour and R. W. Spekkens, The resource theory of quantum reference frames: Manipulations and monotones, New J. Phys. **10**, 033023 (2008).

[19] I. Marvian and R. W. Spekkens, Extending Noether's theorem by quantifying the asymmetry of quantum states, Nat. Commun. **5**, 3821 (2014).

[20] I. Marvian and R. W. Spekkens, How to quantify coherence: Distinguishing speakable and unspeakable notions, Phys. Rev. A **94**, 052324 (2016).

[21] C. Zhang, B. Yadin, Z.-B. Hou, H. Cao, B.-H. Liu, Y.-F. Huang, R. Maity, V. Vedral, C.-F. Li, G.-C. Guo, and D. Girolami, Detecting metrologically useful asymmetry and entanglement by a few local measurements, Phys. Rev. A **96**, 042327 (2017).

[22] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), Vol. 3.

[23] S. L. Braunstein and C. M. Caves, Statistical Distance and the Geometry of Quantum States, Phys. Rev. Lett. **72**, 3439 (1994).

[24] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Generalized uncertainty relations: Theory, examples, and Lorentz invariance, Ann. Phys. **247**, 135 (1996).

[25] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Springer Science & Business Media, Berlin, 2011), Vol. 1.

[26] T. J. Proctor, P. A. Knott, and J. A. Dunningham, Networked quantum sensing, arXiv:1702.04271.

[27] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, Optimal and secure measurement protocols for quantum sensor networks, Phys. Rev. A **97**, 042337 (2018).

[28] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Distributed Quantum Metrology with Linear Networks and Separable Inputs, Phys. Rev. Lett. **121**, 043604 (2018).

[29] T. J. Proctor, P. A. Knott, and J. A. Dunningham, Multiparameter Estimation in Networked Quantum Sensors, Phys. Rev. Lett. **120**, 080501 (2018).

[30] S. Altenburg and S. Wölk, Multi-parameter estimation: Global, local and sequential strategies, Phys. Scr. **94**, 014001 (2019).

[31] J. Rubio, P. A. Knott, T. J. Proctor, and J. A. Dunningham, Quantum sensing networks for the estimation of linear functions, J. Phys. A: Math. Theor. **53**, 344001 (2020).

[32] J. Gross and C. Caves, One from many: Estimating a function of many parameters, J. Phys. A: Math. Theor. **54**, 014001 (2021).

[33] D. Triggiani, P. Facchi, and V. Tamma, Heisenberg scaling precision in the estimation of functions of parameters, Phys. Rev. A **104**, 062603 (2021).

[34] C. Oh, L. Jiang, and C. Lee, Distributed quantum phase sensing for arbitrary positive and negative weights, Phys. Rev. Res. **4**, 023164 (2022).

[35] M. Malitesta, A. Smerzi, and L. Pezzè, Distributed quantum sensing with squeezed-vacuum light in a configurable network of Mach-Zehnder interferometers, arXiv:2109.09178.

[36] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, Heisenberg-scaling measurement protocol for analytic functions with quantum sensor networks, Phys. Rev. A **100**, 042304 (2019).

[37] J. Bringewatt, I. Boettcher, P. Niroula, P. Bienias, and A. V. Gorshkov, Protocols for estimating multiple functions with quantum sensor networks: Geometry and performance, Phys. Rev. Res. **3**, 033011 (2021).

[38] T. Qian, J. Bringewatt, I. Boettcher, P. Bienias, and A. V. Gorshkov, Optimal measurement of field properties with quantum sensor networks, Phys. Rev. A **103**, L030601 (2021).

[39] Thus, the Hilbert space under consideration is a $(d + n_a)$-qubit Hilbert space of dimension 2^{d+n_a}, where n_a is the number of ancillas.

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

[41] J. Liu, H. Yuan, X. Lu, and X. Wang, Quantum fisher information matrix and multiparameter estimation, J. Phys. A: Math. Theor. **53**, 023001 (2020).

[42] S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Springer, Berlin, 1985).

[43] Y. Yang, G. Chiribella, and M. Hayashi, Attaining the ultimate precision limit in quantum state estimation, Commun. Math. Phys. **368**, 223 (2019).

[44] J. Suzuki, Nuisance parameter problem in quantum estimation theory: Tradeoff relation and qubit examples, J. Phys. A: Math. Theor. **53**, 264001 (2020).

[45] J. Suzuki, Y. Yang, and M. Hayashi, Quantum state estimation with nuisance parameters, J. Phys. A: Math. Theor. **53**, 453001 (2020).

[46] For instance, the conditions in Eqs. (6) and (7) are equivalent to the so-called global parameter orthogonality condition discussed in Sec. 5.5 of Ref. [45].

[47] It is worth pointing out that it is not strictly necessary to reduce the problem to single-qubit phase estimation. The reason we consider disentangling all qubits is to reduce fully to the single-qubit phase estimation problem of the robust phase estimation papers in Refs. [51–53], described below. However, one could apply essentially equivalent protocols by forgoing the disentangling of the qubits and simply performing parity measurements on the final catlike state. Such parity measurements can be carried out by simply measuring all qubits individually.

[48] B. Higgins, D. Berry, S. Bartlett, M. Mitchell, H. Wiseman, and G. Pryde, Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements, New J. Phys. **11**, 073023 (2009).

[49] M. Hayashi, S. Vinjanampathy, and L. Kwek, Resolving unattainable Cramer-Rao bounds for quantum sensors, J. Phys. B: At., Mol. Opt. Phys. **52**, 015503 (2019).

[50] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, π-Corrected Heisenberg Limit, Phys. Rev. Lett. **124**, 030501 (2020).

[51] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92**, 062315 (2015).

[52] S. Kimmel, G. H. Low, and T. J. Yoder, Erratum: Robust calibration of a universal single-qubit gate set via robust phase estimation [Phys. Rev. A **92**, 062315 (2015)], Phys. Rev. A **104**, 069901(E) (2021).

[53] F. Belliardo and V. Giovannetti, Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error, Phys. Rev. A **102**, 042613 (2020).

[54] J. Farkas, Theorie der einfachen ungleichungen, J. Reine Angew. Math. **1902**, 1 (1902).

[55] N. Dinh and V. Jeyakumar, Farkas' lemma: Three decades of generalizations for mathematical optimization, TOP **22**, 1 (2014).

[56] These gates are not strictly necessary. See note [47].

[57] Code is available upon request.

[58] Note that the above block of equations relies on the fact that we are using the fixed Hilbert space of qubit sensors. Were one to extend this derivation to photonic sensors with indefinite particle number, the results would not immediately follow.

[59] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).