Efficient Bayesian Phase Estimation

Nathan Wiebe[1] and Chris Granade[2,3]

[1]*Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA*
[2]*Centre for Engineered Quantum Systems, Sydney, New South Wales 2006, Australia*
[3]*School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia*

We introduce a new method called rejection filtering that we use to perform adaptive Bayesian phase estimation. Our approach has several advantages: it is classically efficient, easy to implement, achieves Heisenberg limited scaling, resists depolarizing noise, tracks time-dependent eigenstates, recovers from failures, and can be run on a field programmable gate array. It also outperforms existing iterative phase estimation algorithms such as Kitaev's method.
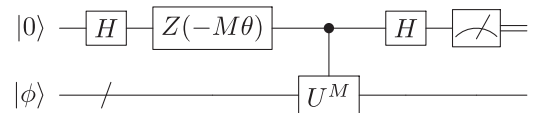
Introduction.—Eigenvalue estimation has been a cornerstone of physics since the dawn of spectroscopy. In recent years, ideas from quantum information have revolutionized the ways that we estimate these values by providing methods that require exponentially fewer experiments than statistical sampling and total experimental time that saturates the Heisenberg limit [1–8], which is the best possible scaling allowed by quantum mechanics. This quantum approach, known as phase estimation, is crucial to achieving many of the celebrated speedups promised by quantum computing [9–13].

While great progress has been made towards optimizing phase estimation, a number of gaps remain. Most notably, existing approaches either require a large number of ancilla qubits, are classically inefficient, or fail to extract information about the phase as efficiently as their classically inefficient brethren. Moreover, the majority of these methods are not analyzed in the presence of noise or other experimental imperfections. Finally, these approaches typically assume that the phase is static, or quasistatic, throughout the experiment. This precludes phase estimation from tracking time-dependent drifts in phase. These gaps mean that many phase estimation algorithms are not known to be robust to common errors present in modern day experiments.

Here we address these shortcomings by proposing an approach to iterative phase estimation that is based on a new inference method that we call *rejection filtering*, and by providing an efficient heuristic for adaptively selecting near-optimal experiments. We call the phase estimation algorithm that follows from rejection filtering phase estimation, or *RFPE*. Our algorithm numerically approximates Bayesian inference on the measurement outcomes to estimate the most likely phase and its uncertainty. Unlike the numerical methods for Bayesian inference studied in [14,15], the method has miniscule memory requirements which permit it to be employed in a memory limited environment like a field programmable gate array

(FPGA). Our algorithm also inherits from Bayesian inference the ability to follow drifting eigenphases without needing to explicitly model the dynamics of the phase. Finally, we propose a novel method for detecting and correcting inference errors that allows the method to track a state's depolarization in real time. These new capabilities constitute an important advance for the field and underscore the relevance of approximate Bayesian inference to physics.

Before going into detail about our algorithm we will first review the iterative phase estimation circuit and Bayesian approaches to phase estimation. Iterative phase estimation learns the eigenvalue of a given eigenvector of a unitary matrix $U \in \mathbb{C}^{2^n \times 2^n}$ from a set of experiments that are performed on the quantum circuit [16]

where H is the Hadamard gate, $U|\phi\rangle = e^{i\phi}|\phi\rangle$ for an unknown eigenphase $\phi \in \mathbb{R}$, $|\phi\rangle \in \mathbb{C}^{2^n}$, $M \in \mathbb{N}$ specifies how many applications of U are applied conditioned on the ancilla qubit being 1, and where $Z(M\theta) = e^{iM\theta Z}$ for the Pauli operator Z is a rotation that can be used to compare the eigenphase of U to a known reference value. Most iterative phase estimation algorithms use this circuit to infer the bits in a binary expansion of ϕ. The process can be made optimal (up to $\log^*$ factors [8]) and also classically efficient.

Bayesian phase estimation.—Bayesian phase estimation involves performing a set of experiments and then updating a prior distribution over the true eigenphase using Bayes' rule [2,8,17]. For example, if an experiment is performed using M repetitions of U and using $Z(-M\theta)$, if a measurement outcome $E \in \{0, 1\}$ is observed then Bayes' rule states that the *posterior probability* distribution for ϕ after observing the datum is

$$P(\phi|E;\theta,M) = \frac{P(E|\phi;\theta,M)P(\phi)}{\int P(E|\phi;\theta,M)P(\phi)d\phi}. \qquad (1)$$

The likelihood function $P(0|\phi;\theta,M)$, is

$$P(0|\phi;\theta,M) = \frac{1+\cos(M[\phi+\theta])}{2},$$

$$P(1|\phi;\theta,M) = \frac{1-\cos(M[\phi+\theta])}{2}. \qquad (2)$$

Note that we have assumed perfect visibility of measurements in this likelihood function; that assumption can be readily relaxed [18,19], such that we consider the perfect visibility case for clarity. After using (1) to update the posterior distribution we then set the prior distribution to equal the posterior distribution. This process is then repeated for each of the experiments in the data set.

Bayesian inference returns a posterior distribution over the phase rather than a point estimate. The mean and standard deviation of this distribution provide an estimate of the true eigenvalue and the algorithm's uncertainty in that value. More sophisticated estimates of uncertainty, such as credible regions, can also be extracted from the posterior distribution [14,20].

Approximate Bayesian phase estimation.—Exact Bayesian inference is impossible if the eigenphases are continuous so approximations are needed to make the procedure tractable. Rather than naïvely discritizing the prior distribution, modern methods discretize by sampling from the prior and then perform Bayesian inference on the discrete set of samples (often called "particles"). These particle filter methods are quite powerful and have become a mainstay in computer vision and machine learning [21–23].

Despite their power, these methods are hard to implement and even harder to deploy in a memory limited environment such as on an embedded controller or an FPGA. With the increasing use of FPGAs in the control of quantum information experiments [24–26], overcoming this limitation presents a significant advantage.

RFPE follows a much simpler approach. Rather than using a set of hypotheses that implicitly define a model for the system, we posit a prior model and directly update it to find a model for our posterior distribution. We achieve this by using a Gaussian distribution with mean μ and variance σ^2 to model our initial prior, perform a Bayesian update on samples drawn from the distribution, and then refit the updated samples to a Gaussian distribution. This strategy is used in a number of particle filter methods such as the extended Kalman filter and assumed density filtering [21,27]. We further optimize this process by approximating the Bayesian inference step using rejection sampling. This can reduce the memory required by a factor of 1000 or more, as we need only consider a single sample at a time. Our algorithm is described in the following, and the pseudocode is given in the Supplemental Material [28]: (1) Perform the

experiment for given θ, M and observe the outcome $E \in \{0, 1\}$. (2) Draw m samples from $\mathcal{N}(\phi|\mu, \sigma^2)$. (3) For each sample ϕ_j, assign ϕ_j to Φ_{accept} with probability $P(E|\phi_j;\theta,M)/\kappa_E$, where $\kappa_E \in (0, 1]$ is a constant such that $P(E|\phi_j;\theta,M)/\kappa_E \leq 1$ for all ϕ_j, E. (4) Return $\mu = \mathbb{E}(\Phi_{\text{accept}})$ and $\sigma = \sqrt{\mathbb{V}(\Phi_{\text{accept}})}$.

The resultant samples are equivalent to those drawn from the posterior distribution $P(\phi|E;M,\theta)$. To see this, note that the probability density of a sample being accepted at $\phi = \phi_j$ is $P(E|\phi;\theta,M)\mathcal{N}(\phi|\mu,\sigma^2)$. Equation (1) then implies

$$P(E|\phi;\theta,M)\mathcal{N}(\phi|\mu,\sigma^2) \propto P(\phi|E;\theta,M), \qquad (3)$$

which implies that the accepted samples are drawn from the posterior distribution.

Although it is difficult to concretely predict the value of m needed to make the error in the inference small, we show in the Supplemental Material [28] that m must scale at least as the inverse square of the relative fluctuations in the likelihood function. Similarly, Markov's inequality shows that m must scale at least as $\kappa_E/\int P(E|\phi;\theta,M)P(\phi)d\phi$ to ensure that, with high probability, the mean can be accurately estimated from the accepted samples. We introduce κ_E to compensate for small expected likelihoods.

The main issue that remains is how to optimally choose the parameters θ and M. One approach is to locally optimize the Bayes risk [14], but the resulting calculation can be too expensive to carry out in online experiments that provide experimental results at a rate of tens of megahertz or faster. This can be avoided using precomputed measurement settings [35] or *policies* [7,36], which can substantially reduce the computational time needed to perform an online inference. These approaches *a priori* limit the number of measurements to be performed and can require significant memory to store the policy.

Fortunately, the particle guess heuristic (PGH) can give expedient and near-optimal experiments for this class of likelihood functions in an online fashion, and with no precomputation required [37],

$$M = \left\lceil \frac{1.25}{\sigma} \right\rceil, \qquad \theta \sim P(\phi). \qquad (4)$$

The constant 1.25 is chosen as it works well empirically. It also approximately corresponds to the optimal choice of $M = \sqrt{2+2\sqrt{5}}/2\epsilon \approx 1.27/\epsilon$ under the assumption that the Bayesian Cramér-Rao bound is saturated and Gaussian priors are used (see the Supplemental Material [28]).

Figure 1 shows the error incurred using RFPE. The most obvious feature is that the error shrinks exponentially with the number of experiments (which is proportional to the evolution time under the PGH) for $m > 100$. Roughly 150 experiments are needed for the method to provide 32 bits of
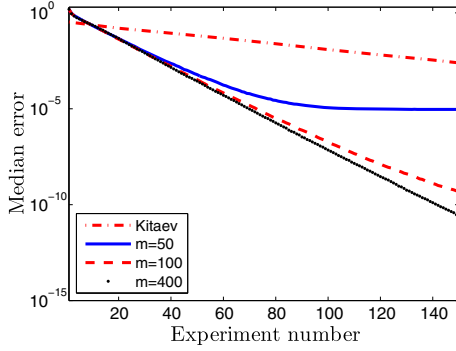
FIG. 1.   Median errors in phase estimation for 10 000 random initial choices of the true eigenphase.

FIG. 2.   Comparison of RFPE to ITPE for $t = 10\,000$ with 100 samples for $\phi_{\text{true}} = 2\pi k/t$ at each measurement.

accuracy in the median case. We discuss the scaling in the mean in the Supplemental Material [28].

The number of experiments needed by RFPE scales as $O\big(\log(1/\epsilon)\big)$, where ϵ is the target uncertainty, rather than the $O\big(\log(1/\epsilon)\log\log(1/\epsilon)\big)$ scaling of Kitaev's method [1,3]. RFPE should therefore offer advantages for implementing Shor's algorithm in small quantum devices. Concretely, after 150 experiments the median error for Kitaev's algorithm (with $s = 10$) is roughly 10^7 times that of RFPE. $O\big(\log(1/\epsilon)\big)$ scaling is further observed over 100 orders of magnitude in ϵ in the Supplemental Material [28] and is optimal (as seen from a simple counting argument).

Figure 2 compares RFPE to the information-theory phase estimation (ITPE) method of Svore *et al.* [8]. Although ITPE is inefficient and is nonadaptive, it is exact and so it is a natural benchmark to compare RFPE against. We find ITPE requires nearly 5 times the applications of U if $\phi = 2\pi k/t$ for integer $k < t$ and $t = 10\,000$.

Conversely, ITPE requires only 25 measurements to identify the phase with 50% probability, whereas RFPE requires 51 experiments if k is an integer. If the true value of k is real valued and ITPE is left unmodified, then ITPE fails to learn in the median because the long evolution times chosen lead to contradictory possibilities that prevent ITPE from converging to a unique answer. We correct this by choosing $M \to \lceil M/2 \rceil$ in such cases, which increases the number of experiments to 35 but also reduces the experimental time below that of unmodified ITPE. All three methods therefore trade-off experimental and computational resources differently. This is especially salient because ITPE is inefficient unlike RFPE.

In contrast to Shor's algorithm, in metrological and simulation applications the scaling of the error with the total number of applications of U needed (or in the case where M is continuous, the total evolution time) is the most significant metric. The Heisenberg bound states that the error optimally scales as $1/T = 1/\sum_i M_i$ here. Figure 3 shows that this scaling is saturated for RFPE. We further observe that the mean error, the root-mean square error, and a 95% confidence interval for the error follow an optimal
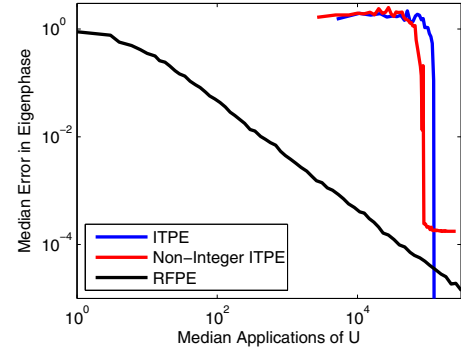
$O(1/T)$ scaling. RFPE therefore saturates optimal scaling in this setting as well.

Phase estimation with depolarizing noise.—A criticism that has been levied lately at phase estimation methods is that they can be impractical to execute on non–fault-tolerant quantum hardware [38–40]. This is because phase estimation attains its quadratic advantage over statistical sampling by using exponentially long evolutions. Decoherence causes the resultant phases to become randomized as time progresses, ultimately resulting in $\lim_{M\to\infty}P(0|\phi;\theta,M) = \frac{1}{2}$. Measurements convey no information in this limit. It similarly may be tempting to think that T_2 fundamentally limits the accuracy of phase estimation. Bayesian inference can, however, learn ϕ using $M \approx T_2$ [14] for some noise models.

We model the effect of decoherence on the system by assuming the existence of a *decoherence time* T_2 such that

$$P(0|\phi) = e^{-M/T_2}\left(\frac{1 + \cos(M[\phi - \theta])}{2}\right) + \frac{1 - e^{-M/T_2}}{2},$$

$$P(1|\phi) = e^{-M/T_2}\left(\frac{1 - \cos(M[\phi - \theta])}{2}\right) + \frac{1 - e^{-M/T_2}}{2}.$$
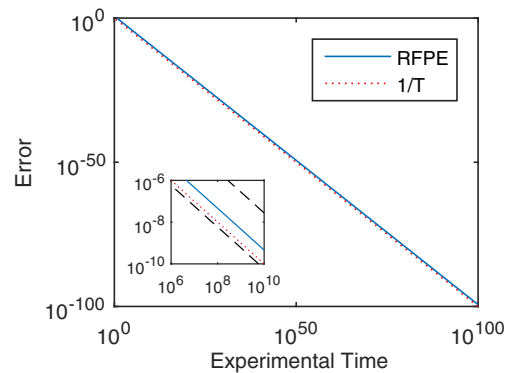
$$(5)$$

FIG. 3.   Median error vs evolution time for rejection filter phase estimation for the case where M is continuous and $T = \sum_i M_i$. The median error scales roughly as $4.7/T$, and the dashed lines in the inset figure gives a 95% confidence interval.
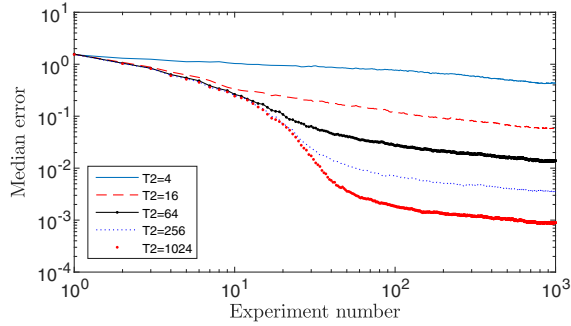
FIG. 4. Median errors for phase estimation in decohering systems for experiments constrained to use $M \leq T_2$. We take $m = 12\,000$ and use 1000 random samples to generate the data above. The initial state is taken to be a randomly chosen eigenstate in all cases.

This model is appropriate when a controlled U gate is slow relative to the H- and Z-rotation gates.

Since T_2 places a limitation on our ability to learn, we propose a variation to (4),

$$M = \min\left\{ \left\lceil \frac{1.25}{\sigma} \right\rceil, T_2 \right\}. \tag{6}$$

The experiments yielded by (6) resemble locally optimized experiments for frequency estimation, which saturate the Cramér-Rao bound [18]; however, (6) requires nearly 100-fold less computing time to select an experiment. The choice to cut off M at T_2 can be understood using the following argument. The Cramér-Rao bound for frequency estimation scales as $O(M^{-2})$ [41] in the absence of decoherence. Equation (5) suggests that decoherence causes the posterior variance to increase as $O(\exp(2M/T_2))$. Calculus reveals that $M = T_2$ optimally trades off these tendencies [18], which justifies (6).

Figure 4 shows that RFPE smoothly transitions between the exponential scaling expected at short times and the polynomial scaling expected when decoherence becomes significant. The error scales roughly as $1/N^{0.6}$ in this polynomial regime, which is comparable to the $1/\sqrt{N}$ scaling expected from statistical sampling. Depolarizing

noise, or diminished visibility [36], therefore does not *necessarily* limit the accuracy of phase estimation.

Tracking eigenphases.—Figure 4 shows the performance of RFPE when the initial quantum state is an eigenstate and is discarded after each experiment. Performing phase estimation in this way minimizes the number of experiments, but can be expensive if initial preparation is costly. In such cases, it makes sense to follow the standard prescription for phase estimation by keeping the quantum state until it is clear that the initial eigenstate has been depolarized. These depolarizations can cause RFPE to become confused because the new data that come in are only consistent with hypotheses that have been ruled out. We address this by performing inexpensive experiments to assess whether the state has depolarized and then restart the learning process. Restarting is also valuable when $T_2 = 0$ to recover from inference errors (see the Supplemental Material [28]).

The following procedure addresses this issue: (1) After each update with probability e^{-M/T_2} perform an experiment with $\theta = \mu$ and $M = \tau/\sigma$ for $\tau < 1$. (2) If result = 1 then prepare initial state and reset σ.

These steps perform a one–sided test of whether the prior distribution is consistent with the current state. If the prior probability distribution is correct, then the probability of measuring 0 is

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos^2\left(\frac{(\mu - x)\tau}{2\sigma} \right) e^{-[(\mu-x)^2/2\sigma^2]} dx$$
$$= \frac{1 + e^{-\tau^2/2}}{2}. \tag{7}$$

If $\tau = 0.1$ then this probability is approximately 0.998 and hence measuring 1 implies that the hypothesis that the prior is correct can be rejected at $p \leq 0.002$. A Bayesian analysis of our reset rule is given in the Supplemental Material [28].

The test process also permits the eigenvalue of an eigenstate in a decohering system to be estimated in real time. Figure 5 shows that the restarting algorithm can rapidly detect a transition from the instantaneous eigenstate and infer the system's new eigenphase.
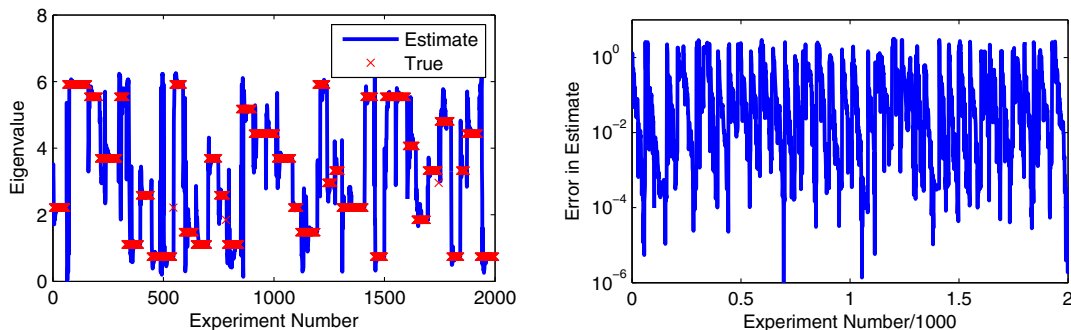

FIG. 5. Instantaneous estimate of the eigenphase for a system with 16 eigenvalues, $\Delta = 0$, $\tau = 0.1$, and $T_2 = 10^4$.

Conclusion.—Our efficient methods for Bayesian phase estimation represent an important step forward for iterative phase estimation (IPE). They provide a new platform upon which online inferences can be performed that scale optimally over thousands of experiments and can therefore be used to make tracking time-dependent properties of the system practical.

We also observe that our methods inherit the robustness to experimental imperfections that Bayesian approaches enjoy. This is especially critical as current experiments push past the classical regime and IPE becomes increasingly impractical in lieu of fault tolerance. In particular, the ability of our algorithms to learn in the presence of decoherence provides an efficient alternative to the variational eigensolvers used in present day experiments [38–40].

Looking at the problem of phase estimation more generally, it is clear that approximate Bayesian inference provides a powerful and flexible platform to perform phase estimation within. It is our firm belief that by combining such tools with clever experimental design, further improvements can be seen not only in phase estimation but also quantum metrology in general.

[1] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, Electron. Colloq. Comput. Complex. **3** (1996).

[2] D. W. Berry and H. M. Wiseman, Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry, Phys. Rev. Lett. **85**, 5098 (2000).

[3] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Providence, 2002), Vol. 47.

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

[5] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, How to perform the most accurate possible phase measurements, Phys. Rev. A **80**, 052114 (2009).

[6] G. Giedke, J. M. Taylor, D. D'Alessandro, M. D. Lukin, and A. Imamoğlu, Quantum measurement of a mesoscopic spin ensemble, Phys. Rev. A **74**, 032316 (2006).

[7] A. Hentschel and B. C. Sanders, Machine Learning for Precise Quantum Measurement, Phys. Rev. Lett. **104**, 063603 (2010).

[8] K. M. Svore, M. B. Hastings, and M. Freedman, Faster phase estimation, Quantum Inf. Comput. **14**, 306 (2014).

[9] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[10] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, Quantum amplitude amplification and estimation, Contemp. Math. **305**, 53 (2002).

[11] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).

[12] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[13] B. P. Lanyon *et al.*, Towards quantum chemistry on a quantum computer, Nat. Chem. **2**, 106 (2010).

[14] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online Hamiltonian learning, New J. Phys. **14**, 103013 (2012).

[15] M. P. V. Stenberg, Y. R. Sanders, and F. K. Wilhelm, Efficient Estimation of Resonant Coupling between Quantum Systems, Phys. Rev. Lett. **113**, 210404 (2014).

[16] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[17] D. W. Berry, H. M. Wiseman, and J. K. Breslin, Optimal input states and feedback for interferometric phase estimation, Phys. Rev. A **63**, 053804 (2001).

[18] C. Ferrie, C. E. Granade, and D. G. Cory, How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies, Quantum Inf. Process. **12**, 611 (2013).

[19] P. Cappellaro, Spin-bath narrowing with adaptive parameter estimation, Phys. Rev. A **85**, 030301 (2012).

[20] C. Ferrie, High posterior density ellipsoids of quantum states, New J. Phys. **16**, 023006 (2014).

[21] S. Haykin, *Kalman Filtering and Neural Networks* (John Wiley & Sons, New York, 2004), Vol. 47.

[22] A. Smith, A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer Science & Business Media, New York, 2013).

[23] M. Isard and A. Blake, CONDENSATION—Conditional density propagation for visual tracking, Int. J. Comput. Vis. **29**, 5 (1998).

[24] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby, Suppressing qubit dephasing using real-time Hamiltonian estimation, Nat. Commun. **5**, 5156 (2014).

[25] S. Casagrande, Masters' thesis, University of Waterloo, 2014, https://hdl.handle.net/10012/8281.

[26] J. M. Hornibrook, J. I. Colless, I. D. Conway Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J. D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly, Cryogenic Control Architecture for Large-Scale Quantum Computing, Phys. Rev. Applied **3**, 024010 (2015).

[27] M. Opper and O. Winther, A Bayesian approach to on-line learning, in *On-line Learning in Neural Networks*, edited by David Saad (Cambridge University Press, New York, 1998), p. 363.

[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.010503, which includes Refs. [29–34], for additional analysis of our reset rules, derivation of the Bayesian Cramer-Rao bound for RFPE, stability analysis and additional numerical experiments.

[29] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Quantum Hamiltonian learning using imperfect quantum resources, Phys. Rev. A **89**, 042314 (2014).

[30] C. Ferrie and C. E. Granade, Likelihood-free Methods for Quantum Parameter Estimation, Phys. Rev. Lett. **112**, 130402 (2014).

[31] S. Goodman, A dirty dozen: Twelve p-value misconceptions, Semin. Hematol. **45**, 135 (2008).

[32] R. Hoekstra, R. D. Morey, J. N. Rouder, and E.-J. Wagenmakers, Robust misinterpretation of confidence intervals, Psychon. Bull. Rev. **21**, 1157 (2014).

[33] R. D. Gill and B. Y. Levit, Applications of the van Trees inequality: A Bayesian Cramèr-Rao bound, Bernoulli **1**, 59 (1995).

[34] C. Granade, C. Ferrie, and S. T. Flammia, Practical adaptive quantum tomography, arXiv:1605.05039.

[35] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis, Phys. Rev. A **84**, 052315 (2011).

[36] A. J. F. Hayes and D. W. Berry, Swarm optimization for adaptive phase measurements with low visibility, Phys. Rev. A **89**, 013838 (2014).

[37] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Hamiltonian Learning and Certification using Quantum Resources, Phys. Rev. Lett. **112**, 190501 (2014).

[38] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. OBrien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[39] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, Exploiting locality in quantum computation for quantum chemistry, J. Phys. Chem. Lett. **5**, 4368 (2014).

[40] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[41] N. Wiebe, C. Granade, and D. G. Cory, Quantum bootstrapping via compressed quantum Hamiltonian learning, New J. Phys. **17**, 022005 (2015).