

# Multiple-phase quantum interferometry—real and apparent gains of measuring all the phases simultaneously

Wojciech Górecki<sup>1</sup> and Rafał Demkowicz-Dobrzański<sup>1</sup>

<sup>1</sup>Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

We characterize operationally meaningful quantum gains in a paradigmatic model of lossless multiple-phase interferometry and stress insufficiency of the analysis based solely on the concept of quantum Fisher information. We show that the advantage of the optimal simultaneous estimation scheme amounts to a constant factor improvement when compared with schemes where each phase is estimated separately—contrary to a widely cited results claiming a better precision scaling in terms of the number of phases involved.

*Introduction and the summary of results.* Quantum metrology aims at identifying optimal ways of utilizing quantum systems as sensing probes [1–9]. When  $N$  quantum probes are used independently, estimation variance decreases inversely proportional to the number of probes—the standard quantum limit (SQL). The hallmark of quantum metrology is the potential quadratic scaling improvement over the SQL—the Heisenberg limit (HL) [10–20].

In *multiple-parameter* estimation scenarios, simultaneous estimation of  $p$  parameters in a single experiment may additionally provide an improved performance when compared with strategies where each parameter is estimated separately [21–27].

A paradigmatic model to study the potential of multi-parameter quantum enhanced metrological protocols is the multiple-phase estimation problem. The goal is to estimate all the relative phase shifts in a multiple-arm interferometer with the best precision possible given a constraint on the total number of photons used, see Fig. 1.

The most common tool to analyze the potential of quantum metrological strategies is the Quantum Fisher Information (QFI), inverse of which lower-bounds the variance of any locally unbiased estimator  $\tilde{\theta}$  via the famous quantum Cramer-Rao (CR) bound [28–30]. In the single parameter case the CR bound takes the form:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{kF(\rho_\theta^n)}, \quad (1)$$

where  $k$  is the number of repetitions of an experiment, while  $F(\rho_\theta^n)$  is the QFI computed on the  $n$ -probe output state on which the parameter  $\theta$  has been encoded. In case of the standard two-arm optical interferometry, a single relative phase between the two arms is being estimated, and  $F(\rho_\theta^n) = n$  for  $n$  uncorrelated photons sent into the interferometer, while the maximum value  $F(\rho_\theta^n) = n^2$  is obtained for optimally entangled state of  $n$  photons—the  $n00n$  state—resulting in the  $1/n^2$  Heisenberg scaling (HS) of precision [5, 12–19]. In general, this bound is operationally saturable provided one takes the asymptotic limit  $k \rightarrow \infty$ , while keeping the  $n$  fixed. Such a case corresponds to an experimental realization, where the amount of resources used in a single realization is large but limited, while experiment may be repeated arbitrary number of times.

However, a fundamental question is what is the true HL for precision if the total amount of resources  $N = n \cdot k$  is restricted and the  $N \rightarrow \infty$  limit is taken. Since the scaling of

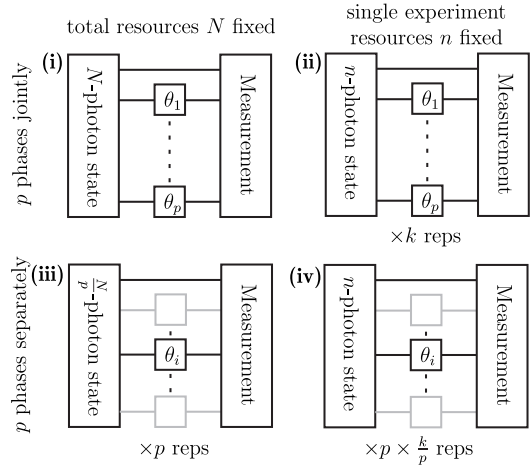


FIG. 1. Multiple-phase estimation schemes, where a constraint is imposed on the total resources used (left column) or resources used in a single experiment (right column). The top and bottom rows represent respectively the protocols where all the phases are measured jointly or the estimation procedure is repeated for each of the phases separately.

$\Delta^2 \tilde{\theta} \geq$	Independent probes SQL	Entangled probes	
		$N = kn$ probes HL	$n$ probes, $k$ reps HS
single phase	$\frac{1}{N}$	$\frac{\pi^2}{N^2}$	$\frac{1}{kn^2}$
$p$ -phases jointly	$\frac{p^2}{4N}$	(i) $\frac{c p^3}{N^2}$	(ii) $\frac{p^2}{4kn^2}$
$p$ -phases separately	$\frac{p^2}{N}$	(iii) $\frac{\pi^2 p^3}{N^2}$	(iv) $\frac{p^2}{kn^2}$

TABLE I. Asymptotically achievable lower bounds on the sum of variances of estimated phases. The main result of the paper is the bolded formula representing the proper multiple-phase Heisenberg limit (HL) and demonstrating the  $p^3$  scaling with the number of estimated phases. All bounds are tight in the asymptotic limit  $N \rightarrow \infty$  (or  $k \rightarrow \infty$ ) while in case of joint phases estimation tightness requires an additional  $p \rightarrow \infty$  limit ( $c = 1.89$  yields a universally valid bound, while  $c = 2$  yields an asymptotically achievable cost). The SQL column can be regarded as a special case of the HS column when  $n = 1$ , and  $k = N$ .

precision is quadratic in  $n$  and linear in  $k$ , the optimal choice

appears to be  $n = N, k = 1$ . However, in this case one can not use saturability arguments based on the many-repetition scenario and the predictions of the QFI may be misleading with respect to the choice of the optimal probe states as well as the asymptotically achievable precision limit. This becomes clear when the results are contrasted with the ones obtained the minimax [31], Bayesian [32] or information theoretic approach [33].

The use of a single N00N state (note  $N$  instead of  $n$  is intentional) is clearly not an operationally meaningful strategy when discussing the HL, as it is not capable of discriminating phases that differ by a multiple of  $2\pi/N$ . There is clearly a need to sacrifice part of the resources to get rid of the arising ambiguity, and this leads to a  $\pi^2$  increase in the asymptotically saturable bound, which can be rigorously shown within a Bayesian estimation framework [32, 34–39]. Therefore, in order to avoid confusion, we will introduce a clear distinction between the two approaches and refer to them as

- (HS) the amount resources used in single repetition of experiment  $n$  large but finite, and whole experiment may be repeated  $k \rightarrow \infty$  times:  $\Delta^2 \tilde{\theta} \propto \frac{1}{kn^2}$
- (HL) the total amount of resources  $N$  is restricted, and no repetitions of an experiment are assumed:  $\Delta^2 \tilde{\theta} \propto \frac{1}{N^2}$ .

The two approaches may only be reconciled provided one is able to properly account for the scaling of the required number of repetitions  $k$  with the increasing number of probes  $n$  used in a single experiment, that guarantees saturability within the HS approach. This is, however, hardly ever possible and typically the issue is simply ignored in the literature.

In the multiparameter case a rigorous study of the achievable HL is much more challenging and the common approach is to rather work in the HS paradigm where efficiently computable multiparameter generalizations of CR bounds are used [23, 27, 40–45]. Rigorous analysis of the actual saturable HL are typically restricted to Bayesian framework case studies utilizing some underlying group structure of the problem [46–53]. However, quite surprisingly, the actual analytical form of the asymptotically saturable HL for the paradigmatic multiple phase estimation problem is missing (see [54] for a recent numerical attempt to tackle the problem).

In this paper, we employ an operationally meaningful minimax approach to derive an asymptotically saturable HL for the multiple-phase estimation problem and demonstrate that it manifests a  $p^3$  scaling, with the number of parameters involved, rather than the  $p^2$  that is advocated when following the HS approach [55]. We also clarify apparent gains that may be obtained thanks to simultaneous phase estimation, when compared with strategies that estimate all the phases separately. We show that the advantage amounts to a constant factor gain and, contrary to the claims of [55–57], does not lead to a better scaling of precision with the number of parameters involved. We explain the apparent contradiction by pointing out the improper use of saturability arguments that are often invoked when following the HS approach—Tab. I summarizes the main results of this paper.

*Multiple-phase estimation problem.* Consider a multiple-phase estimation problem as depicted in Fig. 1, where the goal is to estimate the value of  $p$ -phases  $\theta = [\theta_1, \dots, \theta_p]$ —relative phase delays in the  $i$ -th arm of an interferometer with respect to the reference arm. For a general  $n$ -photon state at the input, the output state with phase information encoded will have the form:

$$|\Psi_\theta^n\rangle = \sum_{\mathbf{m}: \sum_{i=1}^p m_i \leq n} c_{\mathbf{m}} e^{i\mathbf{m} \cdot \theta} |m_1, m_2, \dots, m_p\rangle, \quad (2)$$

with  $\mathbf{m} = [m_1, \dots, m_p]$ , where  $m_i$  represents the number of photons in the  $i$ -th ‘signal’ arm, while remaining  $n - \sum_{i=1}^p m_i$  photons are sent through reference arm. A general quantum measurement is then performed in order to extract information on the encoded phases, mathematically specified by a set of positive operators  $\{M_x\}$ ,  $\sum_x M_x = \mathbb{1}$ , where  $x$  labels a measurement outcome observed with probability  $p_\theta(x) = \langle \Psi_\theta^n | M_x | \Psi_\theta^n \rangle$ . The measurement outcome(s) is then fed into an estimator function which yields the inferred values of the phases. In the scenario (i), the estimator  $\tilde{\theta}(x)$  is a function of just a single measurement outcome (in this case  $n = N$  as all resources are used in a single shot); in (ii) the experiment is repeated  $k$  times and the estimator is a function of all  $k$  measurement outcomes  $\tilde{\theta}(x^1, \dots, x^k)$ ; in (iii)  $p$  separate protocols, involving  $p$  different states (each containing  $N/p$  photons) and different measurements, are performed yielding measurement outcomes  $x_1, \dots, x_p$  where each outcome  $x_i$  feeds the estimator of the  $i$ -th phase  $\tilde{\theta}_i(x_i)$ ; finally, in (iv) each of the  $p$  separate protocol is repeated  $k/p$  times, yielding in total  $p \times k/p = k$  measurement outcomes and resulting in  $p$  separate estimators of each phase  $\tilde{\theta}_i(x_i^1, \dots, x_i^{k/p})$ ,  $i \in \{1, \dots, p\}$ .

Respectively of which scenario is considered, the figure of merit to be minimized is the sum of squared errors of estimated phases

$$\Delta^2 \tilde{\theta} = \int d\mathbf{x} p_\theta(\mathbf{x}) (\tilde{\theta}(\mathbf{x}) - \theta)^2, \quad (3)$$

where  $\int d\mathbf{x}$  formally represents integration over all (possibly continuous) measurement outcomes and  $(\tilde{\theta}(\mathbf{x}) - \theta)^2 = \sum_{i=1}^p (\tilde{\theta}_i(\mathbf{x}) - \theta_i)^2$ . As this is a point-wise figure of merit (calculated at a given  $\theta$ ), in order to make the minimization task meaningful one needs to impose additional constraints on the estimator function—otherwise a trivial solution  $\tilde{\theta}(\mathbf{x}) = \theta$  yields zero cost.

The most common one is the locally unbiasedness condition, which is also the key assumption behind the derivation of the CR-type bounds [28–30, 45, 57]. This assumption itself may not be sufficient to obtain an operationally saturable bounds, as in principle the region where the use of the local-unbiased estimator makes sense may shrink when taking the asymptotic limit  $N \rightarrow \infty$  [31–33].

Alternatively, one may follow the so called *minimax* approach, and define a region  $\Theta$  inside which the true value of  $\theta$  is guaranteed to be and then consider the estimator which gives the best results in the most pessimistic scenario, i.e.

which minimizes the cost maximized over all  $\theta \in \Theta$ :

$$\Delta^2 \tilde{\theta}_{\text{minimax}} \equiv \inf_{M_x, \tilde{\theta}(x)} \sup_{\theta \in \Theta} \int d\mathbf{x} p_\theta(x) (\tilde{\theta}(x) - \theta)^2. \quad (4)$$

The advantage of the approach is that  $\Theta$  is fixed while taking the asymptotic limit  $N \rightarrow \infty$  and hence no region shrinking issues arise. We now proceed to derive an asymptotically saturable lower bound on the above cost in the most fundamental scenario (i) and then contrast it with scenarios (ii), (iii), (iv).

*Derivation of the multiple-phase HL.* Below we present the sketch of the proof. For more formal derivation see Supplemental Material (SM) A [29, 32, 52, 58–65]. First, we consider an extension of the model by replacing discrete variables  $m_i \in \{0, 1, \dots, N\}$  with continuous ones  $\frac{m_i}{N} \rightarrow \mu_i \in [0, 1]$ , and the sums by the respective integrals. Note that such an extension may only decrease the minimal achievable cost, as the discrete model may be always arbitrary well approximated as a special case of the continuous model. The probe-state is now characterized by a  $p$ -dimensional wave function  $f(\boldsymbol{\mu})$ :

$$|\Psi_{f, \theta}^N\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} f(\boldsymbol{\mu}) e^{iN\boldsymbol{\mu}\theta} |\mu_1, \mu_2, \dots, \mu_p\rangle. \quad (5)$$

Next, as we argue in detail in the SM, the asymptotic bound for any finite region  $\Theta$  is equivalent, up to the leading  $1/N^2$  order, to the cost when the region is unbounded  $\Theta = \mathbb{R}^p$ . In the latter case the problem is covariant with respect to the translation group and the optimal measurement can be restricted to the class of covariant measurements [29] (thanks to the generalization of the Hunt-Stein lemma [59, 60] for non-compact groups [52, 61, 62]) and without loss of generality may be chosen to be the momentum projection measurement,  $M_{\tilde{\theta}} = |\chi_{\tilde{\theta}}\rangle\langle\chi_{\tilde{\theta}}|$ , where

$$|\chi_{\tilde{\theta}}\rangle = \frac{1}{\sqrt{(2\pi/N)^p}} \int d\boldsymbol{\mu} e^{iN\boldsymbol{\mu}\tilde{\theta}} |\boldsymbol{\mu}\rangle. \quad (6)$$

Note that we have implicitly replaced the measurement outcomes  $x$  with the actual estimated values  $\tilde{\theta}$ . Minimization of the resulting lower bound over the probe state wave function  $f(\boldsymbol{\theta})$  leads to the following lower bound on the cost  $\Delta^2 \tilde{\theta}$ :

$$\min_f \int_{\mathbb{R}^p} d\tilde{\theta} |\langle\chi_{\tilde{\theta}}|\Psi_{f, \theta}^N\rangle|^2 (\tilde{\theta} - \theta)^2 = \frac{1}{N^2} \min_f \int_{\mathbb{R}^p} d\tilde{\theta} |\hat{f}(\tilde{\theta})|^2 \tilde{\theta}^2 \quad (7)$$

where  $\hat{f}$  is the Fourier transform of  $f$  and we dropped the irrelevant dependence on  $\theta$ . Going back to the  $\boldsymbol{\mu}$ -representation, the minimization problem takes the following form:

$$\frac{1}{N^2} \min_f \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} f^*(\boldsymbol{\mu}) \left( \sum_{k=1}^p -\partial_{\mu_k}^2 \right) f(\boldsymbol{\mu}),$$

$$\text{with } \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} |f(\boldsymbol{\mu})|^2 = 1,$$

$$f(\boldsymbol{\mu}) = 0 \quad \text{for } \boldsymbol{\mu} \text{ on the boundary } (\mu_i = 0 \vee \sum_i \mu_i = 1). \quad (8)$$

This problem is therefore equivalent to identifying the ground state energy of a quantum particle in a  $p$ -dimensional simplex-shaped infinite potential well, which in general has not known analytical solution (apart from specific cases [66–68]). Still, it may be easily bounded from below in following way.

Since the problem enjoys an inherent symmetry with respect to permuting the  $p$  ‘phase’ arms (the reference arm is distinguished by the choice of the cost function), and the total number of photons in  $p$  ‘phase’ arms is  $\leq N$ , the expectation value of the number of photons in each single ‘phase arm’ is  $\leq N/p$ . Now we will neglect the fact, that the distribution of photons in each single arm comes from the multi-arm distribution of  $N$  photons and keep only the constraint on the photon expectation value. Such a constraint is a weaker one than the original one, but as it refers just to a single ‘phase’ mode it allows for an effective separation of variables. This allows us to lower bound the total cost by  $p$  times the minimal single-phase estimation cost given mean number of photons in the mode  $N/p$ :

$$\Delta^2 \tilde{\theta} \geq p \times \frac{1}{N^2} \min_g \int_0^\infty d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial \mu^2} \right) g(\mu) \quad (9)$$

with constraints:

$$g(0) = 0, \quad \int_0^\infty d\mu |g(\mu)|^2 = 1, \quad \int_0^\infty d\mu |g(\mu)|^2 N\mu = N/p. \quad (10)$$

The single mode problem may be solved using the standard Lagrange multiplier method, and we obtain the solution  $g(\mu)$  to be the Airy function  $\text{Ai}(\cdot)$  (see also [69, 70], where the same solution appeared in a single phase estimation context), leading to the final bound

$$\Delta^2 \tilde{\theta} \geq \frac{p^3}{N^2} \frac{4|A_0|^3}{27} \approx \frac{1.89p^3}{N^2}, \quad (11)$$

where  $A_0 \approx -2.34$  is the first zero of  $\text{Ai}(\cdot)$ . The most important feature of the bound is the  $p^3$  scaling. This bound is valid, even if one considers the most general adaptive strategy with arbitrary large ancilla is allowed, see SM A.

Note, that an analogue reasoning could not be performed to bound the QFI, as the QFI may be arbitrary large when only the constraint on the *mean* (and not the maximal) number of photons in the sensing arm is imposed, and leads to some operationally unjustified claims of sub-Heisenberg estimation strategies [71, 72], as discussed in [73].

*Comparison of different approaches.* When following the (ii) approach and minimizing the trace of the inverse of the QFI matrix of the output state, one obtains the following bound on the total cost arising from the application of the multi-parameter version of the CR bound [55] (see also [23] for justification of fundamental optimality):

$$\Delta^2 \tilde{\theta} \geq \frac{1}{k} \min_{|\Psi^n\rangle} \text{Tr}[F^{-1}(|\Psi_\theta^n\rangle)] = \frac{(1 + \sqrt{p})^2 p}{4kn^2} \stackrel{p \gg 1}{\approx} \frac{p^2}{4kn^2}, \quad (12)$$

where the optimal input state has the form

$$|\Psi^n\rangle = \beta |n, 0, \dots, 0\rangle + \alpha (|0, n, \dots, 0\rangle + \dots |0, 0, \dots, n\rangle), \quad (13)$$

with  $\alpha = 1/\sqrt{p + \sqrt{p}}, \beta = 1/\sqrt{1 + \sqrt{p}}$ . The most visible difference between the two approaches is the scaling of the cost with the number of parameters estimated,  $p^3$  in (i) vs  $p^2$  in (ii). In order to avoid contradiction, this implies that when considering  $k$  repetitions in the (ii) scenario, the actual number of repetitions required to saturate the CR bound will in fact need to increase at least linearly with  $p$ . This fact lies at the heart of the discrepancy between the claims of [55] and ours. Interestingly, when considering the Gaussian states only, the QFI based study [74] yields results qualitatively equivalent to ours ( $p^3$  cost scaling for both joint and separate strategies) which should be attributed to the fact that the saturability of the CR-type bounds in Gaussian models is guaranteed already at the single shot level without invoking the multiple repetition argument [29, 45].

In scenario (iii), one separately sends  $N/p$  photons states, into the  $i$ -th and the reference arm, in order to measure a given  $\theta_i$  phase, using the optimal state for sensing a single completely unknown phase [34–36]:

$$|\Psi_i^{N/p}\rangle = \sqrt{\frac{2}{N/p+2}} \sum_{m=0}^{N/p} \sin\left[\frac{(m+1)\pi}{N/p+2}\right] |N/p-m\rangle_0 |m\rangle_i, \quad (14)$$

where  $|N/p-m\rangle_0 |m\rangle_i$  denotes a state where  $m$  photons is sent into the  $i$ -th arm and  $N/p-m$  into the reference arm. The resulting bound on the total variance is therefore lower bounded by  $p$  times the single phase estimation variance [34–36]:

$$\Delta^2 \tilde{\theta} \stackrel{N/p \gg 1}{\gtrsim} p \times \frac{\pi^2}{(N/p)^2} = \frac{p^3 \pi^2}{N^2}. \quad (15)$$

We see the same scaling as in the joint phase estimation protocol (i), which implies that the largest possible gain coming from joint phase estimation amounts to a constant factor  $\leq \pi^2/1.89$ . In order to show that the gain over the separate strategy is indeed achievable, we need to find a state wavefunction  $f(\boldsymbol{\mu})$  that manifests an advantage over the separate strategy, when plugged into the joint estimation cost formula Eq. (8). We propose a simple ansatz for the structure of the state that satisfies the boundary conditions

$$f(\boldsymbol{\mu}) \propto \left( \prod_{i=1}^p \mu_i \right)^\alpha \left( 1 - \sum_{i=1}^p \mu_i \right)^\beta. \quad (16)$$

The minimal cost

$$\Delta^2 \tilde{\theta} = \frac{p(1+2\sqrt{p})^2 \sqrt{p}(4p+2\sqrt{p}-1)}{(8\sqrt{p}-4)N^2} \stackrel{p \gg 1}{\approx} \frac{2p^3}{N^2}. \quad (17)$$

is obtained for  $\alpha = 3/2, \beta = \sqrt{p}$ . For large  $p$  the cost approaches closely the fundamental bound (2 vs. 1.89 coefficient), demonstrating that the  $\pi^2/2 \approx 4.93$  advantage of joint phase estimation over separate strategies is achievable. Note that although this result was obtained for the problem with continuous variables  $\mu_i$ , it may be arbitrary well approximated within the original discrete model Eq. (2) with increasing  $N$  (in the same spirit as discussed in [75] for the single parameter

case). See also SM B [34–36, 75], for the details of computation, more discussion on the structure of the state and numerical investigation of the convergence of the discrete model to the continuous one when  $N$  is being increased.

Finally, the optimal strategy in (iv) is to use subsequently  $p$   $n$ 00n states

$$|\Psi_i^n\rangle = \frac{1}{\sqrt{2}} (|0\rangle_0 |n\rangle_i + |n\rangle_0 |0\rangle_i) \quad (18)$$

each designed to sense the  $i$ -th phase optimally. Since the total number of repetitions is  $k$ , each phase will be sensed  $k/p$  times and hence the final cost resulting from the application of the CR bound reads:

$$\Delta^2 \tilde{\theta} \geq p \times \frac{1}{k/p} \times \frac{1}{n^2} = \frac{p^2}{kn^2}. \quad (19)$$

Comparing this result with Eq. (12) we see that joint phase estimation offers again just a constant factor improvement over separate strategies. This result contrasts the claims of [55] where a scaling improvement ( $p^2$  vs  $p^3$ ) was claimed. Indeed, if  $n/p$  photons were used in a single phase estimation experiment, instead of considering  $k/p$  uses of the  $n$ -photon state, one would obtain the bound on the cost in the form:  $\Delta^2 \tilde{\theta} \geq p \times \frac{1}{k} \frac{1}{(n/p)^2} = \frac{p^3}{kn^2}$ . This latter calculation, however, does not reflect the cost of the optimal separate strategy in the framework that there is some fixed number of photons  $n$  used in a single experiment and the experiment is repeated  $k$  times. The optimal strategy is captured by the former reasoning, leading to the  $p^2$  scaling, as we can always regard this strategy as an equivalent  $k$  repetitions of an experiment using a mixed state  $\rho^n = \frac{1}{p} \sum_{i=1}^p |\Psi_i^n\rangle \langle \Psi_i^n|$ , where the factor  $\frac{1}{p}$  in formula for  $\rho^n$  express the fact, that in each repetition we measure only one parameter, with equal probability for each of them. Effectively  $k/p$  repetition for each parameter is performed, with  $n$  resources each time.

*Conclusions and discussion.* In this paper we have clarified the relation between formulas for the optimal cost in multiple-phase interferometry obtained within different paradigms, involving either fixing the total resources used or the resources used in a single experiment. Doing so, we have shown that within both paradigms joint phases estimation leads to at most constant factor improvement over the optimal separate strategies. This constant factor improvement may be attributed to the fact, that in the limit of many phases being sensed, the number of photons needed to be sent into the reference arm becomes negligible compared with the total number of photons used, whereas in the separate strategy it effectively consumes half of the resources available (see SM B for broader discussion). This claim remains valid also in the lossy optical interferometry case (where, however, HS does not occur) as shown in [76].

Note, that similar issues regarding the apparent scaling advantage of joint vs. separate parameter estimation may arise in other multiparameter estimation problems [22, 23] and in order to arrive at operationally meaningful conclusions, one should avoid implicit switching between (i-iv) paradigms and be aware of non-trivial saturability issues when following the QFI based approach.

## ACKNOWLEDGMENTS

We thank Dominic Berry for sharing his argument that we use in Supplemental Material A1, Francesco Albarelli for nu-

merous fruitful discussions, as well as Howard M. Wiseman, Animesh Datta and Luca Pezzè for useful comments. This work was supported by the National Science Center (Poland) grant No. 2016/22/E/ST2/00559.

- 
- [1] V. Giovannetti, S. Lloyd, and L. Maccone, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [2] M. G. A. Paris, *Int. J. Quantum Inf.* **07**, 125 (2009).
- [3] V. Giovannetti, S. Lloyd, and L. Maccone, *Nat. Photonics* **5**, 222 (2011).
- [4] G. Toth and I. Apellaniz, *J. Phys. A: Math. Theor.* **47**, 424006 (2014).
- [5] R. Demkowicz-Dobrzanski, M. Jarzyna, and J. Kołodyński, in *Prog. Optics*, Vol. 60, edited by E. Wolf (Elsevier, 2015) pp. 345–435.
- [6] R. Schnabel, *Phys. Rep.* **684**, 1 (2017).
- [7] C. L. Degen, F. Reinhard, and P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [8] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, *Rev. Mod. Phys.* **90**, 035005 (2018).
- [9] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, *Nat. Photonics* **12**, 724 (2018).
- [10] C. M. Caves, *Phys. Rev. D* **23**, 1693 (1981).
- [11] M. Holland and K. Burnett, *Phys. Rev. Lett.* **71**, 1355 (1993).
- [12] H. Lee, P. Kok, and J. P. Dowling, *J. Mod. Optic.* **49**, 2325 (2002).
- [13] D. Wineland, J. Bollinger, W. Itano, F. Moore, and D. Heinzen, *Phys. Rev. A* **46**, R6797 (1992).
- [14] K. McKenzie, D. A. Shaddock, D. E. McClelland, B. C. Buchler, and P. K. Lam, *Phys. Rev. Lett.* **88**, 231102 (2002).
- [15] J. Bollinger, W. M. Itano, D. Wineland, and D. Heinzen, *Phys. Rev. A* **54**, R4649 (1996).
- [16] D. Leibfried, M. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. Itano, J. Jost, C. Langer, and D. Wineland, *Science* **304**, 1476 (2004).
- [17] V. Giovannetti, S. Lloyd, and L. Maccone, *Science* **306**, 1330 (2004).
- [18] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *Phys. Rev. Lett.* **79**, 3865 (1997).
- [19] M. de Burgh and S. D. Bartlett, *Phys. Rev. A* **72**, 042301 (2005).
- [20] L. Pezzè and A. Smerzi, *Phys. Rev. Lett.* **102**, 100401 (2009).
- [21] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, *Phys. Rev. A* **94**, 052108 (2016).
- [22] T. Baumgratz and A. Datta, *Phys. Rev. Lett.* **116**, 030801 (2016).
- [23] H. Yuan, *Phys. Rev. Lett.* **117**, 160801 (2016).
- [24] M. Tsang, R. Nair, and X.-M. Lu, *Phys. Rev. X* **6**, 031033 (2016).
- [25] J. Liu and H. Yuan, *Phys. Rev. A* **96**, 042114 (2017).
- [26] R. Nichols, P. Liuzzo-Scorpo, P. A. Knott, and G. Adesso, *Phys. Rev. A* **98**, 012114 (2018).
- [27] W. Górecki, S. Zhou, L. Jiang, and R. Demkowicz-Dobrzański, *Quantum* **4**, 288 (2020).
- [28] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press, 1976).
- [29] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
- [30] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [31] M. Hayashi, *Communications in mathematical physics* **304**, 689 (2011).
- [32] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, *Phys. Rev. Lett.* **124**, 030501 (2020).
- [33] M. J. W. Hall and H. M. Wiseman, *Phys. Rev. X* **2**, 041006 (2012).
- [34] A. Luis and J. Peřina, *Phys. Rev. A* **54**, 4564 (1996).
- [35] V. Bužek, R. Derka, and S. Massar, *Phys. Rev. Lett.* **82**, 2207 (1999).
- [36] D. W. Berry and H. M. Wiseman, *Phys. Rev. Lett.* **85**, 5098 (2000).
- [37] B. Higgins, D. Berry, S. Bartlett, M. Mitchell, H. Wiseman, and G. Pryde, *New Journal of Physics* **11**, 073023 (2009).
- [38] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, *Phys. Rev. A* **80**, 052114 (2009).
- [39] T. Kafal and R. Demkowicz-Dobrzański, *Phys. Rev. A* **90**, 062313 (2014).
- [40] K. Matsumoto, *J. Phys. A.: Math. Theor.* **35**, 3111 (2002).
- [41] T. Baumgratz and A. Datta, *Phys. Rev. Lett.* **116**, 030801 (2016).
- [42] M. Gessner, L. Pezzè, and A. Smerzi, *Phys. Rev. Lett.* **121**, 130503 (2018).
- [43] N. Kura and M. Ueda, *Phys. Rev. A* **97**, 012101 (2018).
- [44] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, *Phys. Rev. Lett.* **121**, 043604 (2018).
- [45] R. Demkowicz-Dobrzański, W. Górecki, and M. Guță, *Journal of Physics A: Mathematical and Theoretical* **53**, 363001 (2020).
- [46] E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach, *Phys. Rev. Lett.* **85**, 5230 (2000).
- [47] E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach, *Phys. Rev. A* **63**, 052309 (2001).
- [48] G. Chiribella, G. M. D’Ariano, P. Perinotti, and M. F. Sacchi, *Phys. Rev. Lett.* **93**, 180503 (2004).
- [49] E. Bagan, M. Baig, and R. Muñoz Tapia, *Phys. Rev. A* **70**, 030301 (2004).
- [50] G. Chiribella, G. M. D’Ariano, and M. F. Sacchi, *Phys. Rev. A* **72**, 042338 (2005).
- [51] M. Hayashi, *Physics Letters A* **354**, 183 (2006).
- [52] M. Hayashi, *Communications in Mathematical Physics* **347**, 3 (2016).
- [53] J. Kahn, *Phys. Rev. A* **75**, 022326 (2007).
- [54] V. Gebhart, A. Smerzi, and L. Pezzè, *Phys. Rev. Applied* **16**, 014035 (2021).
- [55] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, *Phys. Rev. Lett.* **111**, 070403 (2013).
- [56] M. Szczykulska, T. Baumgratz, and A. Datta, *Advances in Physics: X* **1**, 621 (2016).
- [57] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, *Journal of Physics A: Mathematical and Theoretical* **53**, 023001 (2019).
- [58] D. Berry, private communication.
- [59] A. Holevo, *Reports on Mathematical Physics* **16**, 385 (1979).
- [60] M. Ozawa, *Reports on Mathematical Physics* **18**, 11 (1980).
- [61] M. Ozawa, *On the noncommutative theory of statistical decision* (Tokyo Institute of Technology, 1980).
- [62] N. Bogomolov, *Theory of Probability & Its Applications* **26**, 787 (1982).

- [63] M. Tsang, F. Albarelli, and A. Datta, *Phys. Rev. X* **10**, 031023 (2020).
- [64] F. Tanaka, *arXiv preprint arXiv:1410.3639* (2014).
- [65] J. R. P. Boas and M. Kac, *Duke Mathematical Journal* **12**, 189 (1945).
- [66] W.-K. Li, *Journal of Chemical Education* **61**, 1034 (1984).
- [67] H. R. Krishnamurthy, H. S. Mani, and H. C. Verma, *Journal of Physics A: Mathematical and General* **15**, 2131 (1982).
- [68] J. W. Turner, *Journal of Physics A: Mathematical and General* **17**, 2791 (1984).
- [69] G. S. Summy and D. T. Pegg, *Opt. Commun.* **77**, 75 (1990).
- [70] T. J. Baker, S. N. Saadatmand, D. W. Berry, and H. M. Wiseman, *Nature Physics* **17**, 179 (2021), [arXiv:2009.05296 \[quant-ph\]](#).
- [71] P. M. Anisimov, G. M. Raterman, A. Chiruvelli, W. N. Plick, S. D. Huver, H. Lee, and J. P. Dowling, *Phys. Rev. Lett.* **104**, 103602 (2010).
- [72] A. Rivas and A. Luis, *New Journal of Physics* **14**, 093052 (2012).
- [73] L. Pezzé, *Phys. Rev. A* **88**, 060101 (2013).
- [74] C. N. Gagatsos, D. Branford, and A. Datta, *Phys. Rev. A* **94**, 042342 (2016).
- [75] H. Imai and M. Hayashi, *New Journal of Physics* **11**, 043034 (2009).
- [76] F. Albarelli and R. Demkowicz-Dobrzanski, *arXiv e-prints*, [arXiv:2104.11264](#) (2021), [arXiv:2104.11264 \[quant-ph\]](#).

### Appendix A: Formal derivation of the bound Eq. (11) and its application for local estimation problem

In this appendix we discuss the problem of  $p$ -phase shift estimation for the continuous model (i.e. after replacing  $\frac{m_i}{N} \rightarrow \mu_i \in [0, 1]$ ). Moreover, we consider the most general adaptive strategy, when the amount of resources  $N$  is the number of application of an elementary gate  $U_\theta$ , where arbitrary large ancilla is allowed as well as the action on the state with additional unitaries  $V_i$  during the evolution.

Formally, let us consider  $\mathcal{H}_S = L^2(\{[\mu_1, \dots, \mu_p] : \forall_i \mu_i \geq 0, \sum_{i=1}^p \mu_i \leq 1\})$  (square-integrable functions on the set  $\{[\mu_1, \dots, \mu_p] : \forall_i \mu_i \geq 0, \sum_{i=1}^p \mu_i \leq 1\} \subset \mathbb{R}^p$ ). We define  $U_\theta$  to be a unitary gate, which action on the state  $\int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu f(\mu) |\mu\rangle \in \mathcal{H}_S$  is defined as:

$$U_\theta \left( \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu f(\mu) |\mu\rangle \right) = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu e^{i\mu\theta} f(\mu) |\mu\rangle. \quad (\text{A1})$$

Consider the most general adaptive strategy, where one chooses an arbitrary initial state  $|\Psi_{\text{in}}\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$  (where  $\mathcal{H}_A$  may be arbitrary large) and acts on it with a unitary gate  $U_\theta \otimes \mathbb{1}_A$   $N$  times and unitary controls  $V_i$  in between:

$$|\Psi_\theta^N\rangle = V_N(U_\theta \otimes \mathbb{1}_A)V_{N-1}\dots V_i(U_\theta \otimes \mathbb{1}_A)|\Psi_{\text{in}}\rangle. \quad (\text{A2})$$

In particular, if one chooses  $\mathcal{H}_A = \mathbb{C}$ ,  $V_i = \mathbb{1}$  and  $|\Psi_{\text{in}}\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu f(\mu) |\mu\rangle$  they receive the state from Eq. (5):

$$|\Psi_\theta^N\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu e^{iN\mu\theta} f(\mu) |\mu\rangle. \quad (\text{A3})$$

Let us name by  $\text{minimax}(\Theta, N)$  the minimax cost for  $\theta \in \Theta$ , optimized over initial state, unitary controls during evolution and the measurement:

$$\text{minimax}(\Theta, N) \equiv \inf_{M_{\tilde{\theta}}, \{V_i\}, |\Psi_{\text{in}}\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A} \sup_{\theta \in \Theta} \int d\tilde{\theta} \langle \Psi_\theta^N | M_{\tilde{\theta}} | \Psi_\theta^N \rangle (\tilde{\theta} - \theta)^2. \quad (\text{A4})$$

Then the following theorems hold:

**Theorem 1.** For completely unknown parameters  $\Theta = \mathbb{R}^p$ , the optimal cost may be obtained by applying covariant measurement, without invoking unitary controls during the evolution (e.i.  $\forall_i V_i = \mathbb{1}$ ), resulting in ecref:jointenergy:

$$\begin{aligned} \text{minimax}(\mathbb{R}^p, N) &= \frac{1}{N^2} \min_f \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu f^*(\mu) \left( \sum_{k=1}^p -\partial_{\mu_k}^2 \right) f(\mu), \\ &\text{with} \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu |f(\mu)|^2 = 1, \\ &f(\mu) = 0 \quad \text{for } \mu \text{ on the boundary } (\mu_i = 0 \vee \sum_i \mu_i = 1). \end{aligned} \quad (\text{A5})$$

**Theorem 2.** The  $\text{minimax}(\mathbb{R}^p, N)$  may be bounded from below by Eq. (11):

$$\text{minimax}(\mathbb{R}^p, N) \geq \frac{p^3}{N^2} \underbrace{\frac{4|A_0|^3}{27}}_{c \approx 1.89}. \quad (\text{A6})$$

where  $A_0 \approx -2.34$  is the first zero of the Airy function  $\text{Ai}(\cdot)$ .

**Theorem 3.** Assuming that  $\Theta$  is not degenerated in any direction, i.e. there exists finite-size cube  $[-\delta/2, +\delta/2]^p \in \Theta$ , the leading term  $\sim \frac{1}{N^2}$  of  $\text{minimax}(\Theta, N)$  does not depend on  $\Theta$ :

$$\forall_{\Theta} \lim_{N \rightarrow \infty} N^2 \text{minimax}(\Theta, N) = \lim_{N \rightarrow \infty} N^2 \text{minimax}(\mathbb{R}^p, N). \quad (\text{A7})$$

Moreover, for finite  $N$ , satisfying  $pN\delta \geq 2$ , the following lower bound holds:

$$\text{minimax}(\Theta, N) \geq \text{minimax}([-\delta/2, +\delta/2]^p, N) \geq \frac{cp^3}{N^2} \left( 1 - \frac{8p \log(pN\delta)}{N\delta} \right). \quad (\text{A8})$$

### 1. Irrelevance of adaptiveness in deriving the fundamental bound in estimation of completely unknown phases (proof of Theorem 1)

We start the proof by presenting an argument that in case of estimation of completely unknown phases, general adaptive estimation strategies do not provide any fundamental advantage compared with the strategy where all the photons are sent through the interferometer together (provided they can be prepared in an arbitrary multimode entangled state) — i.e. the strategy where all  $V_i = \mathbb{1}$ . This argument was shared with us by Dominic Berry [58].

First, let us discuss the structure of the output state Eq. (A2) (this is analogous to what was considered in [32] in the case of single parameter estimation). As each action of the gate  $U_{\theta}$  multiplies its eigenvectors by the factor  $e^{i\mu\theta}$ :  $U_{\theta} = \int d\mu e^{i\mu\theta} |\mu\rangle \langle \mu|$ , the final state Eq. (A2) may be written as:

$$\begin{aligned} |\Psi_{\theta}^N\rangle &= V_N(U_{\theta} \otimes \mathbb{1}_A) V_{N-1} \dots V_1(U_{\theta} \otimes \mathbb{1}_A) |\Psi_{\text{in}}\rangle = \\ &= \int d\mu_{(N)} \dots \int d\mu_{(1)} e^{i(\mu_{(N)} + \dots + \mu_{(1)})\theta} |\mu_{(N)}\rangle \left( \prod_{i=1}^{N-1} \langle \mu_{(i+1)} | V_i | \mu_{(i)} \rangle \right) \langle \mu_{(1)} | \Psi_{\text{in}} \rangle = \\ &= \int d(\mu_{(1)} + \dots + \mu_{(N)}) e^{i(\mu_{(1)} + \dots + \mu_{(N)})\theta} \int d\mu^{\perp} |\mu_{(N)}\rangle \left( \prod_{i=1}^{N-1} \langle \mu_{(i+1)} | V_i | \mu_{(i)} \rangle \right) \langle \mu_{(1)} | \Psi_{\text{in}} \rangle, \quad (\text{A9}) \end{aligned}$$

where we have split the integral into integration over  $(\mu_{(1)} + \dots + \mu_{(N)})$  and  $\mu^{\perp}$  which represents all directions for which  $(\mu_{(1)} + \dots + \mu_{(N)})$  is constant and also includes any necessary normalization factors. Note that in the above notation  $\langle \mu_{(1)} | \Psi_{\text{in}} \rangle \in \mathcal{H}_A$ , as well as  $\langle \mu_{(i+1)} | V_i | \mu_{(i)} \rangle$  is an operator acting on  $\mathcal{H}_A$ . Next, by rescaling the variable in the first integral  $\mu = \frac{\mu_{(1)} + \dots + \mu_{(N)}}{N}$  and naming the last one by  $\frac{1}{N} c(\mu) |g_{\mu}\rangle$  (where  $\langle g_{\mu} | g_{\mu} \rangle = 1$ ) we end with:

$$|\Psi_{\theta}^N\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\mu e^{iN\mu\theta} c(\mu) |g_{\mu}\rangle. \quad (\text{A10})$$

Note, that  $|g_{\mu}\rangle$  constructed in such a way are not necessary mutually orthogonal, which in general makes the class of the states of the form Eq. (A10) larger than that in Eq. (A3). For a given measurement  $\{M_{\tilde{\theta}}\}$  the cost maximized over  $\theta \in \mathbb{R}^p$  is given by:

$$\max_{\theta \in \mathbb{R}^p} \int d\tilde{\theta} \langle \Psi_{\tilde{\theta}}^N | M_{\tilde{\theta}} | \Psi_{\theta}^N \rangle (\tilde{\theta} - \theta)^2. \quad (\text{A11})$$

Next we show, that for any state of the form Eq. (A10) and the measurement  $\{M_{\tilde{\theta}}\}$  there exists a non-adaptive strategy with at least the same final precision. Consider the input state of the form  $\propto \int c(\mu) |\mu\rangle d\mu$ . After a free evolution, as a part of the measurement procedure, we entangle each  $|\mu\rangle$  with an ancillary system in  $|g_{\mu}\rangle$  state (this is possible since  $|\mu\rangle$  basis is orthogonal) obtaining:

$$|\Psi_{\theta}^{N'}\rangle \propto \int c(\mu) e^{iN\mu\theta} |\mu\rangle \otimes |g_{\mu}\rangle d\mu \quad (\text{A12})$$

and perform a measurement in the Fourier basis on the initial system:

$$|\varphi\rangle = \frac{1}{\sqrt{(2\pi/N)^p}} \int d\boldsymbol{\mu} e^{iN\boldsymbol{\mu}\varphi} |\boldsymbol{\mu}\rangle. \quad (\text{A13})$$

The resulting state, given outcome  $\varphi$ , reads

$$|\Psi_{\boldsymbol{\theta}}^{N''}(\varphi)\rangle \propto \int d\boldsymbol{\mu} c(\boldsymbol{\mu}) e^{iN\boldsymbol{\mu}(\boldsymbol{\theta}-\varphi)} |g_{\boldsymbol{\mu}}\rangle = |\Psi_{\boldsymbol{\theta}-\varphi}^N\rangle \quad (\text{A14})$$

and is obtained with probability distribution  $p(\varphi) = \langle \Psi_{\boldsymbol{\theta}}^{N'} | \varphi \rangle \langle \varphi | \otimes \mathbb{1} | \Psi_{\boldsymbol{\theta}}^{N'} \rangle$ . Next we apply the measurement which is optimal for the output state Eq. (A10), and properly shift the indication of estimator by the measured value  $\varphi$ . In consequence, the mean cost for the true value of parameter  $\boldsymbol{\theta}$  will be the same as the cost in the previously considered adaptive strategy Eq. (A10) for the value of parameter  $\boldsymbol{\theta} - \varphi$ , averaged with probability distribution  $p(\varphi)$ :

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^p} \int d\tilde{\boldsymbol{\theta}} \int d\varphi p(\varphi) \langle \Psi_{\boldsymbol{\theta}-\varphi}^N | M_{\tilde{\boldsymbol{\theta}}-\varphi} | \Psi_{\boldsymbol{\theta}-\varphi}^N \rangle (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^2. \quad (\text{A15})$$

As averaging the cost may only *decrease* the minimax bound, the statement is proven.

Therefore, from now on, when looking for the optimal strategy, we may focus on the output states of the form Eq. (A3). Moreover, for technical reasons, we will treat the function  $f(\boldsymbol{\mu})$  as a function defined on the whole  $\mathbb{R}^p$ , but with the condition, that it is equal to 0 outside of the region  $\forall \mu_i \geq 0, \sum_i \mu_i \leq 1$ .

Note that when only non-adaptive strategy is considered, it is a covariant problem of the group element estimation. Therefore, from the Hunt-Stein lemma [59, 60] (generalized for noncompact group case [52, 61, 62]) it follows that the search for the optimal measurement strategy may be restricted to projective covariant measurements [29]:

$$|\chi_{\tilde{\boldsymbol{\theta}}}\rangle = \frac{1}{\sqrt{(2\pi/N)^p}} \int_{\mathbb{R}^p} d\boldsymbol{\mu} e^{iN\boldsymbol{\mu}\tilde{\boldsymbol{\theta}}} |\boldsymbol{\mu}\rangle. \quad (\text{A16})$$

One can also provide an intuitive argument: as the problem is exactly equivalent to the position-shift estimation (when  $\boldsymbol{\theta}$  is treated as position), or momentum shift estimation problem (when  $\boldsymbol{\theta}$  is treated as momentum), the optimal way is to measure this observable directly using measurement operators projecting on the observable eigenbasis [63].

After applying the covariant measurement we get:

$$\begin{aligned} \Delta^2 \tilde{\boldsymbol{\theta}} &\geq \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} |\langle \chi_{\tilde{\boldsymbol{\theta}}} | \Psi_{f,\boldsymbol{\theta}}^N \rangle|^2 (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 = \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} |\langle \chi_{\tilde{\boldsymbol{\theta}}} | \Psi_{f,0}^N \rangle|^2 \tilde{\boldsymbol{\theta}}^2 = \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} \left| \frac{1}{\sqrt{(2\pi/N)^p}} \int d\boldsymbol{\mu} e^{-iN\boldsymbol{\mu}\tilde{\boldsymbol{\theta}}} f(\boldsymbol{\mu}) \right|^2 \tilde{\boldsymbol{\theta}}^2 \\ &= \min_f \int_{\mathbb{R}^p} d(N\tilde{\boldsymbol{\theta}}) \left| \hat{f}(N\tilde{\boldsymbol{\theta}}) \right|^2 \tilde{\boldsymbol{\theta}}^2 \stackrel{N\tilde{\boldsymbol{\theta}} \rightarrow \tilde{\boldsymbol{\theta}}}{=} \frac{1}{N^2} \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} \left| \hat{f}(\tilde{\boldsymbol{\theta}}) \right|^2 \tilde{\boldsymbol{\theta}}^2, \quad (\text{A17}) \end{aligned}$$

where in the first step we dropped the irrelevant dependence on  $\boldsymbol{\theta}$ ,  $\hat{f}$  is the Fourier transform of  $f$  and at the end we rescaled  $N\tilde{\boldsymbol{\theta}} \rightarrow \tilde{\boldsymbol{\theta}}$  to move the factor  $\frac{1}{N^2}$  in front of the integral. Going back to the position representation, and recalling the constraints on  $f$ , we get Eq. (A5).

## 2. Bound on the ground energy of a particle in a simplex potential well via a single degree of freedom formula (proof of Theorem 2)

Given a general multimode pure state:

$$|\Psi_f^N\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} f(\boldsymbol{\mu}) |\mu_1, \mu_2, \dots, \mu_p\rangle \quad (\text{A18})$$

let us define a single mode reduced density matrix corresponding to the  $i$ -th sensing arm mode:

$$\rho_i(\mu_i; \mu'_i) = \int \left( \prod_{j \neq i} d\mu_j d\mu'_j \delta(\mu_j - \mu'_j) \right) f(\boldsymbol{\mu}) f^*(\boldsymbol{\mu}'). \quad (\text{A19})$$



The formula for the lower bound on the estimation cost Eq. (A5) may be equivalently written in terms of the single mode reduced density matrices:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{N^2} \min_{f(\boldsymbol{\mu})} \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} \sum_{i=1}^p d\mu_i d\mu'_i \delta(\mu_i - \mu'_i) \left( -\frac{\partial^2}{\partial \mu_i^2} \right) \rho_i(\mu_i; \mu'_i). \quad (\text{A20})$$

Due to the inherent symmetry of the problem, we assume all the reduced density matrices to be identical and hence we do not distinguish them by a subscript  $i$ . As discussed in the main text, this symmetry implies also that the expectation value of the number of photons in each single ‘phase arm’ is  $\leq N/p$ . Therefore, looking for the lower bound of the above formula, we may relax the original constraints by ignoring the fact that  $\rho(\mu; \mu)$  comes from the multimode state satisfying conditions from Eq. (A5), and perform optimization with a weaker single mode constraints:

$$\rho(0; 0) = 0, \quad \int d\mu \rho(\mu; \mu) = 1, \quad \int d\mu \rho(\mu; \mu) N\mu \leq N/p, \quad \rho \geq 0. \quad (\text{A21})$$

This allows us to write the lower bound as:

$$\Delta^2 \tilde{\theta} \geq p \times \frac{1}{N^2} \min_{\rho} \int_0^\infty d\mu d\mu' \delta(\mu - \mu') \left( -\frac{\partial^2}{\partial \mu^2} \right) \rho(\mu; \mu') \quad (\text{A22})$$

with the above constraints imposed.

Let  $\bar{\rho}(\mu; \mu')$  be the solution of the above minimization problem. It implies that  $\int d\mu \bar{\rho}(\mu; \mu) N\mu = N/p$  (note that we have replaced the inequality with the equality here, since any function with average photon number  $t \cdot N/p$ , where  $0 < t < 1$ , can always be rescaled  $\bar{\rho}(\mu; \mu') \rightarrow t\bar{\rho}(t\mu; t\mu')$ , which reduces the cost without breaking constraints). If, in addition, the solution corresponds to a pure state  $\bar{\rho}(\mu; \mu') = g(\mu)g^*(\mu')$  then we indeed arrive at formula for the bound Eq. (9) stated in the main text.

$$\Delta^2 \tilde{\theta} \geq p \times \frac{1}{N^2} \min_g \int_0^\infty d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial \mu^2} \right) g(\mu) \quad (\text{A23})$$

with constraints:

$$g(0) = 0, \quad \int_0^\infty d\mu g^*(\mu)g(\mu) = 1, \quad \int_0^\infty d\mu g^*(\mu)g(\mu)N\mu = N/p. \quad (\text{A24})$$

The solution may be found using the standard Lagrange multiplier method,

$$-\frac{\partial^2}{\partial \mu^2} g(\mu) + g(\mu)(\lambda_1 + \mu\lambda_2) = 0 \Rightarrow g(\mu) \propto \text{Ai} \left( \frac{\lambda_1 + \lambda_2 \mu}{\lambda_2^{2/3}} \right), \quad (\text{A25})$$

where  $\text{Ai}(\cdot)$  is the Airy function of the first kind and after taking into account conditions Eq. (A24) we obtain:

$$g(\mu) = \frac{1}{\text{Ai}'(A_0)} \sqrt{\frac{2p|A_0|}{3}} \text{Ai} \left( A_0 + \frac{2p|A_0|}{3} \mu \right), \quad (\text{A26})$$

where  $A_0 \approx -2.34$  is the first zero of the Airy function and  $\text{Ai}'(\cdot)$  is its first derivative. The corresponding bound on the cost reads:

$$\Delta^2 \tilde{\theta} \geq p \times \frac{p^2}{N^2} \frac{4|A_0|^3}{27} = \frac{p^3}{N^2} \underbrace{\frac{4|A_0|^3}{27}}_{c \approx 1.89}. \quad (\text{A27})$$

What remains to be shown is that the assumption of a pure state is justified and that a mixed state cannot provide a lower value for the bound. The solution of the pure state case implies that for the mean photon number  $\bar{N}$  fixed (in this case  $\bar{N} = N/p$ ), the resulting optimal cost (for single phase) is equal  $\frac{1}{N^2} \min_g \int_0^\infty d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial \mu^2} \right) g(\mu) = c/\bar{N}^2$ . Consider now, instead of a pure state  $\bar{\rho} = |g\rangle\langle g|$ , an exemplary mixed state which for concreteness we choose to be a probabilistic mixture of two pure states  $\bar{\rho}' = p_1 |g_1\rangle\langle g_1| + p_2 |g_2\rangle\langle g_2|$ . The constraint on the mean photon number implies that

$$p_1 \bar{N}_1 + p_2 \bar{N}_2 = \bar{N}, \quad (\text{A28})$$

where  $\bar{N}_k$  are mean photon numbers in states  $|g_k\rangle$  respectively. The corresponding cost for the mixed state will be a weighted sum for the costs of the respective pure states,

$$p_1 \frac{c}{\bar{N}_1^2} + p_2 \frac{c}{\bar{N}_2^2} \geq \frac{c}{(p_1 \bar{N}_1 + p_2 \bar{N}_2)^2} = \frac{c}{\bar{N}^2}, \quad (\text{A29})$$

where the inequality follows from the convexity property of the  $1/x^2$  function (Jensen inequality). This reasoning can be trivially generalized to mixed states involving mixtures of more than two states and proves that a mixed state cannot provide a lower value for the bound Eq. (A22) than the optimal pure state.

### 3. Application for the local estimation problem (proof of Theorem 3)

We assume that the prior parameter region  $\Theta$  has a finite volume and hence we can always find  $\delta > 0$  such that  $[-\delta/2, +\delta/2]^p \subset \Theta$ . Replacing  $\Theta$  with its subset may only lower the minimax cost and hence any lower bound we obtain for the easier problem will also be valid for the original one. Therefore, in what follows we assume  $\Theta = [-\delta/2, +\delta/2]^p$ .

For the purposes of the proof, let us imagine for the moment that instead of knowing that  $\theta \in [-\delta/2, +\delta/2]^p$ , we assume that  $\theta$  is completely unknown and hence the region to consider is  $\theta \in \mathbb{R}^p$ . Let us also assume that we have  $N + p \cdot N_0$  gates at our disposal and that we spend  $p \cdot N_0$  resources to find an approximated values of the parameters  $\tilde{\theta}$ , such that the true value of the  $\theta$  (with high probability) lies in its close neighborhood  $[\theta_1 - \delta/2, \tilde{\theta}_1 + \delta/2] \times \dots \times [\tilde{\theta}_p - \delta/2, \tilde{\theta}_p + \delta/2]$ . Next we use the remaining  $N$  in the optimal way to estimate  $\theta$ , taking into account this knowledge. This approach may be suboptimal, from the point o view of the optimal use of all  $N + p \cdot N_0$  resources, and hence we may write the following inequality:

$$\text{minimax}(\mathbb{R}^p, N + p \cdot N_0) \leq \text{minimax}([- \delta/2, +\delta/2]^p, N) + R(p, \delta, N_0), \quad (\text{A30})$$

where in  $\text{minimax}([- \delta/2, +\delta/2]^p, N)$  we have shifted the region to the origin (which does not change the minimax cost). Here the term  $R(p, \delta, N_0)$  corresponds to the risk, that true value of  $\theta$  lies outside of predicted region  $[-\delta/2, +\delta/2]^p$  (formal definition will be given later). Intuitively, as the discrimination error decreases exponentially with  $N_0$ , for large value of  $N$  it is enough to take  $N_0$  sublinear in  $N$  to make the term  $R(p, \delta, N_0)$  small enough to become irrelevant. Therefore, we expect that the minimax costs for the finite and infinite regions approach the same limit in the leading order. Below we discuss the rate of convergence.

#### a. From no prior knowledge to finite knowledge

Consider the following strategy. First, for each phase we use  $N_0$  gates to get the state  $|\Psi_{\theta_i}^{N_0}\rangle = \int_0^1 d\mu_i e^{iN_0\mu_i\theta_i} w(\mu_i) |\mu_i\rangle$  (where all remaining  $\mu_{j \neq i} = 0$ ). After performing the covariant measurement we obtain the result  $\tilde{\theta}_i$  with probability:

$$|\langle \chi_{\tilde{\theta}_i} | \Psi_{\theta_i}^{N_0} \rangle|^2 = |\hat{w}(\theta_i - \tilde{\theta}_i)|^2 = p_{N_0}(\underbrace{\theta_i - \tilde{\theta}_i}_{\phi_i}). \quad (\text{A31})$$

We repeat it for all phases an get  $p_{N_0}(\phi) = \prod_{i=1}^p p_{N_0}(\phi_i)$ . In the next step we need to estimate the value of the difference  $\phi = \theta - \tilde{\theta}$ . Note that while  $\theta$  has a fixed value,  $\phi$  is a random variable and hence we may make use of the Bayesian cost formulas. Since such a procedure of estimating  $\theta$  using  $N + p \cdot N_0$  gates is in general suboptimal, the final mean variance may be bounded by the minimal cost when the total amount of resources  $N + p \cdot N_0$  are used:

$$\text{minimax}(\mathbb{R}^p, N + p \cdot N_0) \leq \underbrace{\min_{M_{\tilde{\phi}}, \rho_{\tilde{\phi}}^N} \int d\phi p_{N_0}(\phi) \int d\tilde{\phi} \text{Tr}(M_{\tilde{\phi}} \rho_{\tilde{\phi}}^N) (\tilde{\phi} - \phi)^2}_{\text{bayesian}(p_{N_0}, N)}. \quad (\text{A32})$$

This bounds from below the minimal obtainable Bayesian cost with a priori distribution  $p_{N_0}(\phi)$  and use of  $N$  gates,  $\text{bayesian}(p_{N_0}, N)$ .

Now we would like to use this result to state the bound for the minimax cost for  $\theta \in [-\delta/2, +\delta/2]^p$ . By definition the minimax cost is greater or equal than the minimal Bayesian cost for any prior with support inside of  $[-\delta/2, +\delta/2]^p$  (formal justification will be given in Eq. (A35)). As  $p_{N_0}(\phi)$  is given by the square of modulus of the Fourier transform of a finite support function  $w(\mu_i)$ ,  $p_{N_0}(\phi)$  will always stick outside any finite size region  $[-\delta/2, +\delta/2]^p$ . However, in principle these tails may be arbitrary small if sufficiently large  $N_0$  is allowed (the exemplary choice of  $w(\mu_i)$  will be given in next section).

For the purpose of the proof, let us define the distribution corresponding to  $p_{N_0}(\phi)$ , but with the tails cut:

$$p_{N_0}^\delta(\phi_i) = \begin{cases} \frac{1}{1-R_1} p_{N_0}(\phi_i) & \text{for } \phi_i \in [-\delta/2, +\delta/2] \\ 0 & \text{for } \phi_i \notin [-\delta/2, +\delta/2] \end{cases}, \quad (\text{A33})$$

where  $\frac{1}{1-R_1}$  is a proper normalization factor, i.e.

$$R_1 = 2 \int_{\delta/2}^{+\infty} d\phi_i p_{N_0}(\phi_i). \quad (\text{A34})$$

For such a distribution we may formally bound the minimax cost using the minimal Bayesian cost. Let  $\rho_\phi^{N, \text{minimax}}, M_\phi^{\text{minimax}}$  be the output state and the measurement optimizing the minimax cost. Then:

$$\begin{aligned} \text{minimax}([-\delta/2, +\delta/2]^p, N) &= \max_{\phi \in [-\delta/2, +\delta/2]^p} \int d\tilde{\phi} \text{Tr}(\rho_\phi^{N, \text{minimax}} M_{\tilde{\phi}}^{\text{minimax}}) (\tilde{\phi} - \phi)^2 \\ &\geq \int_{[-\delta/2, +\delta/2]^p} p_{N_0}^\delta(\phi) d\phi \int d\tilde{\phi} \text{Tr}(\rho_\phi^{N, \text{minimax}} M_{\tilde{\phi}}^{\text{minimax}}) (\tilde{\phi} - \phi)^2 \\ &\geq \min_{M_{\tilde{\phi}}, \rho_\phi^N} \int_{[-\delta/2, +\delta/2]^p} p_{N_0}^\delta(\phi) d\phi \int d\tilde{\phi} \text{Tr}(\rho_\phi^N M_{\tilde{\phi}}) (\tilde{\phi} - \phi)^2 = \text{bayesian}(p_{N_0}, N). \end{aligned} \quad (\text{A35})$$

Alternatively, the above inequality may be derived directly using the fact, that the optimal minimax cost is equal to the minimal Bayesian cost with the least favorable prior [64].

Summarizing Eq. (A32) and Eq. (A35) we have:

$$\begin{aligned} \text{minimax}(\mathbb{R}^p, N + p \cdot N_0) &\leq \text{bayesian}(p_{N_0}, N), \\ \text{bayesian}(p_{N_0}^\delta, N) &\leq \text{minimax}([-\delta/2, +\delta/2]^p, N). \end{aligned} \quad (\text{A36})$$

What remains is to connect  $\text{bayesian}(p_{N_0}, N)$  with  $\text{bayesian}(p_{N_0}^\delta, N)$ .

Let  $\rho_\phi^{\delta, N}, M_\phi^\delta$  be the output state and the measurement minimizing the Bayesian cost for  $p_{N_0}^\delta(\phi)$ . Obviously  $\tilde{\phi}$  that appear in  $M_\phi^\delta$  are only inside of  $[-\delta/2, +\delta/2]^p$ . By applying this state and measurement with the original prior distribution  $p_{N_0}(\phi)$  we get:

$$\text{bayesian}(p_{N_0}, N) \leq \int_{\mathbb{R}^p} d\phi p_{N_0}(\phi) \int_{[-\delta/2, +\delta/2]^p} d\tilde{\phi} \text{Tr}(M_\phi^\delta \rho_\phi^{\delta, N}) (\tilde{\phi} - \phi)^2. \quad (\text{A37})$$

Expanding  $(\tilde{\phi} - \phi)^2 = \sum_i (\tilde{\phi}_i - \phi_i)^2$ , and considering a single term from the sum corresponding to  $\phi_i$ , we can divide the resulting integrals appearing on the RHS of the above equation into three parts:

$$\begin{aligned} &\int_{\mathbb{R}^p} d\phi p_{N_0}(\phi) \int_{[-\delta/2, +\delta/2]} d\tilde{\phi}_i \text{Tr} \left[ \underbrace{\left( \int_{[-\delta/2, +\delta/2]^{p-1}} d\tilde{\phi}_{j \neq i} M_{\tilde{\phi}}^\delta \right)}_{\overline{M}_{\phi_i}^\delta} \rho_\phi^{\delta, N} \right] (\tilde{\phi}_i - \phi_i)^2 = \\ (a) \quad &\int_{[-\delta/2, +\delta/2]^p} d\phi p_{N_0}(\phi) \int_{[-\delta/2, +\delta/2]} d\tilde{\phi}_i \text{Tr} \left[ \overline{M}_{\phi_i}^\delta \rho_\phi^{\delta, N} \right] (\tilde{\phi}_i - \phi_i)^2 + \\ (b) \quad &\int_{[-\delta/2, +\delta/2]} d\phi_i p_{N_0}(\phi_i) \int_{[-\delta/2, +\delta/2]} d\tilde{\phi}_i \text{Tr} \left[ \overline{M}_{\phi_i}^\delta \underbrace{\int_{\mathbb{R}^{p-1} \setminus [-\delta/2, +\delta/2]^{p-1}} d\phi_{j \neq i} p(\phi_j) \rho_\phi^{\delta, N}}_{\overline{\rho}_{\phi_i}^{\delta, N}} \right] (\tilde{\phi}_i - \phi_i)^2 + \\ (c) \quad &\int_{\mathbb{R} \setminus [-\delta/2, +\delta/2]} d\phi_i p_{N_0}(\phi_i) \int_{[-\delta/2, +\delta/2]} d\tilde{\phi}_i \text{Tr} \left[ \overline{M}_{\phi_i}^\delta \int_{\mathbb{R}^{p-1}} d\phi_{j \neq i} p(\phi_j) \rho_\phi^{\delta, N} \right] (\tilde{\phi}_i - \phi_i)^2. \end{aligned} \quad (\text{A38})$$

Part (a) represents the contribution where  $\phi \in [-\delta/2, +\delta/2]^p$ , so after summing over all  $i$  it is exactly equal  $(1 - R_1)^p \cdot \text{bayesian}(p_{N_0}^\delta, N)$ . Part (b) corresponds to the case where  $\phi_i$  belongs to  $[-\delta/2, +\delta/2]$  (probability  $1 - R_1$ ) but at least one of  $\phi_{j \neq i}$  does not (probability  $1 - (1 - R_1)^{p-1} = \text{Tr}(\bar{p}_{\phi_i}^{\delta, N})$ ). It means that this integral may be bounded from above by the worst possible estimation strategy (resulting in squared error  $\delta^2$ ) multiplied by this probability— $(1 - R_1)(1 - (1 - R_1)^{p-1})\delta^2$ . For our purpose it is enough to use a weaker, but simpler, bound  $(1 - R_1)(1 - (1 - R_1)^{p-1})\delta^2 \leq (p - 1)R_1\delta^2$ .

Finally (c), where  $\phi_i$  is outside of  $[-\delta/2, +\delta/2]$ , may be bounded from above by:

$$R_2 = 2 \int_{\delta/2}^{+\infty} d\phi_i p_{N_0}(\phi_i) (\phi_i - (-\delta/2))^2, \quad (\text{A39})$$

where the most pessimistic scenario is assumed (each time when the true value of  $\phi_i$  is to the right of  $[-\delta/2, +\delta/2]$ , our estimator point is at the left border of the region  $-\delta/2$ ; similarly in the opposite situation).

Combining all that was said above and taking the sum over  $p$  parameters, we may bound the RHS of Eq. (A37) by:

$$(1 - R_1)^p \text{bayesian}(p_{N_0}^\delta, N) + \underbrace{p(p - 1)R_1\delta^2 + pR_2}_{R(p, \delta, N_0)}, \quad (\text{A40})$$

and we arrive at:

$$\text{bayesian}(p_{N_0}, N) \leq (1 - R_1)^p \text{bayesian}(p_{N_0}^\delta, N) + R(p, \delta, N_0). \quad (\text{A41})$$

By rearranging this inequality and using  $(1 - R_1) < 1$  we get:

$$\text{bayesian}(p_{N_0}^\delta, N) \geq \text{bayesian}(p_{N_0}, N) - R(p, \delta, N_0). \quad (\text{A42})$$

After applying the above to Eq. (A36) we end up with:

$$\text{minimax}([-\delta/2, +\delta/2]^p, N) \geq \text{minimax}(\mathbb{R}^p, N + p \cdot N_0) - R(p, \delta, N_0), \quad (\text{A43})$$

which was to be proved. What remains is to find  $w(\mu_i)$  (with corresponding  $p_{N_0}(\phi_i)$ ), for which  $R(p, \delta, N_0)$  decrease sufficiently fast with  $N_0$ .

#### b. Irrelevance of the size of $\Theta$ in the limit $N \rightarrow \infty$

Following [32] we choose  $w(\mu_i)$  to be proportional to the self-convolution of the Kaiser window function with window duration  $N_0/2$ . Then  $p_{N_0}(\phi_i)$  is proportional to the fourth power of the Fourier transform of the Kaiser window function (from now on in this section the subscript  $i$  in  $\phi_i$  will be omitted to simplify notation):

$$p_{\alpha, L}(\phi) = \mathcal{N}_\alpha L \text{sinc}^4 \left( \pi \alpha \sqrt{(L\phi/4\alpha)^2 - 1} \right) = \mathcal{N}_\alpha L \frac{\sinh^4 \left( \pi \alpha \sqrt{1 - (L\phi/4\alpha)^2} \right)}{\left( \pi \alpha \sqrt{1 - (L\phi/4\alpha)^2} \right)^4}, \quad (\text{A44})$$

where  $L = 2N_0$  is the bandwidth,  $\alpha$  determines the shape and  $\mathcal{N}_\alpha$  may be bounded by:

$$\mathcal{N}_\alpha \lesssim 4\sqrt{2}\pi^4 \alpha^{7/2} e^{-4\pi\alpha}, \quad (\text{A45})$$

where the bound is tight for big  $\alpha$  [32]. As shown in [32] only exponentially small (with  $\alpha$ ) part lays outside of region  $[-4\alpha/L, 4\alpha/L]$  and, therefore, for our purpose we choose  $\delta/2 = 4\alpha/L$ . We may bound  $R_1$  and  $R_2$  by:

$$\begin{aligned} R_1 &= 2\mathcal{N}_\alpha \int_{4\alpha/L}^{+\infty} d\phi L \text{sinc}^4 \left( \pi \alpha \sqrt{(L\phi/4\alpha)^2 - 1} \right) = 2\mathcal{N}_\alpha \frac{4\alpha}{L} \int_1^{+\infty} dx L \text{sinc}^4 \left( \pi \alpha \sqrt{x^2 - 1} \right) \leq \\ &\leq 8\mathcal{N}_\alpha \alpha \left( \int_1^2 dx + \frac{1}{\pi^4 \alpha^4} \int_2^\infty dx \frac{1}{(x^2 - 1)^2} \right) = 8\mathcal{N}_\alpha \alpha \left( 1 + \frac{1/3 - \log(3)/4}{\pi^4 \alpha^4} \right)^{\alpha > 1/2} \leq 16\mathcal{N}_\alpha \alpha, \end{aligned} \quad (\text{A46})$$

$$\begin{aligned} R_2 &= 2\mathcal{N}_\alpha \int_{4\alpha/L}^{+\infty} d\phi L \text{sinc}^4 \left( \pi \alpha \sqrt{(L\phi/4\alpha)^2 - 1} \right) (\theta + \delta/2)^2 = 2\mathcal{N}_\alpha \frac{4\alpha}{L} \int_1^{+\infty} dx L \text{sinc}^4 \left( \pi \alpha \sqrt{x^2 - 1} \right) (x+1)^2 \left( \frac{4\alpha}{L} \right)^2 \leq \\ &\leq 2\mathcal{N}_\alpha L \left( \frac{4\alpha}{L} \right)^3 \left( \int_1^2 dx (x+1)^2 + \frac{1}{\pi^4 \alpha^4} \int_2^\infty dx \frac{(x+1)^2}{(x^2 - 1)^2} \right) = 2\mathcal{N}_\alpha L \left( \frac{4\alpha}{L} \right)^3 \left( \frac{19}{3} + \frac{1}{\pi^4 \alpha^4} \right)^{\alpha > 1/2} \leq 14\mathcal{N}_\alpha L \left( \frac{4\alpha}{L} \right)^3, \end{aligned} \quad (\text{A47})$$

where inequalities  $\text{sinc}(x) \leq 1$  and  $\text{sinc}(x) \leq 1/x$  were used. We have:

$$R(p, \delta, N_0) = p(p-1)R_1\delta^2 + pR_2 \leq 16p^2\mathcal{N}_\alpha\alpha \left(\frac{8\alpha}{L}\right)^2 + 14p\mathcal{N}_\alpha L \left(\frac{4\alpha}{L}\right)^3 = \mathcal{N}_\alpha \frac{\alpha^3}{L^2} (1024p^2 + 896p) \leq 1920\mathcal{N}_\alpha \frac{\alpha^3}{L^2} p^2, \quad (\text{A48})$$

which after substituting  $L = 2N_0$ ,  $\alpha = N_0\delta/4$ , yields:

$$R(p, \delta, N_0) \leq \frac{15}{2}\mathcal{N}_{N_0\delta/4}N_0\delta^3p^2. \quad (\text{A49})$$

To derive Eq. (A7), we use a trivial fact  $\forall_\Theta \text{minimax}(\mathbb{R}^p, N) \geq \text{minimax}(\Theta, N)$  and Eq. (A43) and we get:

$$\text{minimax}(\mathbb{R}^p, N) \geq \text{minimax}(\Theta, N) \geq \text{minimax}(\mathbb{R}^p, N + p \cdot N_0) - \frac{15}{2}\mathcal{N}_{N_0\delta/4}N_0\delta^3p^2. \quad (\text{A50})$$

Next, as  $\text{minimax}(\mathbb{R}^p, N + p \cdot N_0) = \frac{N^2}{(N+pN_0)^2} \text{minimax}(\mathbb{R}^p, N)$ , we choose  $N_0 = \sqrt{N}$  and apply  $\lim_{N \rightarrow \infty} N^2 \cdot$  to the above inequality in order to get:

$$\lim_{N \rightarrow \infty} N^2 \text{minimax}(\mathbb{R}^p, N) \geq \lim_{N \rightarrow \infty} N^2 \text{minimax}(\Theta, N) \geq \lim_{N \rightarrow \infty} N^2 \text{minimax}(\mathbb{R}^p, N), \quad (\text{A51})$$

which was to be proven. In order to prove Eq. (A8) more subtle choice of  $N_0$  will be required.

### c. Convergence of the bound for finite-size $\Theta$

From Eq. (A6) we have

$$\text{minimax}(\mathbb{R}^p, N + p \cdot N_0) \geq \frac{cp^3}{(N + p \cdot N_0)^2} = \frac{cp^3}{N^2} \left(1 - \frac{2pN_0}{N} + \frac{3(pN_0/N)^2 + 2(pN_0/N)^3}{(1 + pN_0/N)^2}\right), \quad (\text{A52})$$

with  $c = \frac{4|A_0|^{13}}{27} \approx 1.89$ . Therefore, when combining this formula with Eq. (A43), we obtain a corresponding bound for the finite prior case:

$$\text{minimax}([-\delta/2, +\delta/2]^p, N) \geq \frac{cp^3}{N^2} \left(1 - \frac{2pN_0}{N} + \frac{3(pN_0/N)^2 + 2(pN_0/N)^3}{(1 + pN_0/N)^2} - \frac{N^2 R(p, \delta, N_0)}{cp^3}\right). \quad (\text{A53})$$

To make  $R(p, \delta, N_0)$  sufficiently small regardless of  $p$  and  $\delta$ , we choose  $N_0 = \frac{4}{\delta} \log(pN\delta)$ . Next we will show that the difference of two last terms appearing in above equation is strictly positive for sufficiently big  $pN\delta$ , so they may be neglected without breaking inequality. To do that, let us bound it from below by the function of  $pN\delta$ .

Note that  $\frac{3x^2+2x^3}{(1+x)^2}$  is monotonically increasing for positive  $x$  and therefore:

$$\frac{3(pN_0/N)^2 + 2(pN_0/N)^3}{(1 + pN_0/N)^2} \geq \frac{3(N_0/(pN))^2 + 2(N_0/(pN))^3}{(1 + N_0/(pN))^2} = \frac{3(4 \log(pN\delta)/(pN\delta))^2 + 2(4 \log(pN\delta)/(pN\delta))^3}{(1 + 4 \log(pN\delta)/(pN\delta))^2}, \quad (\text{A54})$$

where in the first step  $pN_0/N$  was replaced by  $N_0/(pN)$  everywhere. Moreover, using Eq. (A49):

$$\frac{N^2 R(p, \delta, N_0)}{cp^3} \leq \frac{N^2}{cp^3} \frac{15}{2} \mathcal{N}_{N_0\delta/4} N_0 \delta^3 p^2 \leq \frac{N^2}{c} \frac{15}{2} \mathcal{N}_{N_0\delta/4} N_0 \delta^3 p^2 = \frac{30}{c} (pN\delta)^2 \mathcal{N}_{\log(pN\delta)} \log(pN\delta), \quad (\text{A55})$$

where in the second step the factor  $p^3$  was removed from denominator.

Taking two above together, naming by  $y = pN\delta$  an using the bound for  $\mathcal{N}_{\log(pN\delta)}$  Eq. (A45) we have therefore:

$$\frac{3(pN_0/N)^2 + 2(pN_0/N)^3}{(1 + pN_0/N)^2} - \frac{N^2 R(p, \delta, N_0)}{cp^3} \geq \frac{48(\log(y)/y)^2 + 128(\log(y)/y)^3}{(1 + 4 \log(y)/y)^2} - \frac{120\sqrt{2}\pi^4}{c} y^{2-4\pi} \log^{9/2}(y) \quad (\text{A56})$$

It may be checked numerically, that for  $c \geq 1.89$  it is strictly positive for  $y \geq 2$ . Therefore, for  $pN\delta \geq 2$  (which justify also usage  $\alpha > 1/2$  in Eq. (A46) and Eq. (A47)), Eq. (A53) implies:

$$\text{minimax}([-\delta/2, +\delta/2]^p, N) \geq \frac{cp^3}{N^2} \left(1 - \frac{8p \log(pN\delta)}{N\delta}\right). \quad (\text{A57})$$

d. *Relation with the single-parameter Bayesian estimation bound*

For completeness, we show here how the above reasoning might be used to rederive the Bayesian bound for the single parameter unitary estimation with finite bandwidth prior, Eq. (12) from [32], as well as the minimax analogy of the Bayesian bound for rectangular prior of width  $\delta$ , Eq. (14) from [32]. Note the difference in notation as the  $N$  from this paper corresponds to  $n$  from [32].

Indeed, starting with Eq. (A32), and using the fact that the minimal achievable cost, when estimating a completely unknown single phase using  $N$  phase gates, is given by  $\pi^2/N^2$ , we get:

$$\text{bayesian}(p_{N_0}, N) \geq \frac{\pi^2}{(N + N_0)^2}. \quad (\text{A58})$$

Next, we extend the result to an arbitrary generator with spectrum in  $[\lambda_-, \lambda_+]$  (instead of  $\{0, 1\}$ ), which corresponds to the replacement  $N \rightarrow N(\lambda_+ - \lambda_-)$ ,  $N_0 \rightarrow N_0(\lambda_+ - \lambda_-)$  in RHS of the equation above. Finally, from Eq. (A31), as  $\hat{w}(\theta)$  has a finite bandwidth of size  $N_0(\lambda_+ - \lambda_-)$ , the bandwidth of  $p_{N_0}(\theta)$  is at most  $2N_0(\lambda_+ - \lambda_-)$ . Moreover, from [32, 65], for any non-negative integrable finite bandwidth function  $p_L(\theta)$  there exists a proper function  $\hat{w}(\theta)$  with bandwidth  $L/2$  satisfying  $p_L(\theta) = |\hat{w}(\theta)|^2$ . Therefore:

$$\text{bayesian}(p_L, N) \geq \frac{\pi^2}{(N(\lambda_+ - \lambda_-) + L/2)^2}, \quad (\text{A59})$$

which is exactly Eq. (12) from [32].

Analogously, using Eq. (A57) with  $c \rightarrow \pi^2$ ,  $p \rightarrow 1$ ,  $N \rightarrow N(\lambda_+ - \lambda_-)$  we get:

$$\text{minimax}([- \delta/2, + \delta/2], N) \geq \frac{\pi^2}{[N(\lambda_+ - \lambda_-)]^2} \left( 1 - \frac{8 \log(N(\lambda_+ - \lambda_-)\delta)}{N(\lambda_+ - \lambda_-)\delta} \right), \quad (\text{A60})$$

which is similar to the bound for the Bayesian cost Eq. (14) from [32], but converges faster due to the lack of square root over the last term.

## Appendix B: Demonstrating joint phase estimation advantage over separate strategies using the state given in Eq. (16)

### 1. Joint phase estimation cost

For a given (not normalized) function:

$$f(\boldsymbol{\mu}) = \left( \prod_{i=1}^p \mu_i \right)^\alpha \left( 1 - \sum_{i=1}^p \mu_i \right)^\beta, \quad (\text{B1})$$

(where  $\alpha, \beta \geq \frac{1}{2}$ ) we will first calculate the normalization factor:

$$\mathcal{N} = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} f^*(\boldsymbol{\mu}) f(\boldsymbol{\mu}) \quad (\text{B2})$$

and then the corresponding ‘mean energy’:

$$\mathcal{E} = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} f^*(\boldsymbol{\mu}) \left( \sum_{i=1}^p -\partial_i^2 \right) f(\boldsymbol{\mu}). \quad (\text{B3})$$

The final estimation cost will correspond to the ratio  $1/N^2 \cdot \mathcal{E}/\mathcal{N}$ .

Let us introduce the objects representing ‘(P)roduct’ and ‘1 minus (S)um’ of coordinates:

$$P_k = \prod_{i=1+k}^p \mu_i, \quad S_k = 1 - \sum_{i=1+k}^p \mu_i, \quad (\text{B4})$$

so  $k$  indicates which coordinates are dropped. Then  $f(\boldsymbol{\mu}) = P_0^\alpha S_0^\beta$  and:

$$\mathcal{N} = \int_0^{S_p} d\mu_p \cdots \int_0^{S_2} d\mu_2 \int_0^{S_1} d\mu_1 P_0^{2\alpha} S_0^{2\beta}. \quad (\text{B5})$$

For the first integral we have:

$$\int_0^{S_1} d\mu_1 P_0^{2\alpha} S_0^{2\beta} = \int_0^{S_1} (P_1 \mu_1)^{2\alpha} (S_1 - \mu_1)^{2\beta} d\mu_1 = P_1^{2\alpha} S_1^{2\beta+(1+2\alpha)} \frac{\Gamma(1+2\alpha)\Gamma(1+2\beta)}{\Gamma(1+2\beta+(1+2\alpha))}. \quad (\text{B6})$$

For the  $k^{\text{th}}$  one:

$$\int_0^{S_k} d\mu_k P_{k-1}^{2\alpha} S_{k-1}^{2\beta+(k-1)(1+2\alpha)} = \int_0^{S_k} d\mu_k (P_k \mu_k)^{2\alpha} (S_k - \mu_k)^{2\beta+(k-1)(1+2\alpha)} = P_k^{2\alpha} S_k^{2\beta+k(1+2\alpha)} C_{\alpha,\beta,k}, \quad (\text{B7})$$

where

$$C_{\alpha,\beta,k} = \frac{\Gamma(1+2\alpha)\Gamma(1+2\beta+(k-1)(1+2\alpha))}{\Gamma(1+2\beta+k(1+2\alpha))}. \quad (\text{B8})$$

As  $P_p = 1$ ,  $S_p = 1$ , the  $\mathcal{N}$  is therefore given by:

$$\mathcal{N} = \prod_{k=1}^p C_{\alpha,\beta,k} = \frac{\Gamma(1+2\alpha)^p \Gamma(1+2\beta)}{\Gamma(1+2\beta+p(1+2\alpha))}. \quad (\text{B9})$$

Due to symmetry of  $f(\boldsymbol{\mu})$ , the mean value of  $-\partial_i^2$  operator is the same for each coordinate. Therefore it is sufficient to calculate it for  $i = 1$ . We have

$$|\partial_1 f(\boldsymbol{\mu})|^2 = (\partial_1 (P_1 \mu_1)^\alpha (S_1 - \mu_1)^\beta)^2 = (S_1 - \mu_1)^{2(\beta-1)} (P_1 \mu_1)^{2\alpha} (\alpha S_1 - (\alpha + \beta) \mu_1)^2 \mu_1^{-2}, \quad (\text{B10})$$

so for the first integral we have:

$$\int_0^{S_1} d\mu_1 |\partial_1 f(\boldsymbol{\mu})|^2 = P_1^{2\alpha} S_1^{-1+2\alpha+2\beta} \frac{\alpha\beta\Gamma(-1+2\alpha)\Gamma(-1+2\beta)}{(-1+2\alpha+2\beta)\Gamma(2(-1+\alpha+\beta))} \quad (\text{B11})$$

and for the  $k^{\text{th}}$  one ( $k \geq 2$ ):

$$\int_0^{S_k} d\mu_k P_{k-1}^{2\alpha} S_{k-1}^{2(\beta-1)+(k-1)(1+2\alpha)} = P_k^{2\alpha} S_k^{2(\beta-1)+k(1+2\alpha)} D_{\alpha,\beta,k}, \quad (\text{B12})$$

with

$$D_{\alpha,\beta,k} = \frac{\Gamma(1+2\alpha)\Gamma(1+2(\beta-1)+(k-1)(1+2\alpha))}{\Gamma(1+2(\beta-1)+k(1+2\alpha))}. \quad (\text{B13})$$

Therefore the total energy reads

$$\mathcal{E} = p \times \frac{\alpha\beta\Gamma(-1+2\alpha)\Gamma(-1+2\beta)}{(-1+2\alpha+2\beta)\Gamma(2(-1+\alpha+\beta))} \prod_{k=2}^p D_{\alpha,\beta,k} \quad (\text{B14})$$

where:

$$\prod_{k=2}^p D_{\alpha,\beta,k} = \frac{\Gamma(1+2\alpha)^{p-1} \Gamma(1+2(\beta-1)+(1+2\alpha))}{\Gamma(1+2(\beta-1)+p(1+2\alpha))}. \quad (\text{B15})$$

The variance of the estimator  $\Delta^2 \tilde{\boldsymbol{\theta}} = 1/N^2 \cdot \mathcal{E}/\mathcal{N}$  can be simplified to:

$$\Delta^2 \tilde{\boldsymbol{\theta}} = \frac{1}{N^2} \cdot \frac{\mathcal{E}}{\mathcal{N}} = \frac{1}{N^2} \cdot p \times \frac{(1-\alpha-\beta)(-1+2\beta+p+2\alpha p)(2\beta+p+2\alpha)}{2(-1+2\alpha)(-1+2\beta)}. \quad (\text{B16})$$

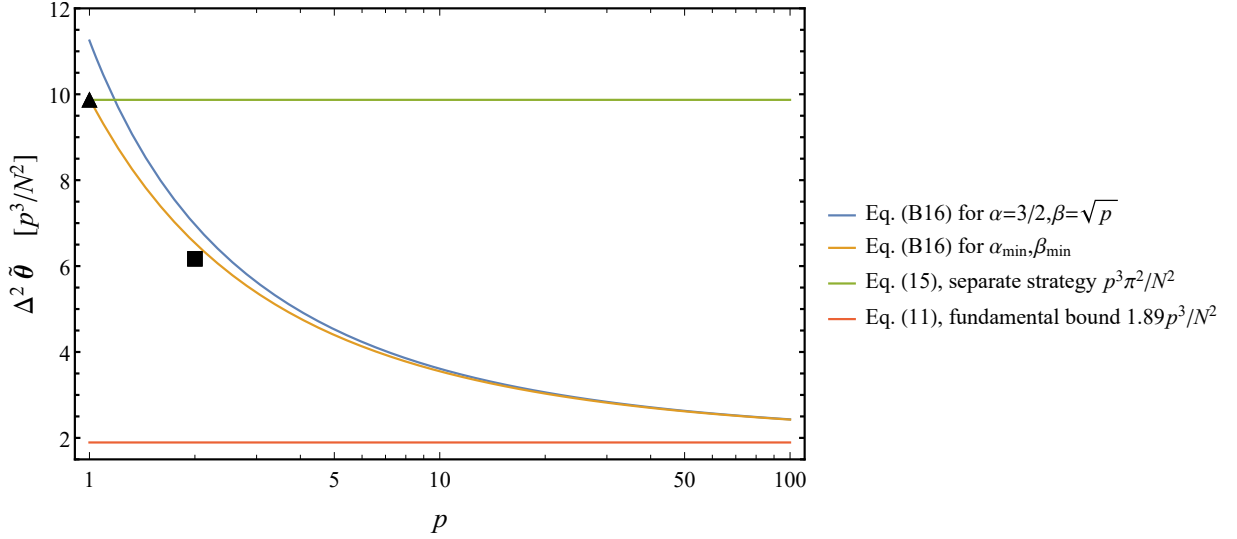


FIG. B1. Comparison between the cost obtained by the ansatz state, by the optimal separate strategy and the fundamental bound. The lines are plotted as continuous only for better visibility (but they are defined only for natural  $p$ ). For  $p = 1$  and  $p = 2$  also analytically found optimal costs are marked respectively by  $\blacktriangle$  and  $\blacksquare$ .

By direct calculation one may find, that  $\alpha, \beta \geq \frac{1}{2}$  which minimize the expression above are

$$\alpha_{\min} = \frac{1}{2} + \sqrt{\frac{4p^2 + 6p + 2 - 4\sqrt{p(1+2p)^2}}{4(p-1)^2}} = \frac{3}{2} + \mathcal{O}\left(\frac{1}{\sqrt{p}}\right), \quad (\text{B17})$$

$$\beta_{\min} = \frac{1 + 2p + \sqrt{2}\sqrt{p(1+2p)^2}\sqrt{\frac{p(3+2p)+1-2\sqrt{p(1+2p)^2}}{(p-1)^2}}}{4p+2} = \sqrt{p} + \mathcal{O}(1). \quad (\text{B18})$$

Looking for the leading term for large  $p$  we may therefore set  $\alpha = \frac{3}{2}$ ,  $\beta = \sqrt{p}$  and then:

$$\Delta^2 \tilde{\theta} = \frac{1}{N^2} \cdot \frac{\mathcal{E}}{\mathcal{N}} = \frac{p(1+2\sqrt{p})^2 \sqrt{p}(4p+2\sqrt{p}-1)}{N^2(8\sqrt{p}-4)} \stackrel{p \gg 1}{\approx} \frac{2p^3}{N^2}. \quad (\text{B19})$$

In Fig. B1 we compare how the cost obtained for the ansatz Eq. (B16) changes with increasing  $p$  for  $\alpha, \beta$  equal respectively  $\frac{3}{2}, \sqrt{p}$  and  $\alpha_{\min}, \beta_{\min}$ . As a reference we also plot the cost corresponding to the separate strategy and the fundamental bound Eq. (11) (not necessary saturable). We see that the difference in the cost obtained by  $\alpha, \beta$  equal  $\frac{3}{2}, \sqrt{p}$  and  $\alpha_{\min}, \beta_{\min}$  is relatively small (almost negligible for  $p \geq 10$ ); therefore, for the simplicity of calculation further we discuss case with  $\alpha = \frac{3}{2}, \beta = \sqrt{p}$ .

It is worth to remind here, that, as mentioned in the main text, the problem given in ecref:jointenergy has a known exact solution for  $p = 1, 2$  [32, 66], which are respectively:

$$\blacktriangle \quad p = 1: \quad f(\mu_1) \propto \sin(\pi\mu_1) \quad \rightarrow \Delta^2 \tilde{\theta} = \frac{\pi^2}{N^2} \quad (\text{B20})$$

$$\blacksquare \quad p = 2: \quad f(\mu_1, \mu_2) \propto \sin(\pi\mu_1) \sin(2\pi\mu_2) + \sin(2\pi\mu_1) \sin(\pi\mu_2) \quad \rightarrow \Delta^2 \tilde{\theta} = \frac{5\pi^2}{N^2}. \quad (\text{B21})$$

From Fig. B1 one can see that for  $p = 1, 2$  the cost obtained by the ansatz is very close to the optimal one.

## 2. The distribution of photons between the arms of the interferometer

For  $\alpha = \frac{3}{2}, \beta = \sqrt{p}$ , chosen to minimize the total cost of estimation, let us investigate some more features of the state Eq. (B1). The distribution of photons in the  $i^{\text{th}}$  sensing arm is given by the diagonal elements of the corresponding single



mode reduced density matrix Eq. (A19). By applying the methods from previous section one may immediately see that it is proportional to:

$$\rho(\mu_i; \mu_i) \propto \mu_i^3 (1 - \mu_i)^{2(2p + \sqrt{p} - 2)}. \quad (\text{B22})$$

The mean number of photons in each sensing arm is therefore given as:

$$\mathbb{E}[N\mu_i] = \int_0^1 d\mu_i \rho(\mu_i; \mu_i) N\mu_i = \frac{4N}{1 + 2\sqrt{p} + 4p} \approx \frac{N}{p}. \quad (\text{B23})$$

From that, the expectation value of the number of photons in the reference arm is:

$$\mathbb{E}[N - \sum_{i=1}^p N\mu_i] = N - p \times \mathbb{E}[N\mu_i] = \frac{(1 + 2\sqrt{p})N}{1 + 2\sqrt{p} + 4p} \approx \frac{N}{2\sqrt{p}}. \quad (\text{B24})$$

We see, that it decreases slower with  $p$  than the mean number of photons for any single sensing arm; still, in the asymptotic regime  $p \rightarrow \infty$  it tends to zero and almost all photons are distributed between the sensing arms. Note that analogous situations occurs for the state maximizing QFI Eq. (13). As mentioned in the main text, this is where the advantage of the joint over separate estimation strategy comes from—by using a single reference arm to measure all the phases, more photons remain for the sensing arms.

An interesting question is if the distributions of photons in different sensing arms are mutually correlated. For the joint distribution of photons in  $i^{\text{th}}$  and  $j^{\text{th}}$  sensing arms we have:

$$\rho(\mu_i, \mu_j; \mu_i, \mu_j) \propto (\mu_i \mu_j)^3 (1 - \mu_i - \mu_j)^{2(2p + \sqrt{p} - 4)}, \quad (\text{B25})$$

$$\text{corr}(N\mu_i, N\mu_j) = \frac{\mathbb{E}[N\mu_i N\mu_j] - \mathbb{E}[N\mu_i]\mathbb{E}[N\mu_j]}{\sqrt{\mathbb{E}[(N\mu_i)^2] - \mathbb{E}[N\mu_i]^2} \cdot \sqrt{\mathbb{E}[(N\mu_j)^2] - \mathbb{E}[N\mu_j]^2}} = -\frac{4}{4p + 2\sqrt{p} - 3} \approx -\frac{1}{p}, \quad (\text{B26})$$

which tends to zero with increasing  $p$  (unlike for the state maximizing QFI, where numbers of photons in different arms are strongly anti-correlated). Together with the fact that the total cost of estimation for the state Eq. (B1) is very close to the fundamental bound, it strongly suggests that this state may have a lot in common with the function optimal for single phase estimation with fixed mean number of photons Eq. (A26). To see it directly, we compare the distributions of photons corresponding to both these cases, i.e. Eq. (B22) with:

$$|g(\mu)|^2 \propto \text{Ai}^2 \left( A_0 + \frac{2p|A_0|}{3} \mu \right). \quad (\text{B27})$$

From Fig. B2 one can see, that for big number of phases  $p$  both distributions are indeed similar.

### 3. Superiority of joint measurement for finite number of photons $N$ —discrete version of Eq. (16)

So far we have shown that for the model where discrete variables from the original problem  $m_i \in \{0, 1, \dots, N\}$  are replaced with continuous ones  $\frac{m_i}{N} \rightarrow \mu_i \in [0, 1]$ , and all the parameters are initially completely unknown  $\theta \in \mathbb{R}^p$ , the state of the form given in Eq. (16) offers a significant advantage:

$$\left[ \Delta^2 \hat{\theta} \right]_{\text{ansatz}}^{\text{jointly, continuous}} = \frac{p(1 + 2\sqrt{p})^2 \sqrt{p}(4p + 2\sqrt{p} - 1)}{N^2(8\sqrt{p} - 4)} \quad (\text{B28})$$

over the optimal separate strategy, for which the cost reads:

$$\left[ \Delta^2 \hat{\theta} \right]_{\text{optimal}}^{\text{separately, continuous}} = \frac{p^3 \pi^2}{N^2}. \quad (\text{B29})$$

Following the reasoning from [75] (where single phase estimation problem was analyzed) we infer that the results obtained for discrete model converge to the above in the limit  $N \rightarrow \infty$ . Still it is an interesting question to investigate this convergence in more detail for finite number  $N$ . Below we present the respective numerical analysis.

The output state reads:

$$|\Psi_{\theta}^N\rangle = \sum_{\mathbf{m}: |\mathbf{m}| \leq N} c_{\mathbf{m}} e^{i\mathbf{m} \cdot \theta} |\mathbf{m}\rangle, \quad \theta \in (-\pi, +\pi]^p. \quad (\text{B30})$$

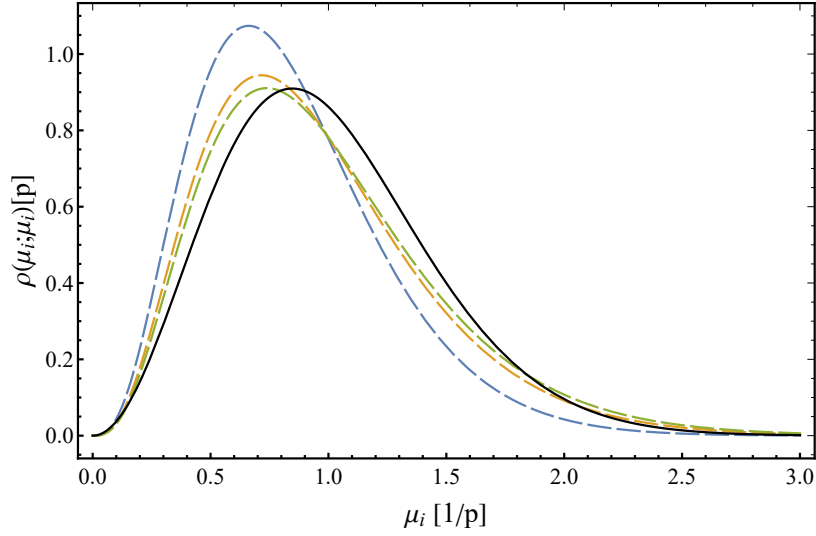


FIG. B2. See how the distribution of photons in single sensing arm changes with increasing number of phases. The black solid line is normalized square of corresponding Airy function Eq. (B27). The dashed lines correspond to single mode photons number distribution Eq. (B22) for  $p$  equal to 10 (blue), 50 (orange) and 250 (green).

Let us consider the covariant cost of the form

$$\sum_{i=1}^p (\min\{|\tilde{\theta}_i - \theta_i|, 2\pi - |\tilde{\theta}_i - \theta_i|\})^2, \quad (\text{B31})$$

which takes into account that two phases  $+\pi - \epsilon$ ,  $-\pi + \epsilon$  should be seen as arbitrary close when  $\epsilon \rightarrow 0$ . We make this choice in order not to keep the cost as similar as possible to the standard quadratic cost but at the same time respect the phase periodicity (covariant cost)—other common choice  $\sum_{i=1}^p 4 \sin^2\left(\frac{\tilde{\theta}_i - \theta_i}{2}\right)$  [34–36], is (except of the point  $|\tilde{\theta}_i - \theta_i| = 0$ ) strictly smaller than the variance and this fact would have an impact on the numerical results and demonstrating the convergence of the results would require going to much higher values of  $N$ . For such a problem, the minimal cost may be obtained by the covariant measurement:

$$|\chi_{\tilde{\theta}}\rangle = \frac{1}{\sqrt{(2\pi)^p}} \sum_{\mathbf{m}: |\mathbf{m}| \leq N} e^{i\mathbf{m} \cdot \tilde{\theta}} |\mathbf{m}\rangle \quad (\text{B32})$$

(note that in this case, unlike in continuous model, the measurement is not projective; still it is well defined as  $\int_{(-\pi, +\pi]^p} d\tilde{\theta} |\chi_{\tilde{\theta}}\rangle \langle \chi_{\tilde{\theta}}| = \mathbb{1}$ ). Then we have:

$$\Delta^2 \tilde{\theta} = \max_{\tilde{\theta} \in \Theta} \int_{(-\pi, +\pi]^p} d\tilde{\theta} |\langle \chi_{\tilde{\theta}} | \Psi_{\tilde{\theta}}^N \rangle|^2 \sum_{i=1}^p (\min\{|\tilde{\theta}_i - \theta_i|, 2\pi - |\tilde{\theta}_i - \theta_i|\})^2 = \sum_{i=1}^p \int_{(-\pi, +\pi]^p} d\tilde{\theta} |\langle \chi_{\tilde{\theta}} | \Psi_0^N \rangle|^2 \tilde{\theta}_i^2. \quad (\text{B33})$$

Let us now focus on  $i^{\text{th}}$  element of above sum:

$$\begin{aligned} & \int_{(-\pi, +\pi]^p} d\tilde{\theta} |\langle \chi_{\tilde{\theta}} | \Psi_0^N \rangle|^2 \tilde{\theta}_i^2 = \\ & = \frac{1}{(2\pi)^p} \int_{(-\pi, +\pi]^p} d\tilde{\theta} \sum_{\mathbf{m}} \sum_{\mathbf{m}'} c_{\mathbf{m}}^* c_{\mathbf{m}'} e^{i(\mathbf{m}' - \mathbf{m}) \cdot \tilde{\theta}} \tilde{\theta}_i^2 = \sum_{\mathbf{m}} \sum_{\mathbf{m}'} c_{\mathbf{m}}^* c_{\mathbf{m}'} \left( \prod_{j \neq i} \delta_{m_j, m'_j} \right) \cdot \begin{cases} (-1)^{m_i - m'_i} \frac{2}{(m_i - m'_i)^2} & \text{for } m_i \neq m'_i, \\ \frac{\pi^2}{3} & \text{for } m_i = m'_i \end{cases}, \end{aligned} \quad (\text{B34})$$

where in the last step discrete Fourier transform of  $\tilde{\theta}_i^2$  was applied.

Now we plug into the above formula the coefficients  $c_{\mathbf{m}}$  coming from the discretization of Eq. (16), and arrive at:

$$\begin{aligned} \left[ \Delta^2 \tilde{\theta} \right]_{\text{ansatz}}^{\text{jointly,discrete}} &= p \cdot \sum_{\mathbf{m}} \sum_{\mathbf{m}'} c_{\mathbf{m}}^* c_{\mathbf{m}'} \left( \prod_{j \neq i} \delta_{m_j, m'_j} \right) \cdot \begin{cases} (-1)^{m_i - m'_i} \frac{2}{(m_i - m'_i)^2} & \text{for } m_i \neq m'_i, \\ \frac{\pi^2}{3} & \text{for } m_i = m'_i \end{cases} \\ &\text{with } c_{\mathbf{m}} \propto f(\mathbf{m}/N) = \left( \prod_{i=1}^p m_i/N \right)^{\frac{3}{2}} \left( 1 - \sum_{i=1}^p m_i/N \right)^{\sqrt{p}}, \quad \sum_{\mathbf{m}} |c_{\mathbf{m}}|^2 = 1. \end{aligned} \quad (\text{B35})$$

In the left panel of Fig. B3, we investigate how Eq. (B35) converges to Eq. (B28) with the increasing number of photons  $N$  for different number of phases  $p$ . From the plot in may be seen that a good convergence (up to 1%) is obtained for  $N/p$  of the order 10.

To make the analysis complete, we should also investigate how the **advantage** of our strategy over optimal separate one converge to the advantage obtained in continuous case. A certain subtlety should be noted here. The minimal cost obtainable for separate strategy is given as:

$$\left[ \Delta^2 \tilde{\theta} \right]_{\text{optimal}}^{\text{separately,discrete}} = p \cdot \min_{c_{\mathbf{m}}} \left( \sum_{m=0}^{N/p} \sum_{m'=0}^{N/p} c_{\mathbf{m}}^* c_{\mathbf{m}'} \cdot \begin{cases} (-1)^{m-m'} \frac{2}{(m-m')^2} & \text{for } m \neq m', \\ \frac{\pi^2}{3} & \text{for } m = m' \end{cases} \right) \quad (\text{B36})$$

and, in principle, it may be **smaller** than Eq. (B29). The reason for this is that for finite  $N$  the cost depend on the size of  $\Theta$  – which is finite in discrete case, and infinite in continuous. As the consequence, superiority of our ansatz over optimal strategy cannot be judged directly from convergence Eq. (B35) to Eq. (B28). Therefore in the right panel of Fig. B3 we present the convergence of the advantage itself, e.i. how  $\frac{[\Delta^2 \tilde{\theta}]_{\text{ansatz}}^{\text{jointly,discrete}}}{[\Delta^2 \tilde{\theta}]_{\text{optimal}}^{\text{separately,discrete}}}$  converge to  $\frac{[\Delta^2 \tilde{\theta}]_{\text{ansatz}}^{\text{jointly,continuous}}}{[\Delta^2 \tilde{\theta}]_{\text{optimal}}^{\text{separately,continuous}}}$ . Comparing both panels one see that convergence of the advantage is indeed slower than convergence of the cost; still, for  $N/p = 16$  the advantage is observed for any number of phases.

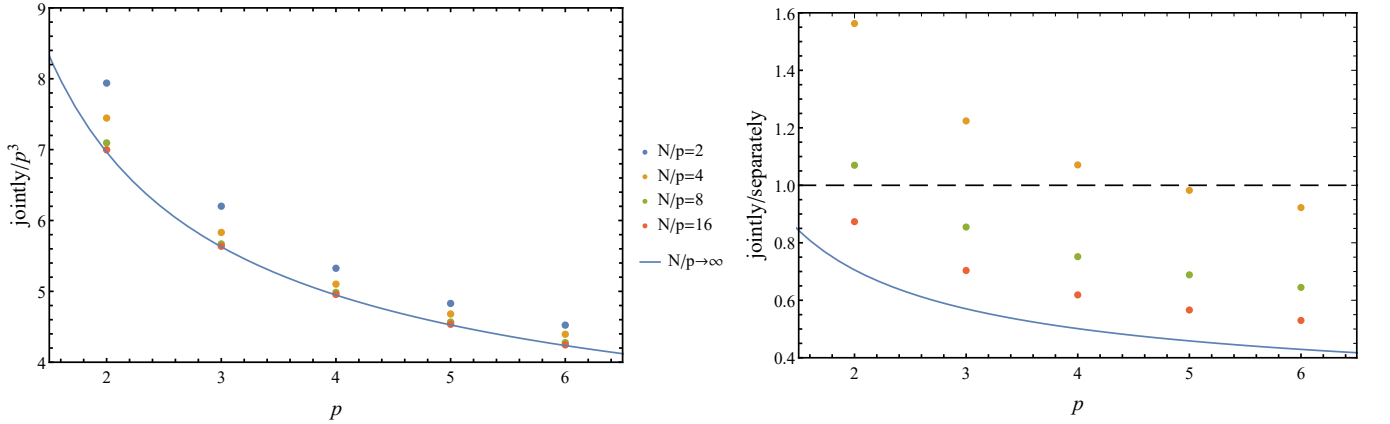


FIG. B3. In the left panel we present how the cost obtained by applying the Ansatz for discrete model  $[\Delta^2 \tilde{\theta}]_{\text{ansatz}}^{\text{jointly,discrete}}$  Eq. (B35) converges to the one from the continuous model  $[\Delta^2 \tilde{\theta}]_{\text{ansatz}}^{\text{jointly,continuous}}$  Eq. (B28) with increasing number of photons  $N$ . The cost obtained for different number of phases is rescaled in the plot by the factor  $1/p^3$  in order to increase the visibility. We see that a good agreement is obtained for  $N/p$  of order 10. In the right panel we perform a similar analysis for the advantage of the ansatz over the optimal separate strategy—  $[\Delta^2 \tilde{\theta}]_{\text{ansatz}}^{\text{jointly,discrete}} / [\Delta^2 \tilde{\theta}]_{\text{optimal}}^{\text{separately,discrete}}$ , where the advantage is achieved for a ratio value below 1 (black dashed line). One can see that for  $N/p = 16$  the advantage is observed for any number of phases  $p$ . In both cases blue lines corresponding to asymptotic limits are plotted as continuous for better visibility (but they are defined only for natural  $p$ ).

#### 4. Applying to local estimation problem

Now let us discuss how one may apply above procedure to initial problem, where  $\theta \in \Theta$  (where  $\Theta$  is some small finite-sized neighborhood of  $\theta_0$ ) and standard quadratic cost is considered. Without loss of generality let as fixed  $\theta_0 = 0$  (otherwise

one could apply proper phase shift  $-\theta_{0_i}$  in each interferometer's arm) and then choose the smallest cube containing  $\Theta$ , e.i.  $\Theta \in [-d/2, d/2]^p$  (in the opposite to  $[-\delta/2, +\delta/2]^p$  laying inside of  $\Theta$ ). Then we apply the covariant measurement Eq. (B32), but each time when the results of measurement point outside of region  $[-d/2, d/2]^p$ , we choose as the indication of the estimator proper point of the border of this cube. More formally, let us label the measurement outcomes by  $\mathbf{x} \in (-\pi, +\pi]^p$ :

$$|\mathbf{x}\rangle = \frac{1}{\sqrt{(2\pi)^p}} \sum_{\mathbf{m}:|\mathbf{m}|\leq N} e^{i\mathbf{m}\cdot\mathbf{x}} |\mathbf{m}\rangle. \quad (\text{B37})$$

The estimator  $\tilde{\boldsymbol{\theta}}(\mathbf{x}) = [\tilde{\theta}_1(\mathbf{x}), \tilde{\theta}_2(\mathbf{x}), \dots, \tilde{\theta}_p(\mathbf{x})]$  is then:

$$\tilde{\theta}_i(\mathbf{x}) = \begin{cases} x_i & \text{for } x_i \in [-d/2, +d/2], \\ -d/2 & \text{for } x_i \in (-\pi, -d/2), \\ +d/2 & \text{for } x_i \in (+d/2, +\pi], \end{cases} \quad (\text{B38})$$

By its construction such procedure for any  $\boldsymbol{\theta} \in [-d/2, +d/2]^p$  (assuming  $d \leq \pi$ ) gives smaller mean cost than the original covariant one discussed previously.

We expect that the difference disappear with increasing  $N$ , as we expect that probability of pointing measurement outside of  $[-d/2, +d/2]$  decrease faster than  $1/N^2$  (however, we did not prove it analytically). Still, for any finite  $N$  results obtained for covariant case bound from above the cost of this strategy.

Therefore, based on the numerical results presented in Fig. B3 and the analytical bound Eq. (11) we may therefore say, that for any non-degenerate  $\Theta$ , in the limit of large  $N$ :

$$\forall_{\Theta} 1.89p^3 \leq \lim_{N \rightarrow \infty} N^2 \Delta^2 \tilde{\boldsymbol{\theta}} \lesssim \frac{p(1 + 2\sqrt{p})^2 \sqrt{p}(4p + 2\sqrt{p} - 1)}{8\sqrt{p} - 4}. \quad (\text{B39})$$