

Minimum Entanglement Protocols for Function Estimation

Adam Ehrenberg,^{*} Jacob Bringewatt,^{*} and Alexey V. Gorshkov

Joint Center for Quantum Information and Computer Science,

NIST/University of Maryland College Park, Maryland 20742, USA and

Joint Quantum Institute, NIST/University of Maryland College Park, Maryland 20742, USA

(Dated: October 18, 2021)

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 networks applications. We demonstrate how to select different protocols from this family under various constraints via linear programming. Focusing 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 entangled cat-like states.

Entanglement is a hallmark of quantum theory and plays an essential role in many applications of quantum technology. Consider single-parameter metrology, where one seeks to determine an unknown phase shift θ that is independently and identically coupled to d sensors via a linear Hamiltonian \hat{H} . Given a probe state ρ , evolution under \hat{H} encodes θ into ρ where it can then be measured. It is well known that if the particles 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 if and 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 is allowed to apply the encoding unitary multiple times) [9, 10], in the presence of decoherence [11–14], when the coupling Hamiltonian is non-linear [15–17], or in reference to resource theories for metrology [18–21].

In this Letter, 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. In particular, we revisit the problem of optimally measuring a single linear function $q(\boldsymbol{\theta})$ [26–35]. This problem is a crucial element of optimal protocols for more general quantum sensor network problems. More specifically, the case of measuring one or multiple analytic functions [36, 37] and the case where the parameters $\boldsymbol{\theta}$ are not independent [38] reduce 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 param-

eters non-trivially 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 whole family of probabilistic protocols that obtain the optimal performance in this setting, and we provide an explicit linear programming method that obtains solutions from this family while taking into consideration experimentally relevant constraints for actual systems. Using this method, we derive solutions that minimize, and, in many cases, even eliminate, the use of highly entangled states. Furthermore, we prove necessary and sufficient conditions on q for the existence of optimal protocols using at most ($k < d$)-partite entanglement: in short, the more uniformly distributed q is amongst the unknown parameters, the more entanglement is required. We also provide a protocol for when these conditions are not satisfied but entanglement resources are still constrained.

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]. We consider a network of d qubit quantum sensors coupled to d independent, unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ via the Hamiltonian

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

with $\hat{\sigma}_i^{x,y,z}$ the Pauli operators acting on qubit i . The term $\hat{H}_c(t)$ is a time-dependent control Hamiltonian that may include coupling to ancilla qubits. This time-dependent control, while potentially useful for creating simpler protocols, is not necessary to achieve an optimal protocol [15, 27] and, therefore, may freely be set to zero, which we do for the rest of the paper [39]. We encode the parameters $\boldsymbol{\theta}$ into a quantum state ρ via the unitary evolution generated by this Hamiltonian. Given some choices of initial probe state, final measurements, and estimator for the quantity of interest, we seek to estimate

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, and we assume without loss of generality that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Ref. [27] established that the fundamental limit for the mean square error \mathcal{M} of q is

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_\infty^2}{t^2}. \quad (2)$$

Eq. (2) holds for estimating q with a single trial (for many trials, simply divide the RHS by the number of trials).

Eq. (2) is derived via the single-parameter quantum Cramér-Rao bound [15, 22–25]. This is somewhat surprising: while we only seek to measure a single quantity $q(\boldsymbol{\theta})$, d parameters control the evolution under Eq. (1), so we do not *a priori* satisfy the condition for the use of the single-parameter quantum Cramér-Rao bound. However, we can justify its validity for our system as follows: 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 in reality, 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. However, for a bound derived this way to be tight there must be some choice(s) of artificially fixing $d-1$ degrees of freedom that gives us no information about $q(\boldsymbol{\theta})$; that is, no *useful* information. In the next section, we will show such a choice exists on information theoretic grounds. However, the existence of explicit protocols in Refs. [27, 32] that saturate Eq. (2) also attests to the tightness of such a bound.

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

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

where \mathcal{F} is the quantum Fisher information, $\hat{g}_q = \partial \hat{H} / \partial q$ (where the partial derivative fixes the other $d-1$ degrees of freedom), and $\|\hat{g}_q\|_s$ is the seminorm given by 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].

Conditions for Saturable Bounds.—While the argument above justifies the validity of applying the single-parameter bound in our multiparameter scenario, it offers no road map for actually constructing optimal protocols. The quantum Fisher information matrix [40] provides an information-theoretic solution to this issue. For pure probe states and unitary evolution for time t under the Hamiltonian in Eq. (1), it has matrix elements

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4t^2 \text{Re}[\langle \hat{g}_i \hat{g}_j \rangle - \langle \hat{g}_i \rangle \langle \hat{g}_j \rangle], \quad (4)$$

where $\hat{g}_i = \partial \hat{H} / \partial \theta_i = \hat{\sigma}_i^{(z)} / 2$, and the expectation values are taken with respect to the probe state. Note that this expression depends on the fact that $[\hat{g}_i, \hat{g}_j] = 0 \forall i, j$. Choosing $d-1$ degrees of freedom to fix in hopes of using the single-parameter bound then corresponds to a basis transformation $\boldsymbol{\theta} \rightarrow \mathbf{q}$, 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 $\mathcal{F}(\mathbf{q}) = J^T \mathcal{F}(\boldsymbol{\theta}) J$.

To obtain the bound in Eq. (2) and have no information about $q(\boldsymbol{\theta})$ from the extra degrees of freedom $q_{j>1}$ (ensuring saturability of the single-parameter quantum Cramér-Rao bound), $\mathcal{F}(\mathbf{q})$ must have the following properties:

$$\mathcal{F}(\mathbf{q})_{11} = \frac{t^2}{\alpha_1^2}, \quad (5)$$

$$\mathcal{F}(\mathbf{q})_{1i} = \mathcal{F}(\mathbf{q})_{i1} = 0 \quad (\forall i \neq 1), \quad (6)$$

where we recall that we have let $|\alpha_1| = \|\boldsymbol{\alpha}\|_\infty$. Via the inverse basis transformation $\mathbf{q} \rightarrow \boldsymbol{\theta}$, we find Eqs. (5)-(6) are satisfied if and only if

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = \mathcal{F}(\boldsymbol{\theta})_{j1} = \frac{\alpha_j}{\alpha_1} t^2, \quad (7)$$

where we assume here and for the rest of the Letter that $|\alpha_1| > |\alpha_j|$ for all $j > 1$ for ease of presentation. Therefore, any valid protocol saturating Eq. (2) must use probe states that obtain this Fisher information matrix with respect to $\boldsymbol{\theta}$. The explicit derivation of Eq. (7), along with the generalization of our results beyond the assumption that $|\alpha_1| > |\alpha_j|$ for $j > 1$, is left to the Supplemental Material [41]. Importantly, our main result (see Theorem 1) is completely unchanged by this assumption, although the proof is slightly more tedious in the general case.

A Family of Optimal Protocols.—We now derive a family of protocols that achieve Eq. (7). Similar to Refs. [27, 32], we allow ourselves to use probabilistic protocols. Note “probabilistic” does not refer to randomness, but instead means that such protocols exploit the convexity of Fisher information matrices to asymptotically prepare a desired overall Fisher information matrix from a collection of N different sub-protocols, each performed a fraction p_n of the time. Mathematically:

$$\mathcal{F}(\boldsymbol{\theta}) = \sum_{n=1}^N p_n \mathcal{F}^{(n)}(\boldsymbol{\theta}). \quad (8)$$

An individual protocol consists of preparing a pure initial state $\rho^{(n)} = |\psi^{(n)}\rangle \langle \psi^{(n)}|$, evolving $\rho^{(n)}$ under the unitary $e^{-i\hat{H}t}$ for time t , performing some choice of measurement (specified by a positive operator-valued measure), and computing an estimator for q from the measurement outcomes. The individual $\mathcal{F}^{(n)}(\boldsymbol{\theta})$ can be computed given $\rho^{(n)}$ via Eq. (4).

At this point, we must specify some choice of states with which to produce a valid probabilistic protocol. We consider the following set \mathcal{T} of $N = 3^{d-1}$ cat-like states:

$$|\psi(\boldsymbol{\tau})\rangle = \frac{1}{\sqrt{2}}(|\boldsymbol{\tau}\rangle + |-\boldsymbol{\tau}\rangle), \quad (9)$$

where $\boldsymbol{\tau} \in \{0, \pm 1\}^d$ are vectors defining the states via

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

and we require that $\tau_1 = 1$, as any optimal protocol must always be sensitive to this parameter. It is straightforward to compute the Fisher information matrix with respect to each $|\psi(\boldsymbol{\tau})\rangle \in \mathcal{T}$ via Eq. (4). In particular, we obtain

$$\mathcal{F}^{(n)}(\boldsymbol{\theta})_{1j} = \mathcal{F}^{(n)}(\boldsymbol{\theta})_{j1} = \tau_j^{(n)} t^2, \quad (11)$$

where $\boldsymbol{\tau}^{(n)}$ corresponds to the n -th state of some enumeration of \mathcal{T} . Eqs. (7)-(8) then yield the condition

$$\frac{\alpha_j}{\alpha_1} = \sum_{n=1}^N p_n \tau_j^{(n)}, \quad (12)$$

for all j . Note that the $j = 1$ case is automatically satisfied for any valid probability distribution specified by the p_n , as $\tau_1^{(n)} = 1$ for all n . If we let \mathbf{p} be the vector of probabilities and define the $d \times N$ matrix T with matrix elements $T_{mn} = \tau_m^{(n)}$, Eq. (12) simplifies to

$$T\mathbf{p} = \frac{\boldsymbol{\alpha}}{\alpha_1}. \quad (13)$$

This means that *any* nonnegative solution (in the sense that $p_n \geq 0 \forall n$) to Eq. (13) specifies a valid set of states to optimally measure the function $q(\boldsymbol{\theta})$. To completely specify a protocol, we must also provide the appropriate measurements, which, here, is a parity measurement on the entangled sensors for all probe states [27, 32].

Because the system in Eq. (13) is highly underconstrained, such protocols do not necessarily use all 3^{d-1} states in \mathcal{T} . For example, the protocol presented in Section 5.1.3 of Ref. [32] uses only d states in \mathcal{T} . We will see, however, that this protocol is not unique.

As an explicit example, consider the case of 2 qubits. The available states are described by the T matrix

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

By Eq. (13), an optimal protocol must satisfy

$$p_1 + p_2 + p_3 = 1, \quad (15)$$

$$p_1 - p_2 = \frac{\alpha_2}{\alpha_1}. \quad (16)$$

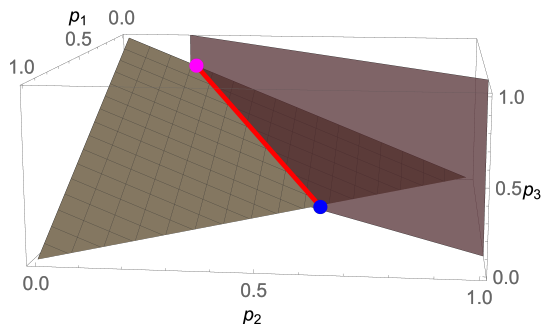


FIG. 1. Family of optimal protocols for $d = 2$, $\alpha_1 = 1$, and $\alpha_2 = -1/5$. Any point on the red line, the intersection of the plane $p_1 - p_2 = \alpha_2$ and the simplex $p_1 + p_2 + p_3 = 1$, subject to $\mathbf{p} \geq 0$, represents an optimal protocol. The blue (magenta) extremal point at the bottom (top) of the figure represents the two-state protocol (not) using exclusively maximally entangled states.

Solving in terms of p_1 leads to the 1-parameter family of solutions $p_2 = p_1 - \frac{\alpha_2}{\alpha_1}$, $p_3 = 1 + \frac{\alpha_2}{\alpha_1} - 2p_1$, and $p_n \in [0, 1]$ for all n . For simplicity, assume $\alpha_1 = 1$. Then non-negativity is achieved by

$$p_1 \in \begin{cases} [\alpha_2, \frac{1+\alpha_2}{2}] & \alpha_2 \geq 0 \\ [0, \frac{1+\alpha_2}{2}] & \alpha_2 < 0 \end{cases}. \quad (17)$$

These solutions are defined by the intersection of $d = 2$ hyperplanes in $\mathbb{R}^{3^{d-1}} = \mathbb{R}^3$ appropriately restricted to valid probability distributions. The first plane is the standard 3D probability simplex and the second plane is defined by $(p_1, p_1 - \alpha_2, p_3)$. See Fig. 1.

If $p_3 = 0$, we obtain $p_1 = \frac{1}{2}(1 + \alpha_2)$ and $p_2 = \frac{1}{2}(1 - \alpha_2)$, which is precisely the $d = 2$ protocol described in Ref. [32], which we now note is simply one solution among many. There is even a two-state protocol that does not require using exclusively maximally entangled states: for $\alpha_2 > 0$ ($\alpha_2 < 0$), let $p_1 = \alpha_2$ (0) so that $p_2 = 0$ ($-\alpha_2$) and $p_3 = 1 - \alpha_2$ ($1 + \alpha_2$). In the Supplemental Material, we generalize this example to the case where $d > 2$ [41].

Minimum Entanglement Solutions.—We now focus on solutions from this family that require the minimum amount of entanglement across the sensors. In stark contrast to the protocols in Refs. [27, 32], many $q(\boldsymbol{\theta})$ admit optimal protocols that do not require d -partite entanglement. Specifically, we can prove necessary and sufficient conditions on $\boldsymbol{\alpha}$ for the existence of a protocol that uses only k -partite entanglement. We have the following theorem, which is the primary technical result of our Letter.

Theorem 1. *Assume, without loss of generality, that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $1 \leq k \leq d$ be an integer. Define $T^{(k)}$ to be the submatrix of T such that 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. Then the system

$$T^{(k)} \mathbf{p}^{(k)} = \boldsymbol{\alpha} / \alpha_1, \quad (18)$$

$$\mathbf{p}^{(k)} \geq 0, \quad (19)$$

has a solution if and only if

$$\|\boldsymbol{\alpha}\|_1 / \|\boldsymbol{\alpha}\|_\infty \leq k. \quad (20)$$

See the Supplemental Material for a detailed example of the theorem for the case of three qubits [41]. The proof requires the following standard linear algebra lemma [42, 43], which, geometrically, is an application of the hyperplane separation theorem [44].

Lemma 1 (Farkas-Minkowski). *Consider the system*

$$A\mathbf{x} = \mathbf{b}, \quad (21)$$

$$\mathbf{x} \geq 0, \quad (22)$$

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

$$A^\top \mathbf{y} \geq 0, \quad (23)$$

$$\langle \mathbf{b}, \mathbf{y} \rangle < 0. \quad (24)$$

We now proceed with the proof of Theorem 1.

Proof. Call Eqs. (18)-(19) System A. Let $\boldsymbol{\alpha}' = \boldsymbol{\alpha} / \alpha_1$ and define System B as

$$(T^{(k)})^\top \mathbf{y} \geq 0, \quad (25)$$

$$\langle \boldsymbol{\alpha}', \mathbf{y} \rangle < 0. \quad (26)$$

By Lemma 1, System A has a solution if and only if System B does not, so it will be sufficient to show that System B (a) has a solution if $\sum_{j>1} |\alpha'_j| > k - 1$; and, (b) does not have a solution if $\sum_{j>1} |\alpha'_j| \leq k - 1$, where we used that $\alpha'_1 = 1$. We prove (a) with an explicit construction. First, observe that $k \geq 1$, and, therefore, $(T^{(k)})^\top$ has a row n^* given by $\boldsymbol{\tau}^{(n^*)} = (1, 0, \dots, 0)$. Thus, Eq. (25) implies that any solution \mathbf{y} to System B has $y_1 \geq 0$. Choose any $y_1 > 0$ and let $y_j = -\text{sgn}(\alpha'_j) y_1 / (k - 1)$ for all $j > 1$. This \mathbf{y} automatically satisfies Eq. (25): any row of $(T^{(k)})^\top$ has at most k non-zero entries, so the smallest any element of $(T^{(k)})^\top \mathbf{y}$ can be is $y_1 - |y_{i_1}| - \dots - |y_{i_{k-1}}| = 0$ for any choice of $k - 1$ indices i_1, \dots, i_{k-1} . This choice of \mathbf{y} substituted into Eq. (26) yields

$$\langle \boldsymbol{\alpha}', \mathbf{y} \rangle = y_1 \left(1 - \frac{1}{k-1} \sum_{j>1} |\alpha'_j| \right) < 0, \quad (27)$$

as y_1 is strictly greater than 0 and $\sum_{j>1} |\alpha'_j| > (k - 1)$. Thus, we have a solution to System B, proving (a).

Now we prove (b): we will assume that $\sum_{j>1} |\alpha'_j| \leq k - 1$ and that a solution \mathbf{y} exists and will arrive at a contradiction. Without loss of generality, we assume

that $|y_j| \geq |y_{j+1}|$ for all $1 < j < d$. Eq. (26) implies $\sum_{j>1} \alpha'_j y_j < -y_1$. As $y_1 \geq 0$, this can be rewritten as $|\sum_{j>1} \alpha'_j y_j| > y_1$, which, by the triangle inequality, implies

$$\sum_{j>1} |\alpha'_j| |y_j| > y_1. \quad (28)$$

Since $|\alpha'_j| \leq 1$ for all j , since $\sum_{j>1} |\alpha'_j| \leq k - 1$, and since $|y_j|$ for $j > 1$ are ordered in descending order, the largest the left-hand-side of Eq. (28) can be is $\sum_{j=2}^k |y_j|$, leading to

$$\sum_{j=2}^k |y_j| > y_1. \quad (29)$$

This directly contradicts Eq. (25) for the row of $T^{(k)}$ given by $\boldsymbol{\tau} = (1, -\text{sgn}(y_2), \dots, -\text{sgn}(y_k), 0, 0, \dots)$. \square

Theorem 1 provides conditions for the existence of solutions to Eq. (13) with limited entanglement, but it is not constructive. When these conditions are satisfied, one can find a solution to Eq. (13) with certain desirable properties (i.e., as determined by experimental constraints) via a linear program that minimizes a cost function $\mathcal{E}(\mathbf{p})$, linear in \mathbf{p} , accounting for these desiderata:

$$\begin{aligned} \min_{\mathbf{p}} \mathcal{E}(\mathbf{p}) \\ \text{s.t. } T\mathbf{p} = \boldsymbol{\alpha} / \alpha_1 \\ p_n \geq 0 \quad \forall n. \end{aligned} \quad (30)$$

The first row of T is the vector of all ones, which ensures that any solution \mathbf{p} is normalized. Using cost functions \mathcal{E} that penalize highly entangled states, we may find solutions that minimize the total amount of entanglement used, weighted by the frequency with which such entanglement appears in the probabilistic protocol. However, the linear program in Eq. (30) is not efficient in the number of sensors, as it takes as input an exponentially large matrix T of size $d \times 3^{d-1}$, and linear programs run in time polynomial in the input size. In the case of limited entanglement, however, Theorem 1 allows us to reduce the size of the input to the linear program. In particular, we know that a solution exists with maximum entanglement of size k if and only if $\|\boldsymbol{\alpha}\|_1 / \|\boldsymbol{\alpha}\|_\infty \leq k$. Therefore, given $\boldsymbol{\alpha}$, we may simply restrict our input to $T^{(k)}$, with $k = \lceil \|\boldsymbol{\alpha}\|_1 / \|\boldsymbol{\alpha}\|_\infty \rceil$. If k scales at most logarithmically with d , then $T^{(k)}$ has size polynomial in d , and the linear program will be efficient. We can then use a cost function \mathcal{E} to encode other parameters that limit the physical system besides the maximum entanglement size used. If there are no other restrictions, any linear cost function will allow us to efficiently pick out a valid protocol using at most k -partite entanglement [45].

Entanglement Constraints.—In the previous section, we derived conditions for which the optimal variance estimation of $q(\theta)$ can be achieved with only k -partite entanglement for $k < d$, despite the fact that the bound in Eq. (2) that we saturate allows for arbitrary states. This raises the question as to what the optimal approach is when these conditions are *not* satisfied, but we are still limited to k -partite entanglement.

We propose the following protocol: 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 $\alpha^{(r)}$ denote α 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 given by

$$\mathcal{M} = \frac{1}{t^2} \sum_{r \in R} \left\| \alpha^{(r)} \right\|_{\infty}^2 =: \frac{\left(\left\| \alpha \right\|_{\infty}^{(k)} \right)^2}{t^2}. \quad (31)$$

The optimal R is a partition of the sensors into contiguous sets (assuming for simplicity that $|\alpha_i| \geq |\alpha_j|$ for $i < j$) such that for all $r \in R$, $\sum_{i \in r} |\alpha_i| / \max_{i \in r} |\alpha_i| \leq k$, satisfying Theorem 1. Note that it is efficient to determine R , taking times $O(d \log d)$ to sort the components of α and $O(d)$ to work through this list and optimally partition it.

We conjecture that this protocol is optimal, but leave the determination of the ultimate lower bound on \mathcal{M} as an open question. Clearly, if partitioning the problem into independent sets is optimal, our protocol is, indeed, optimal. However, one could imagine probabilistic protocols that use different partitions for some fraction of the runs and we cannot rule out that such protocols provide better performance than ours.

Conclusion and Outlook.—In this Letter, we have proven that maximally entangled states are not necessary for the optimal measurement of a linear function with a quantum sensor networks unless the function is sufficiently uniformly supported on the unknown parameters. This result, combined with the general framework of using linear programming to find optimal protocols subject to practical constraints, is of particular relevance to the development of near-term quantum sensor networks, where creating large-scale entangled states may not be practical. We emphasize again that these results are also useful in more general settings, such as the measurement of analytic functions, as these measurements reduce precisely to the case studied here [36–38]. In this work we have only considered a particular class of popular metrological states. We leave as an open question whether these results hold for other classes of states, such as Dicke states [46, 47]. As discussed, we also leave as an open question whether the entanglement-limited protocol presented in the previous section is indeed optimal.

Acknowledgements.—We thank Michael Foss-Feig, Zachary Eldredge, Pradeep Niroula, Luis Pedro García-Pintos, and Michael Gullans for helpful discussions. We acknowledge funding by AFOSR MURI, AFOSR, DoE QSA, DARPA SAVaNT ADVENT, NSF PFCQC program, ARO MURI, DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), U.S. Department of Energy Award No. DE-SC0019449, and the DoE ASCR Quantum Testbed Pathfinder program (award No. DE-SC00119040). 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).

* These two authors contributed equally.

- [1] V. Giovannetti, S. Lloyd, and L. Maccone, *Science* **306**, 1330 (2004).
- [2] V. Giovannetti, S. Lloyd, and L. Maccone, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [3] L. Pezzé and A. Smerzi, *Phys. Rev. Lett.* **102**, 100401 (2009).
- [4] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, *Phys. Rev. A* **85**, 022321 (2012).
- [5] G. Tóth, *Phys. Rev. A* **85**, 022322 (2012).
- [6] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acín, and M. Lewenstein, *Phys. Rev. A* **94**, 012339 (2016).
- [7] G. Tóth and I. Apellaniz, *J. of Phys. A: Math. Theor.* **47**, 424006 (2014).
- [8] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, *Rev. Mod. Phys.* **90**, 035006 (2018).
- [9] A. Luis, *Phys. Rev. A* **65**, 025802 (2002).
- [10] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, *Nature* **450**, 393 (2007).
- [11] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *Phys. Rev. Lett.* **79**, 3865 (1997).
- [12] B. Escher, R. de Matos Filho, and L. Davidovich, *Nat. Phys.* **7**, 406 (2011).
- [13] S. Boixo and C. Heunen, *Phys. Rev. Lett.* **108**, 120402 (2012).
- [14] R. Demkowicz-Dobrzański and L. Maccone, *Phys. Rev. Lett.* **113**, 250801 (2014).
- [15] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, *Phys. Rev. Lett.* **98**, 090401 (2007).
- [16] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves, *Phys. Rev. A* **77**, 012317 (2008).
- [17] T. Tilma, S. Hamaji, W. J. Munro, and K. Nemoto, *Phys. Rev. A* **81**, 022108 (2010).
- [18] G. Gour and R. W. Spekkens, *New J. Phys.* **10**, 033023 (2008).
- [19] I. Marvian and R. W. Spekkens, *Nat. comm.* **5**, 1 (2014).
- [20] I. Marvian and R. W. Spekkens, *Phys. Rev. A* **94**, 052324 (2016).

- [21] C. Zhang, B. Yadin, Z. Hou, H. Cao, B. Liu, Y. Huang, R. Maity, V. Vedral, C. Li, G. Guo, and D. Girolami, *Phys. Rev. A* **96**, 042327 (2017).
- [22] C. W. Helstrom, *Quantum detection and estimation theory*, Vol. 3 (Academic Press, New York, 1976).
- [23] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [24] S. L. Braunstein, C. M. Caves, and G. J. Milburn, *Ann. Phys. (N. Y.)* **247**, 135 (1996).
- [25] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, Vol. 1 (Springer Science & Business Media, Berlin, 2011).
- [26] T. J. Proctor, P. A. Knott, and J. A. Dunningham, *arXiv preprint arXiv:1702.04271* (2017).
- [27] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, *Phys. Rev. A* **97**, 042337 (2018).
- [28] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, *Phys. Rev. Lett.* **121**, 043604 (2018).
- [29] T. J. Proctor, P. A. Knott, and J. A. Dunningham, *Phys. Rev. Lett.* **120**, 080501 (2018).
- [30] S. Altenburg and S. Wölk, *Phys. Scr.* **94**, 014001 (2018).
- [31] J. Rubio, P. A. Knott, T. J. Proctor, and J. A. Dunningham, *J. Phys. A* **53**, 344001 (2020).
- [32] J. Gross and C. Caves, *J. Phys. A: Math. Theor.* **54**, 014001 (2020).
- [33] D. Triggiani, P. Facchi, and V. Tamma, *arXiv preprint arXiv:2103.08564* (2021).
- [34] C. Oh, L. Jiang, and C. Lee, *arXiv preprint arXiv:2108.04119* (2021).
- [35] M. Malitesta, A. Smerzi, and L. Pezzè, *arXiv preprint arXiv:2109.09178* (2021).
- [36] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, *Phys. Rev. A* **100**, 042304 (2019).
- [37] J. Bringewatt, I. Boettcher, P. Niroula, P. Bienias, and A. Gorshkov, *Phys. Rev. Research* **3**, 033011 (2021).
- [38] T. Qian, J. Bringewatt, I. Boettcher, P. Bienias, and A. V. Gorshkov, *Phys. Rev. A* **103**, L030601 (2021).
- [39] It is worth noting that, if non-zero, the term $H_c(t)$ could increase entanglement relative to the initial probe state. If one seeks to minimize the use of entanglement via the protocols explored in this Letter, any entanglement-generating $H_c(t)$ should not be available. In the cases considered here, such $H_c(t)$ are also unnecessary.
- [40] J. Liu, H. Yuan, X. Lu, and X. Wang, *J. Phys. A: Math. Theor.* **53**, 023001 (2019).
- [41] Refer to the Supplemental Material where we generalize beyond the assumption that $|\alpha_1| > |\alpha_j|$ for all $j > 1$, provide a detailed derivation of Eq. (7), construct various optimal protocols with different entanglement structures, and examine Theorem 1 for three qubits.
- [42] J. Farkas, *J. Reine Angew. Math.* **1902**, 1 (1902).
- [43] N. Dinh and V. Jeyakumar, *TOP* **22**, 1 (2014).
- [44] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge University Press, New York, 2004).
- [45] As a simple physical example where we might care about a linear function requiring $\log(d)$ -partite entanglement, consider a 1D array of d evenly spaced, unknown charges θ_i , with each charge coupled independently to a sensor (as is approximately the case if the sensors are close enough to their respective charges). If one seeks to estimate the electrostatic potential at some distance to the right of the chain—a linear function in θ —then $\alpha_i \sim 1/i$, and $\|\alpha\|_1 / \|\alpha\|_\infty \sim \log(d)$.
- [46] Z. Zhang and L. M. Duan, *New J. of Phys.* **16**, 103037 (2014).
- [47] B. Lücke, J. Peise, G. Vitagliano, J. Arlt, L. Santos, G. Tóth, and C. Klempt, *Phys. Rev. Lett.* **112**, 155304 (2014).

SUPPLEMENTAL MATERIAL FOR: MINIMUM ENTANGLEMENT PROTOCOLS FOR FUNCTION ESTIMATION

In this Supplemental Material, we generalize beyond the assumption in the main text that $|\alpha_1| > |\alpha_j|$ for all $j > 1$ and give a formal proof of Eq. (7) from the main text (Sec. I), use our new approach to reconstruct the protocol originally developed in [32] (Sec. II) and derive an optimal protocol using less entanglement (Sec. III), and give an example of Theorem 1 from the main text for the simple case of three qubits (Sec. IV).

CONTENTS

I. Relaxing the Assumption on a Single Maximum Element	7
II. Rederiving the Protocol in Ref. [32]	9
III. Deriving a Less Entangled Protocol	10
IV. Example of Theorem 1 for Three Qubits	12

I. RELAXING THE ASSUMPTION ON A SINGLE MAXIMUM ELEMENT

In this section, 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 than in the main text. In the process, we will rigorously prove Eq. (7) from the main text.

Let $L = \{i \mid |\alpha_i| = |\alpha_1|\}$. The assumption $|\alpha_1| > |\alpha_j|$ for all $j > 1$, stated in the main text, is equivalent to assuming $|L| = 1$. We presented the following set of conditions for the single-parameter bound on $q(\boldsymbol{\theta})$ to be saturable:

$$\mathcal{F}(\mathbf{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{S.1}$$

$$\mathcal{F}(\mathbf{q})_{1i} = \mathcal{F}(\mathbf{q})_{i1} = 0 \quad (\forall i \neq 1), \tag{S.2}$$

where we recall that $\mathcal{F}(\mathbf{q}) = J^T \mathcal{F}(\boldsymbol{\theta}) J$. Here J is the Jacobian for the basis transformation from $\boldsymbol{\theta}$ to \mathbf{q} , where $q_1 = q$ 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. (S.1)-(S.2) are satisfied if and only if

$$\sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji} \lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \tag{S.3}$$

where $\lambda_i \geq 0$ such that $\sum_i \lambda_i = 1$. If $|L| = 1$, this reduces to Eq. (7) of the main text.

It will be important to briefly recount how we obtain the single-parameter bound we are trying to saturate (see Ref. [27] for further details). In particular, referring to Eq. (3) of the main text, we seek a choice of basis that maximizes $\|\hat{g}_q\|_s^2$, 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 $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(d)}\}$, where $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}$. We then have that J^{-1} has columns given by these vectors. Let $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \dots, \boldsymbol{\beta}^{(d)}\}$ be the basis dual to this one. That is, these vectors form the rows of J and satisfy $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$. We can then write

$$\boldsymbol{\theta}^T = \boldsymbol{\theta}^T J^{-1} J = ((J^{-1})^T \boldsymbol{\theta})^T J = \sum_{i=1}^d (\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\theta}) (\boldsymbol{\beta}^{(i)})^T, \tag{S.4}$$

which allows us to rewrite our Hamiltonian in the convenient form

$$\hat{H} = \frac{1}{2} \boldsymbol{\theta}^T \hat{\boldsymbol{\sigma}} = \frac{1}{2} \sum_{i=1}^d (\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\theta}) \boldsymbol{\beta}^{(i)} \cdot \hat{\boldsymbol{\sigma}}, \tag{S.5}$$

where $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1^{(z)}, \dots, \hat{\sigma}_1^{(z)})^T$. Then

$$\hat{g}_q = \frac{\partial \hat{H}}{\partial q} = \frac{\partial \hat{H}}{\partial(\boldsymbol{\alpha}^{(1)} \cdot \boldsymbol{\theta})} = \frac{\boldsymbol{\beta} \cdot \hat{\boldsymbol{\sigma}}}{2}, \quad (\text{S.6})$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$. Therefore, we immediately have that

$$\|\hat{g}_q\|_s = \|\boldsymbol{\beta}\|_1, \quad (\text{S.7})$$

and our tightest bound is given by

$$\begin{aligned} & \min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \\ & \text{s.t. } \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \end{aligned} \quad (\text{S.8})$$

To establish the if and only if condition, we can find all solutions to Eq. (S.8) as follows. In particular, note that

$$1 = \sum_i \alpha_i \beta_i \leq \sum_i |\alpha_i| |\beta_i| \leq |\alpha_1| \sum_i |\beta_i| = |\alpha_1| \|\boldsymbol{\beta}\|_1. \quad (\text{S.9})$$

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 $L = \{i \mid |\alpha_i| = |\alpha_1|\}$. Then the second inequality is tight if and only if

$$\beta_i = 0 \text{ for } i \notin L, \quad (\text{S.10})$$

$$\sum_{i \in L} |\beta_i| = \frac{1}{|\alpha_1|}. \quad (\text{S.11})$$

Any solution $\boldsymbol{\beta}$ specifies the first column of the Jacobian J and allows us to rewrite the conditions in Eq. (S.1) as

$$\mathcal{F}(\mathbf{q})_{11} = \boldsymbol{\beta}^T \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = \frac{t^2}{\alpha_1^2}, \quad (\text{S.12})$$

$$\mathcal{F}(\mathbf{q})_{1i} = \mathcal{F}(\mathbf{q})_{i1} = (\boldsymbol{\beta}^{(i)})^T \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = 0 \quad (\forall i \neq 1). \quad (\text{S.13})$$

The conditions $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$ mean that Eq. (S.13) immediately implies that the vector $\mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta}$ must be proportional to $\boldsymbol{\alpha}$ and Eq. (S.12) specifies the constant of proportionality. In particular, we require

$$\mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = \frac{t^2}{\alpha_1^2} \boldsymbol{\alpha}. \quad (\text{S.14})$$

Invoking Eqs. (S.10)-(S.11) 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 $\lambda_i \geq 0$ for $i \in L$ and $\lambda_i = 0$ for $i \notin L$ such that $\sum_i \lambda_i = 1$. The individual components of Eq. (S.14) imply

$$\sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ij} \text{sgn}(\alpha_i) \lambda_i = \sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ji} \text{sgn}(\alpha_i) \lambda_i = \frac{t^2}{|\alpha_1|} \alpha_j, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0, \quad (\text{S.15})$$

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

$$\sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ij} \lambda_i = \sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji} \lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0, \quad (\text{S.16})$$

which reduced to Eq. (7) of the main text, when $|L| = 1$, as desired.

We can also consider the derivation of the system of equations in Eq. (13) of the main text without the assumption of a single maximum-magnitude element of $\boldsymbol{\alpha}$. In particular, in the general case, we choose our probe states $|\psi(\boldsymbol{\tau})\rangle$ such that $\tau_i = \text{sgn}(\alpha_i) / \text{sgn}(\alpha_1)$ for all $i \in L$, as any optimal protocol must always be sensitive to these parameters, and the sign convention is necessary to ensure that our probe state attains the correct phase. This matches the assumption that $\tau_1 = 1$ when $|L| = 1$. Recalling from the main text that

$$\mathcal{F}^{(n)}(\boldsymbol{\theta})_{ij} = 4t^2 \text{Re} [\langle \hat{g}_i \hat{g}_j \rangle - \langle \hat{g}_i \rangle \langle \hat{g}_j \rangle], \quad (\text{S.17})$$

where the expectation values are taken with respect to the state specified by $\boldsymbol{\tau}^{(n)}$, we have that

$$\mathcal{F}^{(n)}(\boldsymbol{\theta})_{ij} = \mathcal{F}^{(n)}(\boldsymbol{\theta})_{ji} = \tau_i^{(n)} \tau_j^{(n)} t^2. \quad (\text{S.18})$$

Our condition for saturability for a probabilistic protocol then becomes

$$\frac{\alpha_j}{\alpha_1} = \sum_{i \in L} \lambda_i \sum_{n=1}^N p_n \tau_j^{(n)} = \sum_{n=1}^N p_n \tau_j^{(n)}, \quad (\text{S.19})$$

where now $N = 3^{d-|L|}$. This is the generalization of Eq. (13) of the main text.

Finally, we consider how our proof of Theorem 1 in the main text changes without the assumption that $|\alpha_1| > |\alpha_j|$ for all $j > 1$. Note that the theorem statement itself is unchanged when we relax this assumption. For ease of reference we repeat the theorem statement here.

Theorem S.1. *Assume, without loss of generality, that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $1 \leq k \leq d$ be an integer. Define $T^{(k)}$ to be the submatrix of T such that 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. Then the system*

$$T^{(k)} \boldsymbol{p}^{(k)} = \boldsymbol{\alpha} / \alpha_1, \quad (\text{S.20})$$

$$\boldsymbol{p}^{(k)} \geq 0, \quad (\text{S.21})$$

has a solution if and only if

$$\|\boldsymbol{\alpha}\|_1 / \|\boldsymbol{\alpha}\|_\infty \leq k. \quad (\text{S.22})$$

Note that, given our choice that $\tau_i = \text{sgn}(\alpha_i) / \text{sgn}(\alpha_1)$ for all $i \in L$ above, the first $|L|$ rows of $T^{(k)}$ yield redundant equations in Eq. (S.20). Therefore, we can define $T'^{(k)}$ ($\boldsymbol{\alpha}'$) as $T^{(k)}$ ($\boldsymbol{\alpha}$) with all rows $j \in L \setminus \{1\}$ eliminated and define the new system of equations, which we call System A' :

$$T'^{(k)} \boldsymbol{p}'^{(k)} = \boldsymbol{\alpha}' / \alpha_1, \quad (\text{S.23})$$

$$\boldsymbol{p}'^{(k)} \geq 0. \quad (\text{S.24})$$

System A has a solution if and only if System A' does. Most importantly, System A' is identical to System A in the case $|L| = 1$, which is the precise case we proved in the main text. Therefore, once we restrict to System A' , the proof follows exactly as in the main text.

II. REDERIVING THE PROTOCOL IN REF. [32]

In this section, we prove the statement in the main text that we can use an algebraic method to reproduce the protocol from Ref. [32] in any dimension d . Let

$$\bar{T} = \begin{pmatrix} | & | & \cdots & | \\ \text{sgn}(\boldsymbol{\alpha}^{(1)}) & \text{sgn}(\boldsymbol{\alpha}^{(2)}) & \cdots & \text{sgn}(\boldsymbol{\alpha}^{(d)}) \\ | & | & \cdots & | \end{pmatrix}, \quad (\text{S.25})$$

where we define $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}$, the vector determining the function we want to measure, and

$$(\boldsymbol{\alpha}^{(k>1)})_i \equiv \begin{cases} \alpha_i & i < k \\ -\alpha_i & i \geq k \end{cases}. \quad (\text{S.26})$$

These definitions reproduce the states used in the optimal protocol described in Ref. [32]. We will prove that

$$\bar{T}^{-1} = \frac{1}{2} \begin{pmatrix} - & \text{sgn}(\boldsymbol{\gamma}^{(1)}) & - \\ - & \text{sgn}(\boldsymbol{\gamma}^{(2)}) & - \\ \vdots & \vdots & \vdots \\ - & \text{sgn}(\boldsymbol{\gamma}^{(d)}) & - \end{pmatrix}, \quad (\text{S.27})$$

where

$$(\boldsymbol{\gamma}^{(k)})_i \equiv \begin{cases} (\boldsymbol{\alpha}^{(k)})_i & i = k - 1, k, \\ 0 & \text{else} \end{cases}, \quad (\text{S.28})$$

with $i = 0$ corresponding to $i = d$ in the case that $k = 1$. As an example, if $\text{sgn}(\boldsymbol{\alpha}) = (1, 1, -1, -1)$, then we have that

$$\bar{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \end{pmatrix}, \bar{T}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \boldsymbol{p} = \frac{1}{2} \begin{pmatrix} \alpha_1 - \alpha_4 \\ \alpha_1 - \alpha_2 \\ \alpha_2 + \alpha_3 \\ -\alpha_3 + \alpha_4 \end{pmatrix}, \quad (\text{S.29})$$

where \boldsymbol{p} is determined by simply inverting the equation $\bar{T}\boldsymbol{p} = \boldsymbol{\alpha}$, assuming $\alpha_1 = 1$. Note also that the \boldsymbol{p} above is actually non-negative assuming $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4$ (as usual), and that their signs are $(1, 1, -1, -1)$.

Returning to the general case, we show that Eq. (S.27) is indeed the inverse of Eq. (S.25). We have that

$$(\bar{T}^{-1}\bar{T})_{ij} = \sum_{k=1}^d \bar{T}_{ik}^{-1}\bar{T}_{kj} = \frac{1}{2} \sum_k \text{sgn}(\boldsymbol{\gamma}^{(i)})_k \text{sgn}(\boldsymbol{\alpha}^{(j)})_k = \frac{1}{2} \text{sgn}(\boldsymbol{\gamma}^{(i)}) \cdot \text{sgn}(\boldsymbol{\alpha}^{(j)}) \quad (\text{S.30})$$

$$= \frac{1}{2} [\text{sgn}(\boldsymbol{\alpha}^{(i)})_{i-1} \text{sgn}(\boldsymbol{\alpha}^{(j)})_{i-1} + \text{sgn}(\boldsymbol{\alpha}^{(i)})_i \text{sgn}(\boldsymbol{\alpha}^{(j)})_i]. \quad (\text{S.31})$$

If $i = j$, then this is clearly equal to 1 because $\text{sgn}^2(x) = 1$ for $x \neq 0$ (we assume without loss of generality that $\boldsymbol{\alpha}$ contains no zeroes, as these parameters would be completely decoupled from our protocol and can thus be safely ignored). According to Eq. (S.26), $\text{sgn}(\boldsymbol{\alpha}^{(i)})_{i-1} = \text{sgn}(\boldsymbol{\alpha})_{i-1}$ but $\text{sgn}(\boldsymbol{\alpha}^{(i)})_i = -\text{sgn}(\boldsymbol{\alpha})_i$. Consider the case $j > i > 1$ first. $j > i > i - 1$ means that $\text{sgn}(\boldsymbol{\alpha}^{(j)})_i = \text{sgn}(\boldsymbol{\alpha})_i$ and $\text{sgn}(\boldsymbol{\alpha}^{(j)})_{i-1} = \text{sgn}(\boldsymbol{\alpha})_{i-1}$. Thus, the contributions cancel out and we get zero, as desired. If $j > i = 1$, then we have

$$(\bar{T}^{-1}\bar{T})_{ij} = \sum_{k=1}^d \bar{T}_{ik}^{-1}\bar{T}_{kj} = \frac{1}{2} [\text{sgn}(\boldsymbol{\alpha}^{(1)})_d \text{sgn}(\boldsymbol{\alpha}^{(j)})_d + \text{sgn}(\boldsymbol{\alpha}^{(1)})_1 \text{sgn}(\boldsymbol{\alpha}^{(j)})_1] \quad (\text{S.32})$$

$$= \frac{1}{2} [-\text{sgn}(\alpha_d) \text{sgn}(\alpha_d) + \text{sgn}(\alpha_1) \text{sgn}(\alpha_1)] \quad (\text{S.33})$$

$$= 0 \quad (\text{S.34})$$

Notice how the term that contributes the negative sign flips in this case. A similar argument holds for $j < i$.

Having shown \bar{T}^{-1} is indeed the inverse of \bar{T} , we just need to determine \boldsymbol{p} . First consider $i > 1$.

$$\begin{aligned} p_i = (\bar{T}^{-1}\boldsymbol{\alpha})_i &= \frac{1}{2} \text{sgn}(\boldsymbol{\gamma}^{(i)}) \cdot \boldsymbol{\alpha} = \frac{1}{2} \text{sgn}(\boldsymbol{\alpha}^{(i)})_{i-1} \alpha_{i-1} + \frac{1}{2} \text{sgn}(\boldsymbol{\alpha}^{(i)})_i \alpha_i \\ &= \frac{1}{2} \text{sgn}(\boldsymbol{\alpha})_{i-1} \alpha_{i-1} - \frac{1}{2} \text{sgn}(\boldsymbol{\alpha})_i \alpha_i = \frac{1}{2} |\alpha_{i-1}| - \frac{1}{2} |\alpha_i|, \end{aligned} \quad (\text{S.35})$$

which exactly matches the probabilities given in Ref. [32] and our example above. For $i = 1$, we have to be a little bit more careful because $\boldsymbol{\gamma}^{(1)}$ “wraps around” the matrix.

$$p_1 = (\bar{T}^{-1}\boldsymbol{\alpha})_1 = \frac{1}{2} \text{sgn}(\boldsymbol{\gamma}^{(1)}) \cdot \boldsymbol{\alpha} = \frac{1}{2} \text{sgn}(\boldsymbol{\alpha})_d \alpha_d + \frac{1}{2} \text{sgn}(\boldsymbol{\alpha})_1 \alpha_1 = \frac{1}{2} |\alpha_1| + \frac{1}{2} |\alpha_d|, \quad (\text{S.36})$$

which again matches Ref. [32] and our example above.

III. DERIVING A LESS ENTANGLED PROTOCOL

In this section, we demonstrate a protocol that differs crucially from that in Ref. [32] in that not every state used is maximally entangled, once again proving that such a protocol exists for all dimensions d . We again write

$$\bar{T} = \begin{pmatrix} | & | & \cdots & | \\ \text{sgn}(\boldsymbol{\alpha}^{(1)}) & \text{sgn}(\boldsymbol{\alpha}^{(2)}) & \cdots & \text{sgn}(\boldsymbol{\alpha}^{(d)}) \\ | & | & \cdots & | \end{pmatrix}, \quad (\text{S.37})$$

where we again define $\boldsymbol{\alpha}^{(1)} \equiv \boldsymbol{\alpha}$. However, this time, we define

$$(\boldsymbol{\alpha}^{(k>1)})_i \equiv \begin{cases} \alpha_i & i < k \\ 0 & i \geq k \end{cases}. \quad (\text{S.38})$$

Note that only $\boldsymbol{\alpha}^{(1)}$ represents a maximally entangled state. We will prove that

$$\bar{T}^{-1} = \begin{pmatrix} - & \text{sgn}(\boldsymbol{\delta}^{(1)}) & - \\ - & \text{sgn}(\boldsymbol{\delta}^{(2)}) & - \\ \vdots & \vdots & \vdots \\ - & \text{sgn}(\boldsymbol{\delta}^{(d)}) & - \end{pmatrix}, \quad (\text{S.39})$$

where we define

$$(\boldsymbol{\delta}^{(k>1)})_i \equiv \begin{cases} \alpha_i & i = k - 1, \\ -\alpha_i & i = k, \\ 0 & \text{else} \end{cases}, \quad (\text{S.40})$$

and

$$(\boldsymbol{\delta}^{(1)})_i \equiv \begin{cases} \alpha_d & i = d, \\ 0 & \text{else} \end{cases}. \quad (\text{S.41})$$

As an example, if $\text{sgn}(\boldsymbol{\alpha}) = (1, 1, -1, -1)$ again, we have in this case that

$$\bar{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \bar{T}^{-1} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \boldsymbol{p} = \begin{pmatrix} -\alpha_4 \\ \alpha_1 - \alpha_2 \\ \alpha_2 + \alpha_3 \\ -\alpha_3 + \alpha_4 \end{pmatrix}. \quad (\text{S.42})$$

Observe the similarities and differences between the results in Eq. (S.29) and Eq. (S.42). In particular, note the remarkable similarity between the two $\bar{T}^{-1}, \boldsymbol{p}$. We can again calculate

$$(\bar{T}^{-1}\bar{T})_{ij} = \sum_{k=1}^d \bar{T}_{ik}^{-1}\bar{T}_{kj} = \sum_k \text{sgn}(\boldsymbol{\delta}^{(i)})_k \text{sgn}(\boldsymbol{\alpha}^{(j)})_k = \text{sgn}(\boldsymbol{\delta}^{(i)}) \cdot \text{sgn}(\boldsymbol{\alpha}^{(j)}). \quad (\text{S.43})$$

We proceed with a case-wise analysis. We start with the edge cases where one or both of i, j are equal to 1. If we take $i = j = 1$, then this reduces to $\text{sgn}(\boldsymbol{\delta}^{(1)}) \cdot \text{sgn}(\boldsymbol{\alpha}^{(1)}) = \text{sgn}(\alpha_d)^2 = 1$. If $1 = i < j$, then this becomes $\text{sgn}(\boldsymbol{\delta}^{(1)}) \cdot \text{sgn}(\boldsymbol{\alpha}^{(j)}) = \text{sgn}(\alpha_d)(\text{sgn}(\alpha^{(j)})_d) = 0$, as $j \leq d$ for all j . If $1 = j < i$, then we get $\text{sgn}(\boldsymbol{\delta}^{(i)}) \cdot \text{sgn}(\boldsymbol{\alpha}) = \text{sgn}(\alpha_{i-1})^2 - \text{sgn}(\alpha_i)^2 = 0$.

Now we move on to the case where $i, j \neq 1$ such that

$$\text{sgn}(\boldsymbol{\delta}^{(i)}) \cdot \text{sgn}(\boldsymbol{\alpha}^{(j)}) = \text{sgn}(\boldsymbol{\alpha})_{i-1} \text{sgn}(\boldsymbol{\alpha}^{(j)})_{i-1} - \text{sgn}(\boldsymbol{\alpha})_i \text{sgn}(\boldsymbol{\alpha}^{(j)})_i. \quad (\text{S.44})$$

First consider $j = i$. Then the first term on the RHS is 1, and the second term vanishes according to Eq. (S.38). If $i < j$, then both terms are equal to 1, as $j > i$ implies that $(\boldsymbol{\alpha}^{(j)})_i = \alpha_i$ and $(\boldsymbol{\alpha}^{(j)})_{i-1} = \alpha_{i-1}$ according to Eq. (S.38). If $j < i$, then $j \leq i - 1$, which means that both terms on the RHS vanish because.

So, with \bar{T}^{-1} again determined, we just need to calculate \boldsymbol{p} . First consider $i > 1$.

$$p_i = (\bar{T}^{-1}\boldsymbol{\alpha})_i = \text{sgn}(\boldsymbol{\delta}^{(i)}) \cdot \boldsymbol{\alpha} = \text{sgn}(\boldsymbol{\alpha})_{i-1}\alpha_{i-1} - \text{sgn}(\boldsymbol{\alpha})_i\alpha_i = |\alpha_{i-1}| - |\alpha_i| \quad (\text{S.45})$$

which matches our example above. For $i = 1$, we again have to be a little bit more careful.

$$p_1 = (\bar{T}^{-1}\boldsymbol{\alpha})_1 = \text{sgn}(\boldsymbol{\delta}^{(1)}) \cdot \boldsymbol{\alpha} = \text{sgn}(\boldsymbol{\alpha})_d\alpha_d = |\alpha_d|, \quad (\text{S.46})$$

which also matches Eq. (S.42).

IV. EXAMPLE OF THEOREM 1 FOR THREE QUBITS

In this section, we provide a detailed example of Theorem 1 of the main text for the simplest non-trivial case, which occurs for three qubits, when $|\alpha_1| > |\alpha_2|$. Motivating this constraint on $|\alpha_2|$, we note if $|\alpha_2| = |\alpha_1|$, 3-partite entanglement is always needed except in the trivial case that $|\alpha_3| = 0$. For simplicity, we also assume $|\alpha_2| \geq |\alpha_3|$. We first write down the T matrix

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 \end{pmatrix}. \quad (\text{S.47})$$

We claim that the matrix $T^{(2)}$, which is T restricted to columns with at most two non-zero components, admits at least one non-negative vector \mathbf{p} such that $T^{(2)}\mathbf{p} = \boldsymbol{\alpha}/\alpha_1$ as long as $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty \leq 2$, but not if $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty > 2$. In what follows, we will assume without loss of generality that $\alpha_1 = 1$ such that the condition for a non-negative solution becomes whether $|\alpha_2| + |\alpha_3| \leq 1$ or not. We have that

$$T^{(2)} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix}. \quad (\text{S.48})$$

We can show exhaustively that whenever $|\alpha_2| + |\alpha_3| \leq 1$, there exist a choice of three linearly independent vectors that compose a matrix $\bar{T}^{(2)}$ such that $\mathbf{p} = (\bar{T}^{(2)})^{-1}\boldsymbol{\alpha}$ is non-negative. Indeed:

$$\bar{T}^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \implies \mathbf{p} = \begin{pmatrix} \alpha_2 \\ \alpha_3 \\ 1 - \alpha_2 - \alpha_3 \end{pmatrix}, \quad (\text{S.49})$$

$$\bar{T}^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \implies \mathbf{p} = \begin{pmatrix} \alpha_2 \\ -\alpha_3 \\ 1 - \alpha_2 + \alpha_3 \end{pmatrix}, \quad (\text{S.50})$$

$$\bar{T}^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \implies \mathbf{p} = \begin{pmatrix} -\alpha_2 \\ \alpha_3 \\ 1 + \alpha_2 - \alpha_3 \end{pmatrix}, \quad (\text{S.51})$$

$$\bar{T}^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \implies \mathbf{p} = \begin{pmatrix} -\alpha_3 \\ -\alpha_2 \\ 1 + \alpha_2 + \alpha_3 \end{pmatrix}. \quad (\text{S.52})$$

It is manifestly clear that the first solution works for $\alpha_2 \geq \alpha_3 \geq 0$, the second for $\alpha_2 \geq -\alpha_3 \geq 0$, the third for $-\alpha_2 \geq \alpha_3 \geq 0$, and the fourth for $-\alpha_2 \geq -\alpha_3 \geq 0$.

We can also prove that if $|\alpha_2| + |\alpha_3| > 1$, there is no non-negative solution to $T^{(2)}\mathbf{p} = \boldsymbol{\alpha}$.

$$T^{(2)}\mathbf{p} = \boldsymbol{\alpha} \implies \begin{cases} p_1 + p_2 + p_3 + p_4 + p_5 & = 1 \\ p_1 - p_5 & = \alpha_2 \\ p_2 - p_4 & = \alpha_3 \end{cases}. \quad (\text{S.53})$$

We separate into four cases: (1) $\alpha_2 \geq \alpha_3 \geq 0$, (2) $\alpha_2 \geq -\alpha_3 \geq 0$, (3) $-\alpha_2 \geq \alpha_3 \geq 0$, (4) $-\alpha_2 \geq -\alpha_3 \geq 0$. In each case, we either add or subtract the second and third equations from the first to derive a contradiction with the fact that each $p_i \geq 0$. For example, in case (1), we simply subtract the second and third equations to get that $p_3 + 2p_4 + 2p_5 = 1 - \alpha_2 - \alpha_3 < 0$. In case (2), we subtract the second and add the third to get that $2p_2 + p_3 + 2p_5 = 1 - (\alpha_2 - \alpha_3) < 0$. In case (3), we add the second and subtract the third, and, in case (4), we subtract both. Thus, there is no non-negative solution, as is consistent with Theorem 1.