Adaptive Bayesian quantum algorithm for phase estimation

Joseph G. Smith,[1,2,*] Crispin H. W. Barnes,[1,†] and David R. M. Arvidsson-Shukur[2,‡]
[1]*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[2]*Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

(Received 19 May 2023; revised 22 March 2024; accepted 25 March 2024; published 11 April 2024)

Quantum-phase-estimation algorithms are critical subroutines in many applications for quantum computers and in quantum-metrology protocols. These algorithms estimate the unknown strength of a unitary evolution. By using coherence or entanglement to sample the unitary N_{tot} times, the variance of the estimates can scale as $O(1/N_{\text{tot}}^2)$, compared to the best "classical" strategy with $O(1/N_{\text{tot}})$. The original algorithm for quantum phase estimation cannot be implemented on near-term hardware as it requires large-scale entangled probes and fault-tolerant quantum computing. Therefore, alternative algorithms have been introduced that rely on coherence and statistical inference. These algorithms produce quantum-boosted phase estimates without interprobe entanglement. This family of phase-estimation algorithms have, until now, never exhibited the possibility of achieving optimal scaling $O(1/N_{\text{tot}}^2)$. Moreover, previous works have not considered the effect of noise on these algorithms. Here, we present a coherence-based phase-estimation algorithm which can achieve the optimal quadratic scaling in the mean absolute error and the mean squared error. In the presence of noise, our algorithm produces errors that approach the theoretical lower bound. The optimality of our algorithm stems from its adaptive nature: Each step is determined, iteratively, using a Bayesian protocol that analizes the results of previous steps.

DOI: 10.1103/PhysRevA.109.042412

I. INTRODUCTION

Quantum phenomena can be used to improve measurement techniques beyond what is achievable with classical techniques using similar resources. One example of this is quantum phase estimation. Quantum phase estimation is used as the following: a subroutine of many quantum algorithms [1–3]; a component in gravitational-wave detection [4]; a method to measure time [5]; and as a tool to compute ground-state energies [6,7]. Phase estimation is the estimation of an unknown phase, θ, that a unitary operation $\hat{U}(\theta)$ applies to a quantum state. Traditional techniques evolve separable probes via $\hat{U}(\theta)$ and can lead to an estimate with variance bounded by the standard quantum limit $O(1/N_{\text{tot}})$, where N_{tot} is the number of unitary applications [8,9]. If multiple probes are entangled instead, estimates can be made with variance bounded by the Heisenberg limit $O(1/N_{\text{tot}}^2)$ [10]. However, the most trivial applications of entangled states in phase estimation yield ambiguous estimates due to the multivalued nature of the inverse functions used [11]. Precise, unambiguous estimates of θ require the use of a phase-estimation algorithm. Such algorithms achieve what is known as point identification of θ's estimate [12].

The goal of phase-estimation algorithms is to make the best estimate of θ given a certain amount of resources, typically the total number of times $\hat{U}(\theta)$ is applied. The best-known example is the quantum-phase-estimation algorithm (QPEA) [13,14]. The QPEA requires large-scale fully entangled states, the application of several controlled versions of $\hat{U}(\theta)$, and the ability to implement the inverse Fourier-transform algorithm [15]. Although the QPEA attains some Heisenberg scaling errors and point identification, its requirements put stringent limits on which systems or platforms it can be successfully implemented on [16,17]. This stringency has lead to the development of less-cumbersome phase-estimation algorithms based on statistical inference [18–22].

Inference-based algorithms achieve a quantum enhancement by constructing distributions of possible candidate values for the estimate of θ by iteratively sampling multiple circuits. These circuits can either be selected before the algorithm commences [11,20–22] or adaptively as the algorithm runs [23–28]. The circuits constitute several applications of $\hat{U}(\theta)$, but do not necessitate interprobe entanglement, loosening the requirements for physical implementation. A shortcoming of the inference-led phase-estimation algorithms is that few have considered the effect of noise present in realistic devices. Current quantum devices belong to the class of noisy intermediate-scale quantum (NISQ) devices and are prone to environmental noise: qubits suffer from preparation and measurement noise and from circuit-induced decoherence. Such noise often degrades the acquired information and reduces the quantum enhancement from a quadratic boost to a constant-factor improvement [29].

In this paper, we present an adaptive Bayesian phase-estimation algorithm. This algorithm obtains, in the limit of

*jgs46@cam.ac.uk
†chwb101@cam.ac.uk
‡drma2@cam.ac.uk

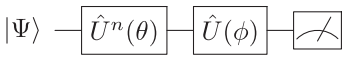

FIG. 1. The quantum circuit (n, ϕ).

large N_{tot}, the Heisenberg scaling of two errors, the mean absolute error, and the mean squared error, when noise is absent, and a theoretical lower bound on errors when noise is present. Our algorithm attains this performance through adaptive iteration, executing a series of circuits with parameters depending on the results of previous measurements. These circuits require one probe at a time and, hence, no interprobe entanglement [30]. We also demonstrate through numerical simulations that our Bayesian algorithm outperforms other phase-estimation algorithms in these error metrics.

II. BAYESIAN PHASE ESTIMATION

We consider the unitary operation $\hat{U}(\theta)$ with at least two eigenstates $|\psi_0\rangle$ and $|\psi_1\rangle$, such that

$$\hat{U}(\theta)|\psi_0\rangle = |\psi_0\rangle\,,$$
$$\hat{U}(\theta)|\psi_1\rangle = e^{i\theta}|\psi_1\rangle\,. \tag{1}$$

Here, $\theta \in \Theta = [0, 2\pi]$ is the unknown phase to be estimated. When measuring θ, it is optimal to do so with probes in an equal superposition of $|\psi_0\rangle$ and $|\psi_1\rangle$ [8]: $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\psi_0\rangle + |\psi_1\rangle)$. Here, we focus on such optimal phase estimation and consider only two-level subspaces of potentially d-dimensional unitaries.

An estimate of θ can be achieved through analyzing the output of the quantum circuit in Fig. 1. A single probe is prepared in the state $|\Psi\rangle$ and evolved coherently through $\hat{U}(\theta)$ a number n times. The probe is then evolved a single time through a known phase-shift $\hat{U}(\phi)$ before being measured in the basis $\{|\Psi\rangle, |\Psi^{\perp}\rangle\}$. [Throughout this paper, we use (n, ϕ) to denote such a circuit]. This measurement process maximizes the quantum Fisher information (QFI) and leads to asymptotically optimal measurements [10,31,32]. The circuit (n, ϕ) is sampled ν times, and the number of probes in the state $|\Psi\rangle$ is recorded as x. This process applies $\hat{U}(\theta)$ a number $n\nu$ times. The probability of x taking a certain value is given by the binomial distribution

$$p(x|n, \phi, \nu, \theta) = \binom{\nu}{x}[p_0(\theta, n, \phi)]^x[1 - p_0(\theta, n, \phi)]^{\nu - x}. \tag{2}$$

Here, $p_0(\theta, n, \phi)$ is the probability of measuring a single probe in the state $|\Psi\rangle$. This work considers the following sources of noise: Depolarization, such that each gate creates the target state with probability $1 - \lambda$, and the maximally mixed state with probability of λ; preparation errors, such that the optimal initial state $|\Psi\rangle$ is prepared with probability $1 - \eta$ and an orthogonal state is incorrectly prepared with probability η; and measurement errors, such that the output state is recorded correctly with a probability of $1 - \xi$. Our noise model leads to a probability of recording the state $|\Psi\rangle$ after the circuit runs of

$$p_0(\theta, n, \phi) = \frac{1}{2} + \frac{\alpha\beta^n}{2}\cos(n\theta + \phi), \tag{3}$$

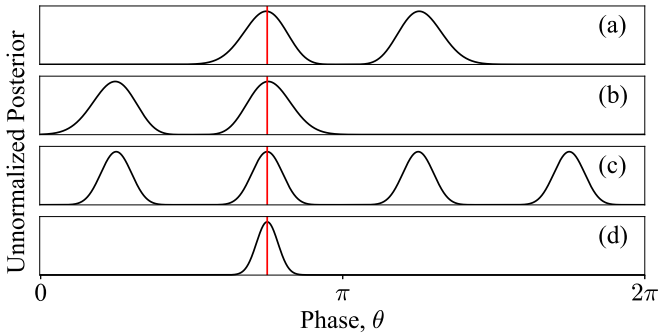

FIG. 2. Unnormalized posterior distribution generated from 20 applications of $\hat{U}(\theta)$ in the circuits (a) (1,0), (b) $(1, \pi/2)$, and (c) (2,0), with $\theta = 3\pi/4$ [vertical line]. (d) By combining the results from (a)–(c), erroneous peaks can be reduced in size.

where $\alpha = (1 - 2\eta)(1 - 2\zeta)$ and $\beta = (1 - \lambda)$. Note that $\alpha = \beta = 1$ in noiseless systems.

After observing x, Bayesian inference gives a posterior distribution for θ [33],

$$p(\theta|n, \phi, \nu, x) = \frac{p(x|n, \phi, \nu, \theta)\pi(\theta)}{\int_{\Theta} p(x|n, \phi, \nu, \theta)\pi(\theta)d\theta}, \tag{4}$$

where $\pi(\theta)$ is any prior knowledge of θ. This posterior can be used to produce an estimator of θ, $\hat{\theta}$. Examples of such estimators include the minimum-mean-squared-error (MMSE) estimator, $\hat{\theta}_{\text{MMSE}} = \mathbb{E}[\theta]$, and the maximum-*a posteriori* (MAP) estimator $\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p(\theta|n, \phi, \nu, x)$. Some estimators can have biases due to the lack point identification [12]. Lack of point identification occurs when an estimator has multiple values: $\hat{\theta}_{\text{MAP}}$ has $2n$ values because $\cos^{-1}(n\theta)$ is multivalued in the interval Θ. This ambiguity is demonstrated in Figs. 2(a) to 2(c): the posterior distribution has $2n$ symmetric peaks centered at either $\theta_c = \theta + \frac{2\pi l_1}{n}$ or $\theta_c = -\theta - \frac{2\phi}{n} + \frac{2\pi l_2}{n}$, where l_1 and l_2 are integers.

Point-identifying $\hat{\theta}_{\text{MAP}}$ (i.e., selecting the correct peak) can be achieved by executing several circuits, $(\boldsymbol{n}, \boldsymbol{\phi}) \equiv \{(n_1, \phi_1), \ldots, (n_m, \phi_m)\}$. Each circuit is executed a number $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)$ times, respectively, resulting in observations $\boldsymbol{x} = (x_1, \ldots, x_m)$. This new measurement strategy applies $\hat{U}(\theta)$ a total $\sum_{j=1}^{m} n_j\nu_j$ times. Combining observations provides the posterior distribution

$$p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) = \frac{\pi(\theta)\prod_{i=1}^{m} p(x_i|n_i, \phi_i, \nu_i, \theta)}{\int_{\Theta}\pi(\theta)\prod_{i=1}^{m} p(x_i|n_i, \phi_i, \nu_i, \theta)d\theta}. \tag{5}$$

Careful selection of $\boldsymbol{n}$, $\boldsymbol{\phi}$, and $\boldsymbol{\nu}$ can ensure that $\hat{\theta}_{\text{MAP}}$ is unique. This is demonstrated in Fig. 2(d), which combines the observations of Figs. 2(a) to 2(c). The erroneous peak at $2\pi - \theta$ can be removed by sampling two circuits with differing values of ϕ [11].

Performance of estimators

The performance of $\hat{\theta}$ can be judged through a distance from the true value of θ, calculated through a loss function $L(\hat{\theta}, \theta)$. Due to the periodicity of phases, $\hat{\theta} = \hat{\theta} + 2\pi l$ for integer l. Distance from θ to $\hat{\theta}$ should be taken as the minimum

loss for each possible $\hat{\theta}$:

$$L(\hat{\theta},\theta)=\min_{l\in\mathbb{Z}}\{L(\hat{\theta}+2\pi l,\theta)\}. \tag{6}$$

In this work, we consider two loss functions: The absolute error, $|\hat{\theta}-\theta|$, with average mean absolute error (MAE), $\mathcal{L}_{\mathrm{MAE}}(\hat{\theta})$; The squared error, $(\hat{\theta}-\theta)^2$, with average mean squared error (MSE), $\mathcal{L}_{\mathrm{MSE}}(\hat{\theta})$. The average of a loss function is taken over the posterior distribution when the true value of θ is unknown:

$$\mathcal{L}(\hat{\theta})=\mathbb{E}[L(\hat{\theta},\theta)]=\int_{\Theta}p(\theta|\boldsymbol{n},\boldsymbol{\phi},\boldsymbol{\nu},\boldsymbol{x})L(\hat{\theta},\theta)d\theta. \tag{7}$$

A lower value of $\mathcal{L}(\hat{\theta})$ indicates that, on average, $\hat{\theta}$ is close to the likely values of θ. Therefore, phase-estimation algorithms aim to minimize $\mathcal{L}(\hat{\theta})$ for a specified $L(\hat{\theta},\theta)$.

$\mathcal{L}(\hat{\theta})$ is reduced by reducing the number of, and width of, the peaks in the posterior distribution. The number of peaks can be reduced through point identification. The widths of the peaks relates to the statistical variance σ_θ^2 which can be estimated by considering the limit of many circuit executions, $\nu\to\infty$: If the single circuit (n,ϕ) is executed ν times, the binomial distribution $p(x|n,\phi,\nu,\theta)$ tends to a normal distribution [34]

$$p(x|n,\phi,\nu,\theta)=N(\bar{x},\sigma_x)=N(\nu p_0,\sqrt{\nu p_0(1-p_0)}). \tag{8}$$

$p(\theta|n,\phi,\nu,x)$ is then a sum of normal distributions with centres θ_c and identical variances σ_θ^2. σ_θ^2 can be estimated through linear error propagation

$$\sigma_\theta^2=\left(\frac{\partial\theta}{\partial\bar{x}}\right)^2\sigma_x^2\equiv\frac{1-\alpha^2\beta^{2n}\cos^2(n\theta+\phi)}{\alpha^2\beta^{2n}\sin^2(n\theta+\phi)n^2\nu}. \tag{9}$$

A normal distribution centered on $\hat{\theta}$ with variance σ_θ^2 has $\mathcal{L}_{\mathrm{MAE}}(\hat{\theta})=\sqrt{2\sigma_\theta^2/\pi}$ and $\mathcal{L}_{\mathrm{MSE}}(\hat{\theta})=\sigma_\theta^2$.

In the absence of noise, $\sigma_\theta^2=1/n^2\nu$ independent of θ and ϕ. Restricting the number of times $\hat{U}(\theta)$ can be applied to N_{tot} times, a depth-one circuit without multiple coherent unitary applications achieves $\sigma_\theta^2=1/N_{\mathrm{tot}}$, the standard quantum limit (SQL). A point-identified estimate with a single peak in the posterior has $\mathcal{L}_{\mathrm{MAE}}(\hat{\theta})=\sqrt{2/\pi N_{\mathrm{tot}}}$ and $\mathcal{L}_{\mathrm{MSE}}(\hat{\theta})=1/N_{\mathrm{tot}}$. Instead, utilizing coherent applications of multiple instances of the unknown unitary, by using a circuit of depth $n\propto N_{\mathrm{tot}}$, achieves a quadratic speed-up with $\sigma_\theta^2=O(1/N_{\mathrm{tot}}^2)$, the optimal Heisenberg limit (HL) [35]. A point-identified estimate with a single peak in the posterior will also demonstrate quadratic speed-ups of $\mathcal{L}_{\mathrm{MAE}}(\hat{\theta})=\sqrt{2/\pi N_{\mathrm{tot}}^2}$ and $\mathcal{L}_{\mathrm{MSE}}(\hat{\theta})=1/N_{\mathrm{tot}}^2$. For large N_{tot}, this scaling requires unrealistic infinitely long circuits. In real devices, circuit depth is restricted by some limiting value n_{lim}. Executing the circuit (n_{lim},ϕ) results in $\sigma_\theta^2=O(1/N_{\mathrm{tot}})$ which is the SQL.

When noise is present ($\beta<1$), $\sigma_\theta^2\to\infty$ for $n\to\infty$. Therefore, deep circuits, and algorithms using them, give estimates with high uncertainties. Instead of letting the circuit depth increase indefinitely with N_{tot}, our algorithm samples the circuit

$$(n_{\mathrm{opt}},\phi_{\mathrm{opt}})\equiv\left(\left\lfloor-\frac{1}{2\ln\beta}\right\rceil,\frac{\pi}{2}-n_{\mathrm{opt}}\theta\right), \tag{10}$$

to minimize σ_θ^2 to $-\frac{2e\ln\beta}{\alpha^2N_{\mathrm{tot}}}$. Then, σ_θ^2 scales with the SQL, suggesting the quantum advantage has become a prefactor rather than a quadratic speedup. The loss of the quadratic speedup is derived for more general noise models in Ref. [29]. A complication with finding this optimal circuit is that the value of ϕ_{opt} depends on the unknown value of θ. Therefore, one has to rely on an adaptive scheme to find ϕ_{opt}.

The aforementioned variance is often quoted as the figure of merit of phase-estimation algorithms [10]. However, this is only appropriate if the correct peak (out of the $2n$ peaks) is picked out with certainty. This selection is typically achieved by sampling shallower circuits alongside the optimal circuit. Doing this requires the applications of $\hat{U}(\theta)$, and thus prevents σ_θ^2 from reaching its minimum value for a given amount of resources.

An alternative way to judge performance is to calculate the probability that θ belongs to a credible interval $\Theta_i\subseteq\Theta$:

$$\Pr[\theta\in\Theta]=\int_{\Theta_i}p(\theta|n,\phi,\nu,x)d\theta. \tag{11}$$

This credible-interval analysis alone is a poor way to judge performance of $\hat{\theta}$ due to the ignorance of the tail of the distribution outside Θ_i. This tail may lead to a finite probability of the distance between the true θ and $\hat{\theta}$ being large. Therefore, the main focus of this paper is producing a phase-estimation algorithm that minimizes an average loss rather than focusing solely on credible intervals.

III. ADAPTIVE PHASE-ESTIMATION ALGORITHM

From the above analysis, circuits with $n>1$ may lead to narrower posterior peaks. However, this requires the execution of shallower circuits to point-identify the estimate. This point-identification process requires the use of resources in a suboptimal way. To reduce the potential waste of resources, previous noiseless phase-estimation algorithms have suggested doubling circuit depths [11,20,22,36,37]. However, these algorithms tend to be nonadaptive, either executing suboptimal circuits excessively or requiring circuits beyond hardware capabilities. Furthermore, some algorithms require measurements of the probes along multiple axes, which may lead to further resources being wasted on these suboptimal circuits.

To prevent potential resource waste, we present an adaptive algorithm. The algorithm uses the outcomes of previous measurements to predict which circuit to execute next to minimize the final value of $\mathcal{L}(\hat{\theta})$ after all remaining resources, N_{left}, are used. This prediction is achieved by calculating $p(\theta|\boldsymbol{n},\boldsymbol{\phi},\boldsymbol{\nu},\boldsymbol{x})$ and $\hat{\theta}$ either after each measurement or after a batch of measurements. The algorithm is outlined in Algorithm1 and pictorially in Fig. 3. As shown in the Appendix, this algorithm achieves quadratic speedups of both $\mathcal{L}_{\mathrm{MAE}}(\hat{\theta})$ and $\mathcal{L}_{\mathrm{MSE}}(\hat{\theta})$ when compared to classical techniques in the noiseless regime.

The point-identification process occurs by executing circuits with doubling depth until some optimal cutoff: At the ith step, the circuit

$$(n_i,\phi_i)=\left(\min\{2^{i-1},n_{\mathrm{opt}},n_{\mathrm{lim}}\},\frac{\pi}{2}-n_i\hat{\theta}\right) \tag{12}$$

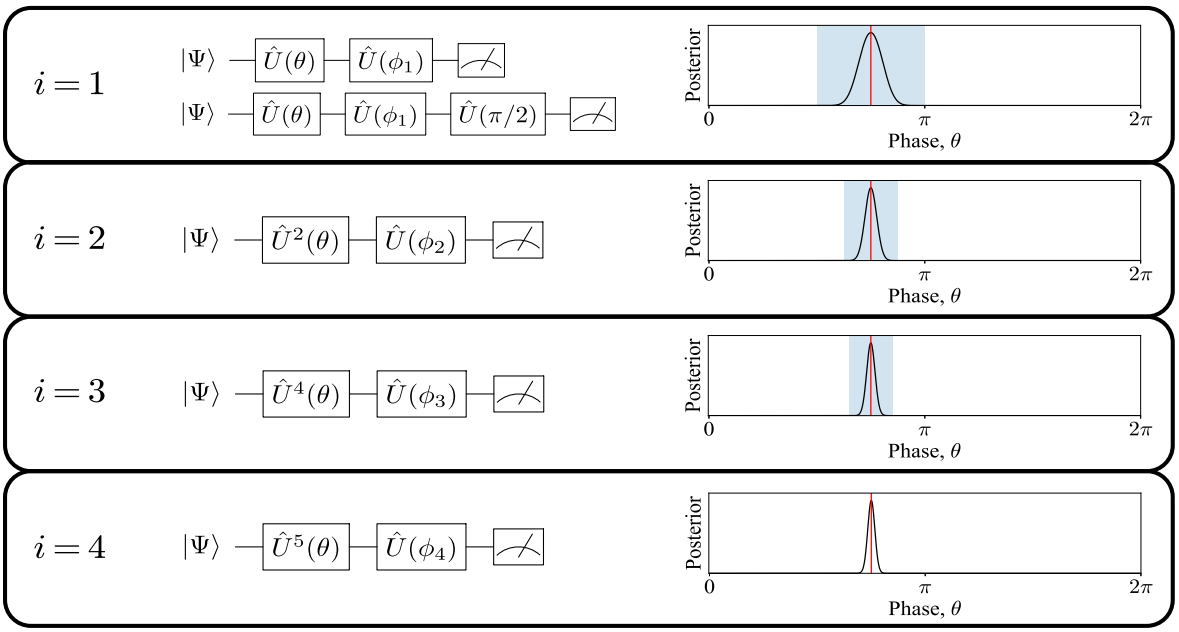

FIG. 3. The circuits executed at each iteration step of our algorithm and the posterior distribution at the end of the step. The initial conditions were $\theta = \frac{2\pi}{3}$ (vertical line), $N_{\text{tot}} = 300$ and $\beta = 0.9$ ($n_{\text{opt}} = 5$). The regions Θ_i are shaded.

is executed enough times to restrict θ to an interval Θ_i with a probability of at least $1 - \epsilon_i$, where

$$\Theta_i = \left[\hat{\theta} - \frac{\pi}{2n_{i+1}}, \hat{\theta} + \frac{\pi}{2n_{i+1}}\right] \tag{13}$$

and $\epsilon_i = (n_i/N_{\text{tot}})^3$. The interval Θ_i is chosen such that the posterior distribution generated from execution of the following circuit, (n_{i+1}, ϕ_{i+1}), contains only one peak, thus, point-identifying the estimate. The fewest executions of the suboptimal shallower circuit (n_i, ϕ_i) required for the above restriction will occur when the posterior peak variance σ_θ^2 is minimized: $\phi_i = \pi/2 - n_i\theta$ and $\theta_c = \theta + \pi l/n_i$ for integer l. As the true θ is unknown, we use $\hat{\theta}$ to determine ϕ. The first step of the algorithm requires two circuits being sampled, $(1, \phi_1)$ and $(1, \phi_1 + \pi/2)$, to remove the initial peak at $2\pi - \theta$. The combination of these two circuits is equivalent to measuring the probe along both the z axis and x axis if probes are single qubits. We also redefine $\Theta = [\hat{\theta} - \pi, \hat{\theta} + \pi]$ for the normalization of intervals to avoid potential issues with discontinuous intervals.

Adaptivity is also used to determine which circuit is executed next after we establish $\Pr[\theta \in \Theta_i] \geqslant 1 - \epsilon_i$. Using the current posterior distribution, we calculate the predicted loss when using all of the remaining resources to execute either the circuit (n_i, ϕ_i), with loss $\hat{\mathcal{L}}_i(\hat{\theta})$, or (n_{i+1}, ϕ_{i+1}), with loss $\hat{\mathcal{L}}_{i+1}(\hat{\theta})$. The circuit with the smallest predicted loss is then executed. If (n_{i+1}, ϕ_{i+1}) is (is not) selected, the iteration proceeds with $i \leftarrow i + 1$ [iteration concludes with the circuit (n_i, ϕ_i) executed until all remaining resources are used up]. Losses are predicted by estimating expected outcomes

$$\hat{x}_i = \int_\Theta \nu_i p_0(\theta, n_i, \phi_i) p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) d\theta, \tag{14}$$

where $\nu_i = \lfloor N_{\text{left}}/n_i \rfloor$ is the maximum number of times the circuit (n_i, ϕ_i) can be executed. $\hat{x}_i$ is used to generate a new posterior distribution from which the expected loss $\hat{\mathcal{L}}_i$ is calculated. Our adaptive algorithm does not suffer from the same shortfalls as other nonadaptive phase-estimation algorithms. (1) Other algorithms that use fixed intervals instead of our adaptively changing single interval Θ_i may require a large number of circuit executions to place θ, with high confidence, within one of the intervals. Indeed if θ lies on the boundary, circuits must be executed infinite number of times [11]. (2) Other algorithms that use predecided constant numbers of circuit executions may waste resources building overconfidence placing θ in an interval if θ is far from an interval boundary. (3) Previous algorithms lack a cutoff depth, such that circuits with depths exceeding n_{opt} are sampled. (4) Previous algorithms do not take into account hardware limitations on circuit depths. (5) Beyond the first circuit, probes are only measured

ALGORITHM 1. Adaptive algorithm.

1: Set $i \leftarrow 1$, $N_{\text{left}} \leftarrow N_{\text{tot}}$, $p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) \leftarrow \pi(\theta)$.
2: **while** $N_{\text{left}} > n_i$ **do**
3: **while** $\Pr[\theta \in \Theta_i] < 1 - \epsilon_i$ **do**
4: Execute circuit (n_i, ϕ_i).
5: If $i = 1$, execute circuit $(1, \phi_i + \pi/2)$.
6: Update N_{left}, $p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x})$ and $\hat{\theta}$.
7: **end while**
8: Calculate $\hat{\mathcal{L}}_i(\hat{\theta})$ and $\hat{\mathcal{L}}_{i+1}(\hat{\theta})$.
9: If $\hat{\mathcal{L}}_{i+1}(\hat{\theta}) < \hat{\mathcal{L}}_i(\hat{\theta})$, set $i \leftarrow i + 1$ and return to step 3. Else, go to step 11.
10: **end while**
11: Execute the circuit (n_i, ϕ_i) a number $\lfloor N_{\text{left}}/n_i \rfloor$ times.
12: Execute circuit $(1, \phi_1)$ a number N_{left} mod n_i times to use up any remaining resources

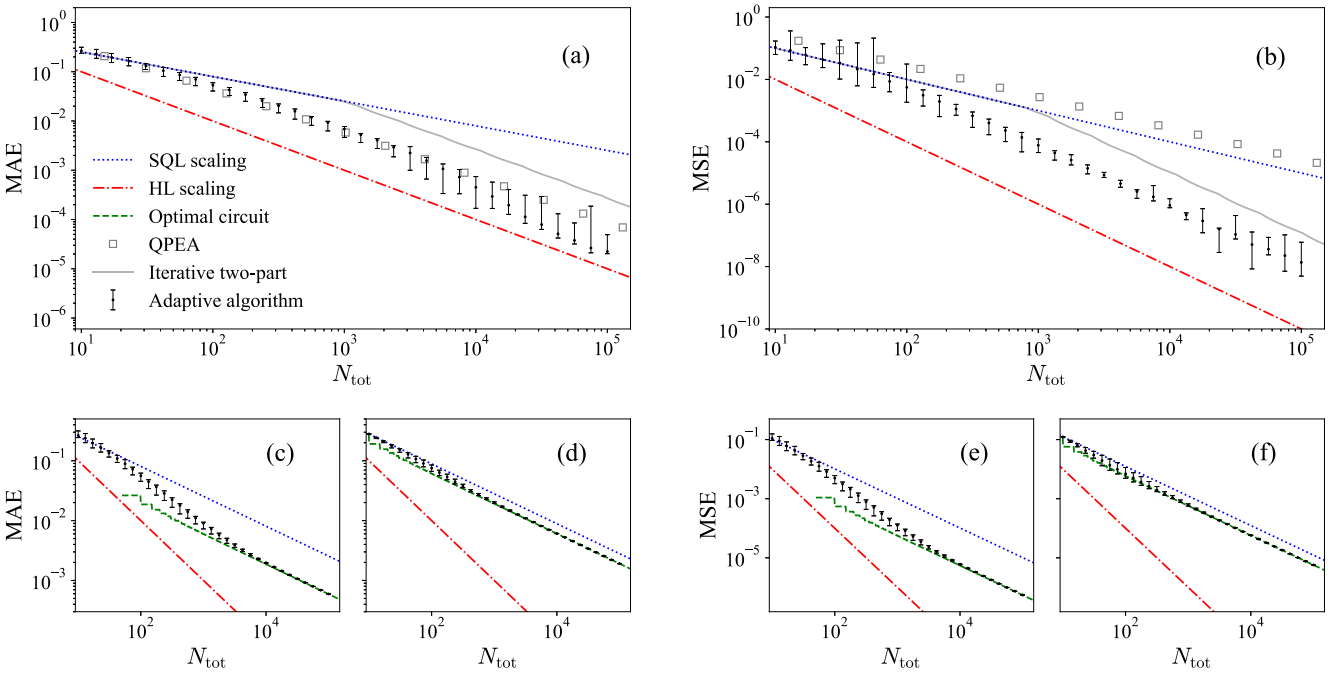

FIG. 4. Simulated values of the MAE and MSE against total number of unitary applications for different phase-estimation algorithms with $\alpha = 1$ and (a), (b) $\beta = 1$; (c), (e) $\beta = 0.99$; (d), (f) $\beta = 0.9$. Error bars are plotted between the minimum and maximum average losses achieved after 1000 samples.

along one axis in our algorithm, so less resources are used for point identification than in other algorithms.

Our algorithm also reduces the classical computing times of other adaptive algorithms. Algorithms based on Bayesian experimental designs calculate an average of $\mathcal{L}(\hat{\theta})$ over all potential outcomes, based on the current posterior [23,24]. This average-loss calculation may require evaluating many costly integrals for each potential measurement. Furthermore, if every circuit is allowed to be executed at any stage of the algorithm, huge classical computing costs after each measurement are expected to calculate all risks and determine which circuit to use. Our algorithm circumvents this issue by evaluating $\mathcal{L}(\hat{\theta})$ of the most likely outcomes for two circuits only, and also delays these less-costly computations until the condition $\Pr[\theta \in \Theta_i] \geqslant 1 - \epsilon_i$ is met.

IV. PERFORMANCE

We benchmark the performance of our adaptive algorithm through both analytic bounds for large N_{tot} and numerical simulation for finite values of N_{tot}. The analytic bounds are derived in the Appendix by considering the maximum contributions from both the bias and width of the posterior. Our adaptive algorithm achieves a quadratic speedup of both $\mathcal{L}_{\text{MAE}}(\hat{\theta})$ and $\mathcal{L}_{\text{MSE}}(\hat{\theta})$ when compared to classical techniques in the noiseless regime. When $\beta < 1$ or n_{lim} is finite, both $\mathcal{L}_{\text{MAE}}(\hat{\theta})$ and $\mathcal{L}_{\text{MSE}}(\hat{\theta})$ achieved by our adaptive algorithm tend towards the performance of a point-identified optimal circuit.

The results of numerical simulations are presented in Fig. 4. The adaptive algorithm is evaluated for $\alpha = 1$, varying β, and 1000 uniformly spread values of $\theta \in [0, 2\pi]$ for a fixed $N_{\text{tot}} \in [10^1, 10^5]$, assuming no hardware limitations and the uniform prior $\pi(\theta) = 1/2\pi$. The target loss is either the absolute error [Figs. 4(a), 4(c) and 4(d)] or the squared error [Figs. 4(b), 4(e) and 4(f)], with $\hat{\theta} = \hat{\theta}_{\text{MMSE}}$ for both cases. Alongside these data, we plot the performance from similar simulations of the quantum phase-estimation algorithm (QPEA) [13] and the nonadaptive iterative two-part protocol [11]. Also plotted are the SQL, the HL, and the theoretical limit from sampling the optimal circuit, all when the posterior has a single peak.

For a noiseless system ($\beta = 1$), our algorithm achieves a sub-SQL average MAE for roughly $N_{\text{tot}} \geqslant 20$ and a sub-SQL average MSE for roughly $N_{\text{tot}} \geqslant 40$. The iterative two-part algorithm has sub-SQL MAE and MSE for N_{tot} greater than approximately 1000. This demonstrates that our adaptive algorithm achieves a quantum advantage with fewer resources used than the nonadaptive algorithm. Although having a sub-SQL MAE at $N_{\text{tot}} \geqslant 31$, the MAE of the QPEA diverges from the HL line showing suboptimal scaling. In addition, the MSE of the QPEA shows SQL-like scaling. These observations of the QPEA are due to the long tail of the posterior distribution [38].

In Figs. 4(c) to 4(f), we plot the MAE and MSE when noise does scale with circuit depth: $\beta < 1$. Our adaptive algorithm tends to to the optimal limit when N_{tot} is large. We also observe a reduced spread of the MAE and MSE achieved for larger N_{tot}, indicating the algorithms ability to find the optimal circuit.

V. CONCLUSION

We present an adaptive algorithm for quantum phase estimation based on Bayesian inference. Our algorithm samples a series of circuits with parameters determined, iteratively,

by the outcomes of previous measurements. The algorithm operates to minimize the predicted expected loss for a given amount of resources. Our algorithm achieves the Heisenberg scaling of both the mean absolute error and the mean squared error in the noiseless setting. Furthermore, in the presence of noise, our algorithm's performance tends to the optimal error limit imposed by the statistical variance. An experimental implementation of this adaptive algorithm is the subject of future research.

ACKNOWLEDGMENTS

The authors thank N. Mertig and W. Salmon for enlightening discussions. This work was supported by Hitachi, Lars Hierta's Memorial Foundation, and Girton College, Cambridge.

APPENDIX: ASYMPTOTIC SCALING

We now derive upper bounds of both the MAE and MSE achieved by our adaptive algorithm when N_{tot} is large and $\pi(\theta)$ is uniform. We assume N_{tot} is large enough for all circuits up to and including the optimum circuit executed. At the conclusion of our algorithm, circuits (n_1, ϕ_1) to (n_m, ϕ_m) are executed ν_1 to ν_m times, with

$$n_i = \begin{cases} \min\{n_{\text{opt}}, n_{\text{lim}}\} & i = m \;\&\; n_{\text{lim}} \neq \infty \text{ or } \beta < 1, \\ 2^{i-1} & \text{otherwise.} \end{cases} \tag{A1}$$

For $\epsilon_i = (n_i/N_{\text{tot}})^3$, the circuit (n_i, ϕ_i) must be executed a large number of times to ensure $\Pr[\theta \in \Theta_i] \geqslant 1 - \epsilon_i$. The posterior distribution generated by the circuit (n_i, ϕ_i) in this limit is the normalized sum of normal distributions

$$\sum_c N(\theta_c, \sigma_{\theta,i}) \equiv \sum_{l=0}^{2n_i-1} N\left(\theta + \frac{\pi l}{n_i}, \frac{1}{\alpha\beta^{n_i} n_i \sqrt{\nu_i}}\right), \tag{A2}$$

as described above, provided $\hat{\theta}$ is a consistent estimator.

After the ith stage of iteration, the circuits up to and including (n_i, ϕ_i) are executed and the value of θ is restricted to the interval Θ_i with a confidence at least $1 - \epsilon_i$. This process discards the interval $\bar{\Theta}_i = \Theta_{i-1} - \Theta_i$. The posterior distribution inside Θ_i at this stage is

$$\prod_{j=1}^{i} N(\theta, \sigma_{\theta,j}) = N(\theta, \Sigma_i) \equiv N\left(\theta, \frac{1}{\sqrt{\sum_{j=1}^{i} \sigma_{\theta,j}^{-2}}}\right), \tag{A3}$$

which will be sharply peaked at the value of θ. After all circuits have been executed, the total average loss is

$$\begin{aligned} \mathcal{L}(\hat{\theta}) &= \int_{\Theta_m} p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) L(\hat{\theta}, \theta) d\theta \\ &\quad + \sum_{i=1}^{m-1} \int_{\bar{\Theta}_i} p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) L(\hat{\theta}, \theta) d\theta \\ &\leqslant \int_{\Theta_m} p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) L(\hat{\theta}, \theta) d\theta \\ &\quad + \sum_{i=1}^{m-1} \max_{\bar{\Theta}_i}\{L(\hat{\theta}, \theta)\} \int_{\bar{\Theta}_i} p(\theta|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{x}) d\theta \\ &\leqslant \mathcal{L}_{\Theta_m}(\hat{\theta}) + \sum_{i=1}^{m-1} \max_{\bar{\Theta}_i}\{L(\hat{\theta}, \theta)\}(\epsilon_{i-1} - \epsilon_i), \end{aligned} \tag{A4}$$

where $\mathcal{L}_{\Theta_m}(\hat{\theta})$ is the average loss from the normal distribution inside Θ_m. The sum over all $\bar{\Theta}_i$ is the maximal contribution of the bias on $\mathcal{L}(\hat{\theta})$. The quantity $\max_{\bar{\Theta}_i}\{L(\hat{\theta}, \theta)\}$ occurs when θ is at the edge of $\bar{\Theta}_i$ as this maximizes distance from $\hat{\theta}$. Using the absolute error as the loss leads to

$$\max_{\bar{\Theta}_i}\{L(\hat{\theta}, \theta)\} = \begin{cases} \pi & i = 1, \\ \frac{\pi}{2n_{i+1}} & i \neq 1, \end{cases} \tag{A5}$$

and using the squared error as the loss leads to the square of this. These results give upper bounds

$$\begin{aligned} \mathcal{L}_{\text{MAE}}(\hat{\theta}) &\leqslant \sqrt{\frac{2}{\pi}}\Sigma_m + \frac{\pi}{2}\left(\epsilon_1 + \frac{\epsilon_{m-1}}{n_m} + \frac{1}{2}\sum_{i=1}^{m-2}\frac{\epsilon_i}{2^i}\right), \\ \mathcal{L}_{\text{MAE}}(\hat{\theta}) &\leqslant \Sigma_m^2 + \frac{\pi^2}{4}\left(3\epsilon_1 + \frac{\epsilon_{m-1}}{n_m^2} + \frac{3}{4}\sum_{i=1}^{m-2}\frac{\epsilon_i}{4^i}\right), \end{aligned} \tag{A6}$$

where Σ_m^2 is the variance of the normal distribution inside Θ_m.

To derive an expression for Σ_m, we must first find a relationship between ϵ_i and ν_i for $i < m$. The maximum value of ν_i required to achieve $\Pr[\theta \in \Theta_i] \geqslant 1 - \epsilon_i$ can be derived using the Chernoff-Cramér bound for normal distributions [14]

$$\Pr[|\hat{\theta} - \theta| \geqslant \delta_i] = \epsilon_i \leqslant 2\exp\left[-\delta_i^2/\Sigma_i^2\right], \tag{A7}$$

where $\delta_i = \frac{\pi}{2n_{i+1}}$ is half the length of the interval Θ_i. By rearranging, we find the largest value of ν_i is when $\Sigma_i^{-2} = \frac{8n_{i+1}^2}{\pi^2}\ln(\frac{2}{\epsilon_i})$. We can also note that $\sigma_{\theta,i}^{-2} = \Sigma_i^{-2} - \Sigma_{i-1}^{-2}$, which leads to

$$\sigma_{\theta,i}^2 = \frac{8n_{i+1}^2}{\pi^2}\left[\ln\left(\frac{2}{\epsilon_i}\right) - \frac{1}{4}\ln\left(\frac{2}{\epsilon_{i-1}}\right)\right]. \tag{A8}$$

Equating this to Eq. (9),

$$\nu_i = \frac{32}{\pi^2\alpha^2\beta^{2n_i}}\left[\ln\left(\frac{2}{\epsilon_i}\right) - \frac{1}{4}\ln\left(\frac{2}{\epsilon_{i-1}}\right)\right]. \tag{A9}$$

The total number of times $\hat{U}(\theta)$ is applied is fixed to N_{tot}. Therefore,

$$\begin{aligned} n_m\nu_m &= N_{\text{tot}} - \sum_{i=1}^{m-1}\frac{2^{i+4}}{\pi^2\alpha^2\beta^{2^i}}\left[\ln\left(\frac{2}{\epsilon_i}\right) - \frac{1}{4}\ln\left(\frac{2}{\epsilon_{i-1}}\right)\right] \\ &= N_{\text{tot}} - \frac{72}{\pi^2\alpha^2}\sum_{i=1}^{m-1}\left(\frac{2}{\beta^2}\right)^i \ln\left(\frac{2N_{\text{tot}}}{2^i}\right). \end{aligned} \tag{A10}$$

$n_m\nu_m$ is used to find Σ_m:

$$\begin{aligned} \Sigma_m^{-2} &= \Sigma_{m-1}^{-2} + \sigma_{\theta,m}^{-2} \\ &= \frac{8n_m^2}{\pi^2}\ln\left(\frac{2N_{\text{tot}}^3}{8^{m-2}}\right) + \alpha^2\beta^{2n_m}n_m^2\nu_m. \end{aligned} \tag{A11}$$

As we only care about the scalings of losses for large N_{tot}, we only focus on the greatest scaling terms. When $\beta < 1$ or n_{lim} is finite, m remains fixed for all large N_{tot}. The sum in Eq. (A10) will scale as $O(\ln N_{\text{tot}})$, leading to $n_n \nu_m = O(N_{\text{tot}})$ and $\Sigma_m^{-2} = O(N_{\text{tot}})$.

If, however, $\beta = 1$ and $n_{\text{lim}} \to \infty$, m varies with N_{tot}. m must scale no larger than $O(\log_2 N_{\text{tot}})$ to ensure ν_m in Eq. (A10) is not negative. Assume now $m = O(\log_2 N_{\text{tot}})$, so $N_{\text{tot}} = O(2^m)$. $\Sigma_m^{-2} = O(2^m N_{\text{tot}})$ and Eq. (A6) leads to

$$\mathcal{L}_{\text{MAE}}(\hat{\theta}) = \begin{cases} O\left(\frac{1}{\sqrt{N_{\text{tot}}}}\right) & \beta < 1 \text{ or } n_{\text{lim}} \neq \infty, \\ O\left(\frac{1}{N_{\text{tot}}}\right) & \text{otherwise,} \end{cases} \tag{A12}$$

for the MAE and

$$\mathcal{L}_{\text{MSE}}(\hat{\theta}) = \begin{cases} O\left(\frac{1}{N_{\text{tot}}}\right) & \beta < 1 \text{ or } n_{\text{lim}} \neq \infty, \\ O\left(\frac{1}{N_{\text{tot}}^2}\right) & \text{otherwise,} \end{cases} \tag{A13}$$

for the MSE. These are optimal scalings [29]. If N_{tot} scales with m as a function larger than $O(2^m)$, these noiseless losses would have a worse scaling. However, these larger scalings require a smaller m and a shallower circuit depth cutoff. The adaptive algorithm would have predicted the outcomes of selecting this smaller m (where the estimates are exact due to the sharpness of the posterio) and decided this smaller m leads to a larger loss than the larger m. Therefore, the larger m would have been selected, leading to the choice of $N_{\text{tot}} = O(2^m)$.

Classical techniques that involve sampling circuits of depth one lead to a posterior that is a normal distribution with variance $O(1/N_{\text{tot}})$ [Eq. (9)]. This normal distribution has a MAE of $O(1/\sqrt{N_{\text{tot}}})$ and a MSE of $O(1/N_{\text{tot}})$. Therefore, our algorithm achieves a quadratic speedup in ideal cases. When $\beta < 1$ or $n_{\text{lim}} \neq \infty$, the speedup is a constant. However, we now argue that the loss tends to the same value achieved by sampling a point-identified optimal circuit alone [Eq. (10)]. The N_{tot} term will outgrow the sum in Eq. (A10) such that $\nu_m \approx N_{\text{tot}}/n_m$ for large-enough N_{tot}. In this case, $\Sigma_m^{-2} \approx \alpha^2 \beta^{2n_m} n_m N_{\text{tot}}$, the same as if all resources were used sampling the optimal circuit. The terms in the loss functions proportional to ϵ_i will scale as $O(1/N_{\text{tot}}^3)$ which is much less than Σ_m. Therefore, the losses are bounded by

$$\begin{aligned} \mathcal{L}_{\text{MAE}}(\hat{\theta}) &\leqslant \sqrt{\frac{2}{\pi}} \Sigma_m = \sqrt{\frac{2}{\pi \alpha^2 \beta^{2n_m} n_m N_{\text{tot}}}}, \\ \mathcal{L}_{\text{MAE}}(\hat{\theta}) &\leqslant \Sigma_m^2 = \frac{1}{\alpha^2 \beta^{2n_m} n_m N_{\text{tot}}}, \end{aligned} \tag{A14}$$

the theoretical lower bound if the optimal circuit is used.

[1] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, Contemp. Math. **305**, 53 (2002).
[2] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, Quantum Inf. Process. **19**, 75 (2020).
[3] P. Shor, *Proceedings of the 35th FOCS* (IEEE, New York, 1994), pp. 124–134.
[4] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboiration), Phys. Rev. Lett. **116**, 061102 (2016).
[5] D. Oblak, P. G. Petrov, C. L. G. Alzar, W. Tittel, A. K. Vershovski, J. K. Mikkelsen, J. L. Sørensen, and E. S. Polzik, Phys. Rev. A **71**, 043807 (2005).
[6] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Mol. Phys. **109**, 735 (2011).
[7] L. Lin and Y. Tong, PRX Quantum **3**, 010318 (2022).
[8] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).
[9] Note that this work concerns asymptotic phase estimation, where N_{tot} is large; nonasymptotic strategies are discussed in [39].
[10] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Ann. Phys. (NY) **247**, 135 (1996).
[11] J. G. Smith, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Phys. Rev. A **106**, 062615 (2022).
[12] A. Lewbel, J. Econ. Lit. **57**, 835 (2019).
[13] A. Yu. Kitaev, arXiv:quant-ph/9511026.
[14] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
[15] Y. S. Weinstein, M. A. Pravia, E. M. Fortunato, S. Lloyd, and D. G. Cory, Phys. Rev. Lett. **86**, 1889 (2001).
[16] H. Mohammadbagherpoor, Y.-H. Oh, A. Singh, X. Yu, and A. J. Rindos, arXiv:1903.07605.
[17] F. Chapeau-Blondeau and E. Belin, Signal Processing **170**, 107441 (2020).
[18] K. M. Svore, M. Hastings, and M. Freedman, Quantum Inf. Comput. **14**, 306 (2014).
[19] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, Phys. Rev. Lett. **118**, 100503 (2017).
[20] T. Rudolph and L. Grover, Phys. Rev. Lett. **91**, 217905 (2003).
[21] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, Phys. Rev. Lett. **118**, 190502 (2017).
[22] F. Belliardo and V. Giovannetti, Phys. Rev. A **102**, 042613 (2020).
[23] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New J. Phys. **14**, 103013 (2012).
[24] F. Belliardo, V. Cimini, E. Polino, F. Hoch, B. Piccirillo, N. Spagnolo, V. Giovannetti, and F. Sciarrino, arXiv:2211.04747.
[25] M. Valeri, V. Cimini, S. Piacentini, F. Ceccarelli, E. Polino, F. Hoch, G. Bizzarri, G. Corrielli, N. Spagnolo, R. Osellame, and F. Sciarrino, Phys. Rev. Res. **5**, 013138 (2023).
[26] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).
[27] M. A. Rodríguez-García, I. P. Castillo, and P. Barberis-Blostein, Quantum **5**, 467 (2021).
[28] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett. **85**, 5098 (2000).
[29] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, Nat. Commun. **3**, 1063 (2012).
[30] The probes themselves may require entanglement between multiple qubits, and so the algorithm cannot be guaranteed as entanglement free.
[31] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[32] An alternative QFI-maximizing measurement involves a Greenberger-Horne-Zeilinger (GHZ) state [8]. However, owing to the extreme noise sensitivity of GHZ states [40], we only consider coherent unitary applications in this paper.

[33] D. V. Lindley, *Bayesian Statistics: A Review* (SIAM, Philadelphia, PA, 1972).

[34] A. Ly, M. Marsman, J. Verhagen, R. P. P. P. Grasman, and E.-J. Wagenmakers, J. Math. Psychol. **80**, 40 (2017).

[35] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. Lett. **105**, 180402 (2010).

[36] M. Dobšíček, G. Johansson, V. Shumeiko, and G. Wendin, Phys. Rev. A **76**, 030306(R) (2007).

[37] D. Grinko, J. Gacon, C. Zoufal, and S. Woerner, npj Quant. Inf. **7**, 52 (2021).

[38] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature (London) **450**, 393 (2007).

[39] W. Salmon, S. Strelchuk, and D. Arvidsson-Shukur, Quantum **7**, 998 (2023).

[40] P. Hyllus, O. Gühne, and A. Smerzi, Phys. Rev. A **82**, 012337 (2010).