

## Hamiltonian Learning and Certification Using Quantum Resources

Nathan Wiebe,<sup>1,2,3</sup> Christopher Granade,<sup>4,3</sup> Christopher Ferrie,<sup>5</sup> and D. G. Cory<sup>6,3,7</sup>

<sup>1</sup>*Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA*

<sup>2</sup>*Department of Combinatorics & Optimization, University of Waterloo, Ontario N2L 3G1, Canada*

<sup>3</sup>*Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada*

<sup>4</sup>*Department of Physics, University of Waterloo, Ontario N2L 3G1, Canada*

<sup>5</sup>*Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA*

<sup>6</sup>*Department of Chemistry, University of Waterloo, Ontario N2L 3G1, Canada*

<sup>7</sup>*Perimeter Institute, University of Waterloo, Ontario N2L 2Y5, Canada*

(Received 15 November 2013; published 14 May 2014)

In recent years quantum simulation has made great strides, culminating in experiments that existing supercomputers cannot easily simulate. Although this raises the possibility that special purpose analog quantum simulators may be able to perform computational tasks that existing computers cannot, it also introduces a major challenge: certifying that the quantum simulator is in fact simulating the correct quantum dynamics. We provide an algorithm that, under relatively weak assumptions, can be used to efficiently infer the Hamiltonian of a large but untrusted quantum simulator using a trusted quantum simulator. We illustrate the power of this approach by showing numerically that it can inexpensively learn the Hamiltonians for large frustrated Ising models, demonstrating that quantum resources can make certifying analog quantum simulators tractable.

DOI: [10.1103/PhysRevLett.112.190501](https://doi.org/10.1103/PhysRevLett.112.190501)

PACS numbers: 03.67.Ac, 03.65.Wj, 03.67.Lx

Quantum information processing promises to dramatically advance physics and chemistry by providing efficient simulators for the Schrödinger or Dirac equations [1–3]. This is important because conventional methods are inefficient, scaling exponentially in the number of interacting subsystems. Consequently, quantum simulations beyond a few tens of interacting particles are generally believed to be beyond the limitations of conventional supercomputers. Although digital quantum simulation is currently limited to systems that can easily be simulated using a laptop [4], current analog simulators already can simulate systems that are difficult to simulate using existing supercomputers [5–7]. A major objection to this avenue of inquiry is that analog simulators are not necessarily trustworthy [8,9] and certification of them is not known to be efficient. Without such certification, an analog simulator can at best only provide hints about the answer to a given computational question. A resolution to this problem is necessary if analog quantum simulators are to compete on an even footing with supercomputers.

This issue is equally important to the problem of quantum hardware validation, where the goal is to decide whether a quantum device is actually implementing the intended quantum dynamics. Without a way to practically characterize the dynamics of a large quantum system, it is difficult to imagine how to build gates to the exacting tolerances required for fault tolerant quantum computation. Our work provides a new way to leverage quantum resources to make characterization practical, and, in turn, to enable the validation of quantum simulators and devices.

There has been substantial progress towards finding an efficient resolution to related problems using tomographic methods [10,11], and recently machine learning and statistical inference methods have been introduced [12–19] to address similar problems in metrology or Hamiltonian learning. Yet none of these approaches provide an efficient and robust method for characterizing generic Hamiltonian dynamics. We overcome these challenges by providing a robust method that can be used to characterize unknown Hamiltonians by unifying statistical inference with quantum simulation. The key insight behind this unification is that Bayesian inference reduces the problem of Hamiltonian estimation to a problem in Hamiltonian simulation that can be efficiently solved using a trusted quantum simulator.

Our algorithm achieves this through the following steps. (i) The user begins by positing a Hamiltonian model for the system and a probability distribution over the parameters of the Hamiltonian model. Specifically, each hypothetical Hamiltonian  $H_j$  is represented by a vector of  $d$  real numbers  $\vec{x}_j$ , such that  $H_j = H(\vec{x}_j)$ . The Hamiltonian model is therefore specified by  $H(\vec{x})$ . The prior distribution is input using a finite particle approximation known as the sequential Monte Carlo approximation (SMC). (ii) The algorithm adaptively chooses experiments based on the current uncertainty in the Hamiltonian. The experiment is then performed and the trusted quantum simulator is used to efficiently compute the likelihood of the measurement outcome occurring if each hypothetical model were true. (iii) The algorithm then updates its knowledge of the Hamiltonian parameter via Bayes rule, resulting in an updated probability distribution, called the posterior distribution. Steps (ii) and (iii) are then

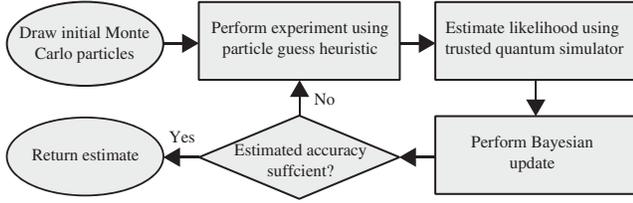


FIG. 1. Flow chart for Hamiltonian learning algorithm.

repeated until the uncertainty in the unknown Hamiltonian parameters becomes small. This iterative process is given in Fig. 1.

We consider three classes of inference experiments given an initial state  $|\psi\rangle$  (typically a pseudorandom state [20]): (a) classical likelihood evaluation (CLE), (b) quantum likelihood evaluation (QLE), and (c) interactive quantum likelihood evaluation (IQLE). CLE is the simplest of these experiments, as it involves only classical resources, and is discussed in detail in [17]. It involves simply picking an experimental time  $t$ , and computing the likelihood  $\Pr(D|\vec{x}_i) = |\langle D|e^{-iH(\vec{x}_i)t}|\psi\rangle|^2$  using a classical computer, where  $D$  is the experimental outcome. This function, known as the likelihood function, is not generally efficient on a classical computer because it involves quantum simulation.

In QLE experiments, a trusted quantum simulator is used to ameliorate these problems. It does so by estimating  $\Pr(D|\vec{x}_i)$  to be the fraction of times outcome  $D$  occurs in a sufficiently large set of simulated experiments, which is efficient if  $\Pr(D|\vec{x}_i)$  is only polynomially small. This approach allows a complex quantum simulator, such as a fault tolerant quantum computer, to act as a certifier for an uncharacterized quantum device. A trusted quantum simulator could also be constructed using a bootstrapping protocol, wherein a smaller trusted analog simulator is the certifier. This is possible if a compressed simulation scheme [21] for the dynamics exists, which is typical for local systems [10,11].

The Loschmidt echo famously shows that, for complex quantum systems, two nearly identical Hamiltonians will typically generate evolutions that diverge exponentially after a short time, before saturating at an exponentially small overlap. For example, if a Hamiltonian  $H_b \in \mathbb{C}^{N \times N}$  differs from a Hamiltonian  $H_a$  by an amount that is large compared to the characteristic scale of its eigenvalue gaps and  $H_a$  has complexity that is characteristic of canonical random matrix ensembles, then there exists times  $t_1$  and  $t_{\text{eq}}$  such that [22]

$$|\langle \psi | e^{iH_a t} e^{-iH_b t} | \psi \rangle|^2 \sim \begin{cases} 1 - O(t^2) & \text{if } t < t_1 \\ e^{-O(t)} & \text{if } t \in [t_1, t_{\text{eq}}] \\ 1/N & \text{for most } t \geq t_{\text{eq}}. \end{cases}$$

Hence, the simulated evolution and actual evolutions will vary substantially in typical QLE experiments if  $t$  is not short. This is not ideal since (a) experiments with small  $t$  tend to be uninformative [17] and (b) the exponentially small likelihoods that occur for  $t > t_{\text{eq}}$  cannot be efficiently estimated.

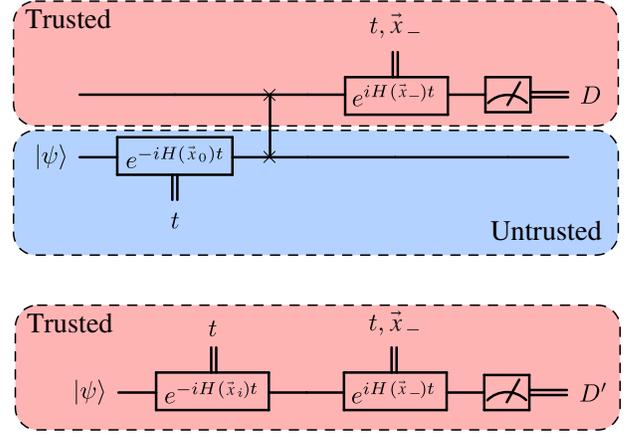


FIG. 2 (color online). IQLE. (Upper) Circuit diagram for drawing experimental data. (Lower) Circuit diagram for likelihood evaluation. QLE is similar but with  $H_- = 0$ , no SWAP gate is used and the untrusted system is measured directly.

We resolve these issues by using interactive QLE experiments, described in Fig. 2. The key point in IQLE experiments is that if a state is evolved forward in time under  $H$  and then backwards in time under  $H_-$  then it will approximately return to the initial state if  $\|H - H_-\|$  is small. Thus, even if the state vector is scrambled by the forward evolution, inverting under a well chosen  $H_-$  will unscramble it, making  $\Pr(D|\vec{x}_i)$  easy to estimate.

This provides a strong separation between the different hypotheses in our prior distribution, since if the inverted state is measured to be the initial state, we have strong evidence that  $H \approx H_-$ , since the Loschmidt echo gives the probability of returning to the initial state for a bad choice of  $H_-$  to be approximately  $1/N \approx 0$ . By acting on the quantum state output by the system under study, IQLE can therefore be highly informative even when QLE is not, due to effects like the Loschmidt echo.

We need to know that the experiment correctly evolved the state under  $H_-$  for such a test to be definitive. This is achieved by SWAP ping the state in the untrusted system into the trusted simulator, then evolving backwards according to  $H_-$  using the trusted simulator. The seemingly trivial SWAP gate in Fig. 2 is actually key to IQLE, as it enables quantum communication between the two registers. No SWAP gate is needed in the likelihood estimation step since all evolutions are performed using the trusted simulator.

More formally, an IQLE experiment is a two outcome experiment: the final state is either measured to be  $|\psi\rangle$ , or it is not. The likelihood function is then completely specified by  $P(\psi|\vec{x}) = |\langle \psi | e^{iH_- t} e^{-iH t} | \psi \rangle|^2$  and  $P(\psi^\perp|\vec{x}) = 1 - |\langle \psi | e^{iH_- t} e^{-iH t} | \psi \rangle|^2$ . Furthermore, we can estimate how rapidly the likelihood decays as  $t$  and  $\|H - H_-\|_2$  increase:

$$|\langle \psi | e^{iH_- t} e^{-iH t} | \psi \rangle| \geq 1 - O(\|H - H_-\|_2 t). \quad (1)$$

If  $H_-$  is within 1 standard deviation of  $H$ , by choosing  $t = 1/\|H - H_-\|_2$ , the likelihoods of both outcomes are then  $O(1)$  for the most likely models. This causes the

experimental results to increase or decrease the posterior probability of the most likely models by constant factors, causing uncertainty in  $H$  to shrink exponentially with the number of updates (if the likelihood function is not flat). We show in the Supplemental Material [32] that these choices are optimal for certain analytically tractable learning problems.

These values of  $H_-$  and  $t$  are difficult to compute directly since  $H$  is unknown. We address this by using the ‘‘particle guess heuristic’’ (PGH), which approximates  $H_-$  and  $\|H(\vec{x}) - H_-\|$  by sampling. The PGH samples two particles,  $\vec{x}_-$  and  $\vec{x}_1$  from the prior distribution  $\Pr(\vec{x})$ . It then chooses  $H_- = H(\vec{x}_-)$  and  $t = 1/\|H(\vec{x}_1) - H(\vec{x}_-)\|_2$ , which, for linear models, can be approximated as  $t = 1/|\vec{x}_1 - \vec{x}_-|$ . As the uncertainty in the estimated parameter shrinks, the PGH will tend to pick longer times to ensure that informative experiments continue to be chosen as certainty about the unknown parameters increases. In practice, this means that  $t$  grows as  $1/\delta$ , similar to phase estimation.

The measurement outcomes yielded by the experiments are immediately processed using Bayesian inference, as described in Fig. 1. This accelerates the learning process by allowing experiments to be chosen using the current knowledge of the correct Hamiltonian. This state of knowledge is represented by a distribution that is called, previous to the next update step, the *prior*. In the cases we consider, the initial prior distribution before any data is observed is taken to be uniform. This encodes a state of maximum ignorance about the correct  $\vec{x}$ . The prior distribution is updated as measurement outcomes are recorded using Bayes’ rule, which gives the proper way of computing the probability of each  $\vec{x}_j$  being correct given the observed data and the prior. It states that if datum  $D$  is recorded then

$$\Pr(\vec{x}_j|D) \propto \Pr(D|\vec{x}_j) \Pr(\vec{x}_j), \quad (2)$$

up to a normalization factor. Here,  $\Pr(\vec{x}_j|D)$  is called the posterior distribution.

Equation (2) can be efficiently computed (for a polynomial number of  $H_j$ ) only if the likelihood function  $\Pr(D|\vec{x}_j)$  is tractable. Unfortunately, there is no known efficient classical algorithm for computing  $\Pr(D|\vec{x}_j)$ . Our algorithm overcomes this by estimating the likelihoods using output sampled from the trusted simulator. In particular, for each  $\vec{x}_j$ ,  $e^{iH_-t} e^{-iH(\vec{x}_j)t} |\psi\rangle$  is simulated. The resultant state is then measured and  $\Pr(D|\vec{x}_j)$  is estimated to be the fraction of times that outcome  $D$  is observed. To reach an error within  $\epsilon$ , it suffices to draw  $O(1/\epsilon^2)$  samples for each  $\Pr(D|\vec{x}_j)$ . This removes the main obstacle to using Bayesian methods to learn the correct  $\vec{x}$ .

We make the update rule tractable by employing the SMC approximation [17,23,24], which approximates the prior as a discrete set of weighted point ‘‘particles’’ that are adaptively moved so as to concentrate particle density around regions of high probability. This procedure is described in detail in the Supplemental Material [32].

The key advantage of our method is that the cost of Hamiltonian inference on a fixed number of IQLE experiments is exponentially smaller than the cost of using CLE, which are the best known methods for certain classes of problems [17]. A natural measure of the cost is the number of quantum simulations needed to estimate the Hamiltonian parameters. The total cost is, therefore,

$$\text{Cost} = N_{\text{steps}}(\delta) \times \text{Cost}(\text{update}; \epsilon). \quad (3)$$

Here,  $N_{\text{steps}}$  is the number of updates needed to make the uncertainty less than  $\delta$  and  $\text{Cost}(\text{update}; \epsilon)$  is the number of samples from the trusted simulator that are needed to update the particle weights using Eq. (2) within error  $\epsilon$  in the 1-norm. We show in the Supplemental Material [32] that, with high probability,  $\text{Cost}(\text{update}; \epsilon)$  scales at most as

$$\frac{|\{\vec{x}_j\}|}{\epsilon^2} \left( \mathbb{E}_{D|H} \left[ \frac{\max_k \Pr(D|\vec{x}_k)(1 - \Pr(D|\vec{x}_k))}{(\sum_k \Pr(D|\vec{x}_k) \Pr(\vec{x}_k))^2} \right] \right). \quad (4)$$

This implies that the update process will be efficient if the number of particles required is small and the resultant probability distribution is not too flat. That is,  $|\{\vec{x}_i\}| \in O(\text{poly}(n))$  and  $\sum_k \Pr(D|H(\vec{x}_k)) \Pr(H(\vec{x}_k)) \in O(1/\text{poly}(n))$ , where  $n$  is the number of interacting systems. SMC algorithms require a number of particles that scales subexponentially in  $d$  [25], which may be independent of  $n$ . This means a small number of particles are required in practice. The robustness of the algorithm to sampling errors is discussed in [26] as well as in the Supplemental Material [32], so relatively large  $\epsilon$  can be tolerated.

If the posterior distribution has converged to a unimodal distribution such that  $\vec{x}$  is within a fixed distance from the mean, then the PGH and (2) ensure that  $\mathbb{E}_{H_-} [|\langle \psi | e^{iH_-t} e^{-iH(\vec{x}_j)t} | \psi \rangle|^2] \in \Theta(1)$ , since  $t \in \Theta(|\vec{x} - \vec{x}_-|^{-1})$ . If a two outcome measurement is used then Markov’s inequality implies that  $\sum_k \Pr(D|H(\vec{x}_k)) \Pr(H(\vec{x}_k)) \in \Theta(1)$  with high probability, implying efficiency. A similar result holds if the *effective* number of outcomes,  $\sum_j \Pr(j|\vec{x}_k)^{-2}$ , is at most  $O(\text{poly}(n))$  for each  $\vec{x}_k$ . In contrast, QLE experiments may not lead to a superpolynomial separation in the cost estimates for generic Hamiltonians and large  $t$  because  $\sum_k \Pr(D|H(\vec{x}_k)) \Pr(H(\vec{x}_k)) \in 2^{-\Theta(n)}$  with high probability for complex quantum systems [22,27]. This can be rectified by choosing small  $t$  as per [10], but such QLE experiments will be much less informative [17].

We can now bound the complexity of performing the inference procedure given these assumptions. It is straightforward to show (see Supplemental Material [32]) that the numerical error in the posterior mean is at most  $\max_i \|H(\vec{x}_i)\| \epsilon$ . Hence, if we wish to learn  $H(\vec{x})$  to within error  $\delta$  it suffices to take  $\epsilon \in \Theta(\delta/\max_i \|H(\vec{x}_i)\|)$ . Finally, given that  $\delta \in O(e^{-\gamma N_{\text{steps}}})$ , where  $\gamma$  is potentially a function of  $\epsilon$  and  $|\{\vec{x}_j\}|$ , (4) implies that the cost of QHL bounded above by

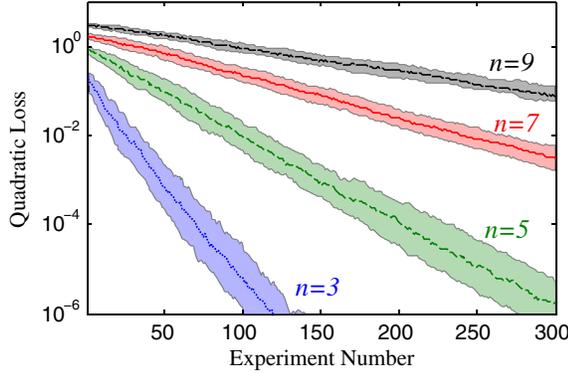


FIG. 3 (color online). The quadratic loss plotted as a function of the number of inversion experiments for Ising models on the complete graph. The shaded areas show a 50% confidence interval for the quadratic loss.

Cost(inference;  $\delta, \gamma$ )

$$\begin{aligned} & \in O\left(\frac{\log(1/\delta)|\{\vec{x}_i\}|\max_k\|H(\vec{x}_k)\|^2}{\gamma\delta^2}\right. \\ & \left.\times\left(\mathbb{E}_{D|H}\left[\frac{\max_k\Pr(D|\vec{x}_k)(1-\Pr(D|\vec{x}_k))}{(\sum_k\Pr(D|\vec{x}_k)\Pr(\vec{x}_k))^2}\right]\right)\right), \quad (5) \end{aligned}$$

also  $\text{Cost}(\text{inference}; \delta) \in \Omega(d \log_2(1/\delta))$ , since IQLE yields one bit per experiment.

Consider the problem of learning  $H(\vec{x})$  using IQLE experiments for an Ising model with no external field:

$$H(\vec{x}) = \sum_{(i,j) \in G} x_{i,j} \sigma_z^{(i)} \sigma_z^{(j)}, \quad (6)$$

where  $G$  is the edge set of an interaction graph on  $n$  qubits. Unless otherwise specified, we take  $x_{i,j} \in [-1/2, 1/2]$  uniformly at random and the initial state is  $|\psi\rangle = |+\rangle^{\otimes n}$ . We choose this Hamiltonian not only because it is physically relevant [28], but also for numerical expediency, since QHL requires thousands of simulated evolutions of the initial state. All measurements are performed in the eigenbasis of  $X^{\otimes n}$ .

Figure 3 shows that the quadratic loss (a generalization of the mean-squared error for multiple parameters) shrinks exponentially with the number of experiments performed; however, the rate at which the error decreases slows as the number of qubits  $n$  increases. This is expected because  $d = n(n-1)/2$  for the case of a complete interaction graph, which implies that the learning problem becomes more difficult as  $n$  increases. The data for interactions on the line are similar and are presented in the Supplemental Material [32] and QLE data are given in the Supplemental Material [32]. Only a few hundred IQLE experiments are needed to learn the Hamiltonian within a quadratic loss of  $10^{-2}$  or smaller for even a nine-qubit system.

The rate at which the learning process slows as  $n$  increases is investigated in Fig. 4. We examine the slowing of the learning problem by fitting the quadratic loss  $\delta$ , in each experiment to  $Ae^{-\gamma N_{\text{steps}}}$ . The median decay exponent, which is the median of the values of  $\gamma$  attained for a set of

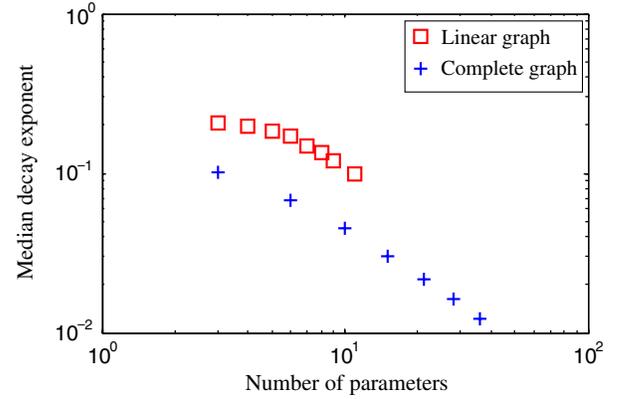


FIG. 4 (color online). The median decay exponent for the quadratic loss as a function of the number of parameters in the Ising model,  $d$ .

experiments with constant  $n$ , measures the typical speed with which QHL learns  $\vec{x}$ . Figure 4 shows that these decay constants scale as  $O(1/d)$  for the complete graph and the line. This implies that  $N_{\text{steps}}(\delta) \in O(d \log(1/\delta))$  for this Hamiltonian, which implies that the inference is efficient. Similarly, the PGH implies that the total simulation time needed (for fixed  $|\{\vec{x}_i\}|$ ) scales as  $N_{\text{steps}}\delta^{-1} \in O(d\delta^{-1/2} \log(1/\delta))$ , which is relevant if the cost of a simulation is dominated by the evolution time.

In Fig. 5, we consider the case of the Ising model on the complete graph where each of the  $x_{i,j}$  is approximately the same value chosen uniformly on  $[0,100]$ , but with small normally distributed fluctuations with mean 0 and variance  $10^{-4}$ . This causes the learning problem to effectively transition from  $d = 1$  to  $d = n(n-1)/2$  once the small fluctuations need to be identified. The transition happens at  $\delta \approx d \times 10^{-4} \approx 10^{-3}$ , which coincides with the point when the slope in Fig. 5 changes. The cost of QHL therefore only implicitly depends on  $n$  through  $d$  since  $\gamma(n=4)/\gamma(n=6) \approx 6 \cdot 5/4 \cdot 3$  after and  $\gamma(n=4)/\gamma(n=6) \approx 1$  before the transition.

Bayesian inference combined with the SMC approximation provides an ideal way to leverage a (potentially

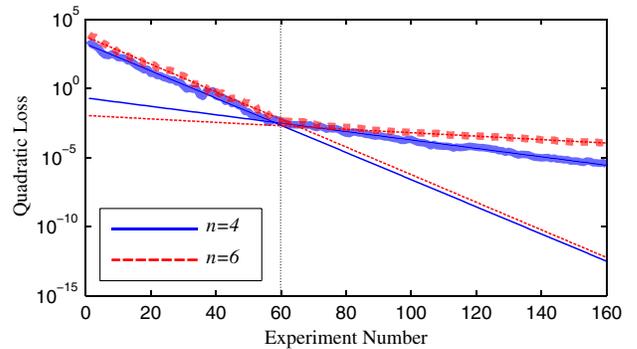


FIG. 5 (color online). An approximate 1 parameter Ising model on the complete graph. The thin lines give the best fits to the exponential decays, which scale as  $e^{-0.07N}$  and  $e^{-0.23N}$  for  $n = 4$  and as  $e^{-0.029N}$  and  $e^{-0.23N}$  for  $n = 6$  qubits.

nonuniversal) quantum simulator to characterize an unknown or unreliable quantum system. We provide theoretical evidence that shows that the update rule, which is at the heart of the learning algorithm, can be performed exponentially faster using quantum resources. We also give upper and lower bounds on the resources needed to estimate the Hamiltonian using QHL and find numerically that QHL using IQLE saturates the scaling predicted by the lower bound in every example that we consider. These results show that our QHL protocol is a powerful and practical method for inferring the dynamics of unknown Hamiltonian systems.

Looking forward, these results say something profound: quantum simulation provides a new way to interrogate quantum systems about their dynamics, thereby revealing patterns that can be used to distinguish models in what would otherwise seem to be random measurement outcomes. Apart from providing efficient methods for modeling quantum systems, these ideas constitute a new way of approaching physics wherein a quantum simulator is more than just a computational device: it is also a powerful experimental tool. Such tools give us the ability to address open questions, such as how to certify analog quantum simulators, and may also provide the first practical methods for bootstrapping a large scale quantum computer, wherein compressed quantum simulators actively control the dynamics of the quantum computer. In short, by drawing on insights from statistical inference, we show that quantum simulation is a powerful experimental resource, and we give an algorithm that exploits this resource to address pressing and immediate problems.

We acknowledge Troy Borneman for suggesting bootstrapping and also Krysta Svore, Allan Geller, and Olivier-Landon Cardinal for useful feedback and discussion. The numerical experiments performed here used SciPy, F2Py and QInfer [29–31]. This work was supported by funding from USARO-DTO, NSERC, CERC, and CIFAR. C. F. was supported in part by NSF Grants No. PHY-1212445 and No. PHY-1005540.

- 
- [1] S. Lloyd, *Science* **273**, 1073 (1996).  
 [2] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, *Science* **309**, 1704 (2005).  
 [3] R. Gerritsma, G. Kirchmair, F. Zähringer, E. Solano, R. Blatt, and C. F. Roos, *Nature (London)* **463**, 68 (2010).  
 [4] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zhringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos, *Science* **334**, 57 (2011).  
 [5] J. Simon, W. S. Bakr, R. Ma, M. E. Tai, P. M. Preiss, and M. Greiner, *Nature (London)* **472**, 307 (2011).  
 [6] J. W. Britton, B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger, *Nature (London)* **484**, 489 (2012).  
 [7] K. Kim, M.-S. Chang, S. Korenblit, R. Islam, E. E. Edwards, J. K. Freericks, G.-D. Lin, L.-M. Duan, and C. Monroe, *Nature (London)* **465**, 590 (2010).  
 [8] P. Hauke, F. M. Cucchietti, L. Tagliacozzo, I. Deutsch, and M. Lewenstein, *Rep. Prog. Phys.* **75**, 082401 (2012).  
 [9] C. Gogolin, M. Kliesch, L. Aolita, and J. Eisert, [arXiv:1306.3995](https://arxiv.org/abs/1306.3995).  
 [10] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, *Phys. Rev. Lett.* **107**, 210404 (2011).  
 [11] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, *Phys. Rev. A* **84**, 012107 (2011).  
 [12] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **104**, 063603 (2010).  
 [13] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **107**, 233601 (2011).  
 [14] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, *Phys. Rev. A* **84**, 052315 (2011).  
 [15] C. Ferrie, C. E. Granade, and D. G. Cory, *Quantum Inf. Process.* **12**, 611 (2013).  
 [16] A. Sergeevich and S. D. Bartlett, *Proceedings of 2012 IEEE Conference on Evolutionary Computation (CEC), Brisbane, Australia, 2012* (IEEE, New York, 2012).  
 [17] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, *New J. Phys.* **14**, 103013 (2012).  
 [18] N. B. Lovett, C. Crosnier, M. Perarnau-Llobet, and B. C. Sanders, *Phys. Rev. Lett.* **110**, 220501 (2013).  
 [19] K. M. Svore, M. B. Hastings, and M. Freedman, *Quantum Inf. Comput.* **14**, 306 (2013).  
 [20] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, *Science* **302**, 2098 (2003).  
 [21] B. Kraus, *Phys. Rev. Lett.* **107**, 250503 (2011).  
 [22] F. Haake, *Quantum Signatures of Chaos* (Springer-Verlag, New York, 2004), 2nd ed.  
 [23] F. Huszar and N. M. T. Houlby, *Phys. Rev. A* **85**, 052120 (2012).  
 [24] A. Doucet and A. M. Johansen, *The Oxford Handbook of Nonlinear Filtering*, edited by D. Crisan and B. Rozovskii (Oxford University Press, Oxford, 2011).  
 [25] A. Beskos, D. Crisan, and A. Jasra, [arXiv:1103.3965](https://arxiv.org/abs/1103.3965)  
 [26] C. Ferrie and C. E. Granade, *Phys. Rev. Lett.* **112**, 130402 (2014).  
 [27] C. Ududec, N. Wiebe, and J. Emerson, *Phys. Rev. Lett.* **111**, 080403 (2013).  
 [28] P. Richerme, C. Senko, S. Korenblit, J. Smith, A. Lee, W. C. Campbell, and C. Monroe, *Phys. Rev. Lett.* **111**, 100506 (2013).  
 [29] E. Jones *et al.*, SciPy: Open source Scientific Tools for Python (2001); <http://www.scipy.org>.  
 [30] P. Peterson, *Int. J. Comput. Sci. Eng.* **4**, 296 (2009).  
 [31] C. Ferrie *et al.*, QInfer: Library for Statistical Inference in Quantum Information (2012); <https://github.com/csferrie/python-qinfer>.  
 [32] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.112.190501>, which includes Refs. [33–35], for additional numerical experiments for Hamiltonian learning using QLE and IQLE. It also provides evidence for the optimality of the particle guess heuristic in analytically tractable cases and provides a proof for robustness of the protocol to sampling error.  
 [33] J. Liu and M. West, *Combined Parameter and State Estimation in Simulation-Based Filtering* (Springer-Verlag, New York, 2000).  
 [34] C. Ferrie, *New J. Phys.* **16**, 023006 (2014).  
 [35] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998), ISBN 9780387985022.