Quantum Hamiltonian learning using imperfect quantum resources

Nathan Wiebe,[1,2,3] Christopher Granade,[3,4] Christopher Ferrie,[5] and David Cory[3,6,7]

[1]*Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA*
[2]*Department of Combinatorics & Optimization, University of Waterloo, Ontario, Canada N2L 3G1*
[3]*Institute for Quantum Computing, University of Waterloo, Ontario, Canada N2L 3G1*
[4]*Department of Physics, University of Waterloo, Ontario, Canada N2L 3G1*
[5]*Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA*
[6]*Department of Chemistry, University of Waterloo, Ontario, Canada N2L 3G1*
[7]*Perimeter Institute, University of Waterloo, Ontario, Canada N2L 2Y5*

Identifying an accurate model for the dynamics of a quantum system is a vexing problem that underlies a range of problems in experimental physics and quantum information theory. Recently, a method called *quantum Hamiltonian learning* has been proposed by the present authors that uses quantum simulation as a resource for modeling an unknown quantum system. This approach can, under certain circumstances, allow such models to be efficiently identified. A major caveat of that work is the assumption of that all elements of the protocol are noise free. Here we show that quantum Hamiltonian learning can tolerate substantial amounts of depolarizing noise and show numerical evidence that it can tolerate noise drawn from other realistic models. We further provide evidence that the learning algorithm will find a model that is maximally close to the true model in cases where the hypothetical model lacks terms present in the true model. Finally, we also provide numerical evidence that the algorithm works for noncommuting models. This work illustrates that quantum Hamiltonian learning can be performed using realistic resources and suggests that even imperfect quantum resources may be valuable for characterizing quantum systems.

I. INTRODUCTION

The challenges faced by experimentalists trying to learn an appropriate Hamiltonian model for a large quantum system can be quite daunting. Traditional techniques, such as tomography, rapidly become infeasible as the number of qubits increases. To make matters worse, the dynamics of such quantum systems cannot even be simulated efficiently using existing methods without making strong assumptions. This raises an important question: How can you decide the properties of a model that is too complex to even simulate?

This is not simply a point of theoretical interest. Present-day experiments are already operating in regimes that are challenging for classical supercomputers to simulate [1], and near-future experiments will soon be well beyond their capabilities [2–5]. The stakes are further compounded by the fact that many of these systems are intended as quantum simulators or demonstrations of the supremacy of quantum information processing over classical information processing, and none of these demonstrations can be considered compelling unless their predictions can be independently verified [6–8]. A resolution to this problem is thus urgently needed.

A natural solution to this problem is to leverage the inherent power of quantum systems as a resource to characterize other quantum systems. The idea behind this approach is simple: In order to learn a Hamiltonian model, you build a trustworthy quantum simulator for that class of models and use it to make dynamical predictions about a hypothetical model for the quantum system. This quantum simulator need not necessarily be a quantum computer, but it must be "trusted" that the dynamical map that it implements is sufficiently close to that of the ideal model.

In Ref. [9], we provided a concrete way of implementing this procedure using Bayesian inference, wherein trusted quantum simulators are used to compute the probability that a hypothetical model would yield the observed measurement outcome. The approach is shown to be remarkably efficient at learning Hamiltonian parameters and resilient to some forms of noise, such as shot noise in the computation of the probabilities. Two implementations of quantum Hamiltonian learning are proposed in Ref. [9]. The first involves simply using a quantum simulator as a resource for computing likelihood functions that appear in Bayesian inference or, more concretely, the Hamiltonian inference protocol in Ref. [10]. The second approach goes beyond this by allowing the quantum simulator to interact with the experimental system via a SWAP gate. This approach is called *interactive* quantum Hamiltonian learning, and it is shown to be more stable and efficient than its noninteractive brethren. Here we go beyond these results and show that interactive quantum Hamiltonian learning is resilient to realistic sources of noise that can arise in the protocol. This not only illustrates that interactive experiments can be performed with realistic quantum resources but also suggests that they could be performed with existing or near-future quantum systems.

Before ending this Introduction with an outline of the paper, we briefly comment on the relation to other learning methods which seek to reduce the cost of characterization and validation. These include identifying stabilizer states [11,12]; tomography for matrix product states [13]; tomography for permutationally invariant states [14]; learning local Hamiltonians [12]; tomography for low-rank states via compressed sensing [15]; and tomography for multiscale entangled states [16]. Several of these methods use quantum resources to accelerate quantum the characterization such as the matrix product state tomography method of [13] and the direct quantum process tomography method of [17]. Direct quantum process

tomography is in some senses analogous to our work because it uses two-qubit (or qudit) interactions to infer the dynamics of an unknown system, thereby removing the need to perform a computationally expensive inversion procedure.

The key difference between prior works and ours is that the above techniques employ efficient *classical* simulation algorithms which propagate efficient representations of the state vector to calculate of the probabilities defining the likelihood function, whereas we evaluate the likelihood function using *quantum* resources.

We lay out the paper as follows. We discuss the theory of Bayesian inference that our quantum Hamiltonian learning scheme (as well as the prior classical result in Ref. [10]) uses in Sec. II. Section III reviews the results in Ref. [9], which provides a method for making Bayesian inference practical by using quantum simulation to evaluate the likelihood function. We also provide an explicit algorithm for the procedure in Sec. III. We then present numerical evidence in Sec. IV that shows the quantum Hamiltonian learning (QHL) algorithm can rapidly infer a model for Ising and transverse Ising model Hamiltonians. Section V provides theoretical evidence that the learning algorithm is robust to depolarizing noise and that realistic noise models for the SWAP gates used in interactive QHL experiments do not prevent the algorithm from learning at an exponential rate. Finally, in Sec. VI, we consider the performance of QHL when the model being used differs from the physics governing an experiment.

II. CLASSICAL HAMILTONIAN LEARNING

Many approaches for learning the Hamiltonian of an uncharacterized quantum system have been considered, but only recently have ideas and methods from statistical inference been applied to this problem despite their ubiquity in machine learning and related fields [10,18–21]. Here we consider a Bayesian approach to parameter estimation. This approach, in essence, uses Bayes' theorem to give the probability that a hypothesis about the Hamiltonian H is true, given the evidence that has been accumulated through experimentation. In other words, it provides an approximation to the Hamiltonian (from a class of Hamiltonians) that is most likely to yield the observed experimental data.

Bayesian inference problems are specified by two quantities.

(1) A *prior* probability distribution $\Pr(H)$ that encodes the *a priori* confidence that a given Hamiltonian model H is correct. This is chosen by the experimenter, but can always be taken to be a least informative (typically uniform) distribution utilizing no prior assumptions.

(2) A *likelihood function* $\Pr(D|H)$ that returns the probability that outcome D is observed given H is the true Hamiltonian. This is not chosen; it is prescribed by quantum mechanics via the Born rule.

Bayes' theorem states that the probability that a given hypothetical Hamiltonian H is correct given the observed data D can be computed from the prior and the likelihood function via

$$\Pr(H|D) = \frac{\Pr(D|H)\Pr(H)}{\Pr(D)} = \frac{\Pr(D|H)\Pr(H)}{\int \Pr(D|H)\Pr(H)\,dH}. \quad (1)$$

The inference process then involves repeating the above process for each observed datum after setting $\Pr(H) \leftarrow \Pr(H|D)$, where each such repetition is known as an update. The probability distribution will typically converge to a sharply peaked distribution about the true model parameters (unless uninformative experiments are chosen or the learning problem is degenerate) as the number of updates increases. This procedure has been shown to be extremely effective at Hamiltonian learning: Only a few hundred experiments can lead to an accurate estimate of a Hamiltonian. In contrast, traditional methods can require billions of measurement outcomes to achieve comparable precision [9,10].

An important feature to note is that Bayesian inference reduces the problem of inference to a problem in simulation. In cases where the likelihood can be easily computed, Bayesian inference will often be well suited for the problem, whereas it is ill suited when the likelihood function is intractable. This is the key insight behind the entire QHL approach.

A typical Hamiltonian inference problem involves evolving a known initial state $|\psi\rangle$ under an unknown Hamiltonian H, then measuring against a fixed basis $\{|D\rangle\}$. The Hamiltonian is then inferred from the measurement statistics. In Bayesian Hamiltonian inference, the likelihood function for such experiments is given by the Born rule as as

$$\Pr(D|H) = |\langle D| e^{-iHt} |\psi\rangle|^2. \quad (2)$$

The final issue that needs to be considered is that the probability distributions must be discretized in order to make Bayesian updating tractable. We use a finite particle approximation to the probability distributions known as the sequential Monte Carlo (SMC) approximation, in which we draw samples from the initial prior distribution. Each such particle i drawn from the initial prior is assigned a weight $w_i = 1/N$, and is then updated using Bayes' rule as data is collected.

A common problem with these methods is that the effective sample size $N_{\text{ess}} = \sum_i 1/w_i^2$ of the approximation becomes small as data are incorporated, such that the approximation becomes *impoverished*. To rectify this and to recover numerical stability, we employ a resampling method proposed by Liu and West that changes the particle positions in order to concentrate particles in regions of high posterior probability by interpreting the posterior at any given time step as a mixture of the SMC-approximated posterior and a multivariate normal having the same mean and covariance [22,23]. The quality parameter a allows the resampling algorithm to smoothly interpolate between sampling from the original SMC distribution and a Gaussian distribution; as $a \to 1$, the resampling algorithm draws particles from the SMC approximation to the posterior, whereas the resampling algorithm draws particles from a normal distribution with the same mean and variance as the posterior distribution in the limit $a \to 0$. Here we find that the approximate normality of the problem allows for us to take $a = 0.9$ such that we need fewer particles for each simulated run of our algorithm. By drawing new particles from the resampling mixture distribution, the first two moments of the posterior are manifestly preserved, but the particle approximation is *refreshed* such that the effective sample size of the particle approximation is increased.

This resampling procedure is essential because, with high probability, none of the discrete particles used to represent the probability distribution will coincide with the correct model parameters. This means that as time progresses, the algorithm will become more sure that each of the particles does not hold the correct model parameters but will not be able to accurately estimate the true parameters. Thus, the refresh of the effective sample size allowed by the resampling step avoids this problem by adaptively changing the discretization to reflect the tighter posterior distributions afforded by experimental knowledge.

Further discussion of the technical details of these methods can be found in Refs. [10,23]. After this discretization process is done, Bayesian updating (and experiment design [10]) can be applied to learn the Hamiltonian.

As an example of how the inference procedure may be used in practice, imagine that an experimentalist suspects that their system is an Ising model with Hamiltonian

$$H = \sum_{i=1}^{n-1} \frac{\pi J_i}{2} \sigma_z^{(i)} \sigma_z^{(i+1)}. \tag{3}$$

Bayesian inference can then be used to find the most likely parameters $\{J_i\}$ given the experimental data and any prior knowledge about these parameters. Here and in the majority of the numerical cases that we consider, we take the pessimistic assumption that the experimentalist is maximally ignorant of the Hamiltonian given some physical constraints and reflects this knowledge by choosing $\Pr(H)$ to be the uniform distribution over all Hamiltonian operators obeying these constraints.

In general, we denote the parameters for a Hamiltonian to be a vector $\boldsymbol{x} \in \mathbb{R}^d$ and that the corresponding Hamiltonian is $H(\boldsymbol{x})$. For example, in this example, $d = n - 1$, $\boldsymbol{x} = [J_1, \ldots, J_{n-1}]$ and $H(\boldsymbol{x}) = \boldsymbol{x} \cdot [\sigma_z^{(1)} \sigma_z^{(2)}, \ldots, \sigma_z^{(n-1)} \sigma_z^{(n)}]$. An appropriate choice for the initial state is $|+\rangle^{\otimes n}$, and similarly computational basis measurements are suitable because these state preparations and measurements lead to different Hamiltonian parametrizations, resulting in different measurement outcomes. In contrast, computational basis-state preparations are inappropriate because the Hamiltonian has a trivial action on them.

Bayesian inference is well suited for learning a concise parametrization of a Hamiltonian within a family of potential models. In addition, region estimates for the true Hamiltonian can be easily found from the posterior distribution, which allows the confidence in the final inferred Hamiltonian [10,24] to be quantified. In particular, we advocate for a concise representation of our uncertainty through the ellipse defined by the posterior covariance [10], which was shown to be nearly optimal in terms of capturing the densest region of posterior probability [24], such that covariance ellipsoids provide a very good approximation to the highest-power credible region estimators. Moreover, by the use of a clustering algorithm, this can be extended to allow for efficient region estimation over multimodal distributions. Thus, in addition to providing a fast method for inferring the form of the Hamiltonian, Bayesian inference also naturally gives an estimate of the uncertainty of the result, unlike most tomographic approaches.

In practice, the posterior distribution tends to converge to a unimodal distribution that is, to a good approximation, Gaussian. Under this assumption, an error ellipsoid that contains a ratio $\mathrm{erf}(Z/\sqrt{2})^d$ of the total *a posteriori* probability is given by the set of all $\boldsymbol{x}$ that obey [10]

$$(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \leqslant Z^2, \tag{4}$$

where $\boldsymbol{\mu}$ is the posterior mean and Σ is the posterior covariance matrix. Computation of the posterior mean and the inverse of the covariance matrix is efficient for fixed particle number, if the matrix is well conditioned, because d is considered to be polylogarithmic in the Hilbert-space dimension of the system. This simple method works well in practice, but in cases where a precise estimate of the error is needed the numerical methods discussed in Refs. [10,24] should be used.

A major drawback of this approach is that the likelihood function may, for certain experiments, be expensive to compute using classical computing resources. Quantum Hamiltonian learning can resolve this problem if efficient quantum simulators exist for the class of models used in the inference procedure.

III. QUANTUM HAMILTONIAN LEARNING

The use of the likelihood function $\Pr(D|H(\boldsymbol{x})) \equiv \Pr(D|\boldsymbol{x})$ implies that the ability to infer H is intimately connected to our ability to simulate dynamics according to H, a fact that is clearly illustrated in Bayesian methods by (1). This allows our implementation to be very general and means that our estimation is always motivated by our knowledge of the underlying physics of a system.

Using classical simulation to implement evaluations of the likelihood function $\Pr(D|\boldsymbol{x})$ is typically extremely difficult for quantum systems. For example, performing a dynamical simulation of random 12-qubit quantum systems by direct exponentiation on a dual-processor Xeon E5-2630 work station requires on average roughly 1300 s. Given that the SMC approximation may require more than 2000 particles to achieve a good approximation, and that 200 experiments may be needed to learn the Hamiltonian parameters with reasonable accuracy, this would require approximately 16.5 yr of computational time on the work station, despite these optimistic assumptions. This is difficult, but not outside the realm of possibility for existing supercomputers. If we wished to scale this up to 100 qubits, as is applicable to current proposals for experimental quantum information processing devices [25,26], then roughly 5×10^{80} yr would be required just to process the 20 kbits of data produced by the experiment. Clearly, a better approach for characterizing the dynamics of mesoscopic quantum systems is needed.

A natural solution to this limitation is to use a quantum simulator to estimate the required likelihoods. Efficient quantum simulators exist, in principle, for a wide range of physical systems [25–33]. Such simulators allow the user to draw a sample from $\Pr(D|\boldsymbol{x})$ using energy and time that scale at most polynomially with the number of interacting particles. In contrast, the best known classical methods require time that scales polynomially in the Hilbert-space dimension and hence cannot practically simulate generic large quantum systems (particular properties of certain systems can nonetheless be estimated using alternative techniques such as matrix product states (MPS), density-matrix renormalization group
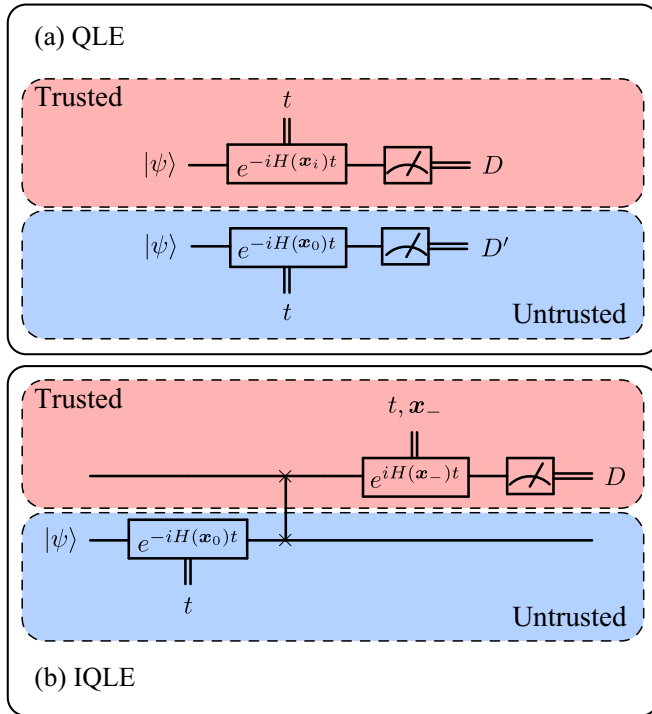
FIG. 1. (Color online) QLE and IQLE experiments. (a) QLE, wherein the untrusted and trusted simulator act in parallel and the outputs are compared; (b) IQLE, wherein the state of the simulators is swapped and the trusted simulator attempts to invert the evolution.

calculations, or Lieb-Robinson bounds [34]). Such quantum simulators can take many forms. It could be a universal quantum computer or it could be a special-purpose analog quantum simulator. Ultimately, the only things that matters are that it can, approximately, sample from $\Pr(D|\boldsymbol{x})$ for any $\boldsymbol{x}$ and that its state can be swapped with that in the uncharacterized quantum system.

Two ways have been proposed to use quantum information to accelerate Bayesian inference: quantum likelihood evaluation (QLE) experiments and interactive quantum likelihood evaluation (IQLE) experiments. A QLE experiment involves repeating the experiment that was performed on the uncharacterized quantum system a large number of times using a simulator for $H(\boldsymbol{x})$. If datum D was recorded for the experiment, then $\Pr(D|\boldsymbol{x})$ is set to be the fraction of times outcome D is observed in the simulations. IQLE experiments are similar, except they involve swapping the quantum state out of the uncharacterized system using a noisy SWAP gate and then approximately inverting the evolution of the uncharacterized system by applying e^{iH_-t}. These methods are illustrated in Fig. 1.

IQLE experiments have many advantages over QLE experiments [9]. First, if $H_- \approx H(\boldsymbol{x}_{\text{true}})$, where $\boldsymbol{x}_{\text{true}}$ are the true parameters, and the noise is negligible then the simulation will approximately map $|\psi\rangle \mapsto |\psi\rangle$. This is useful because it gives a firm condition to check to see if the current hypothesis about the Hamiltonian is correct. In many cases, these benefits can outweigh the increased complexity of IQLE experiments and, in particular, we show that the Hamiltonian parameters can be

learned even given realistic noise in the gate that swaps the states of the trusted simulator and the untrusted system.

IQLE experiments inherit their robustness in part from the use of the *particle guess heuristic* (PGH), which is an adaptive method for choosing a reasonable experiment based on the current knowledge about the unknown quantum system. The heuristic works by drawing two different "particles" from the prior distribution $\Pr(\boldsymbol{x})$, $\boldsymbol{x}_-$ and $\boldsymbol{x}'$. The experimental time is chosen to be $t = \|H(\boldsymbol{x}_-) - H(\boldsymbol{x}')\|^{-1}$. This heuristic has several remarkable properties [9].

(1) The typical value of t used for an experiment is the inverse of the *current* uncertainty in the Hamiltonian. Intuitively, this means that the guess heuristic will (on average) choose an evolution time that causes the majority of the potential models under consideration to have different dynamics.

(2) If $\Pr(\boldsymbol{x})$ has converged to a unimodal distribution centered near $\boldsymbol{x}_{\text{true}}$, then with high probability the measurement outcome will be $|\psi\rangle$, i.e., $|\langle\psi| e^{iH(\boldsymbol{x}_-)t} e^{-iH(\boldsymbol{x}_{\text{true}})t} |\psi\rangle|^2 \in O(1)$.

(3) If $\|H(\boldsymbol{x}_{\text{true}}) - H(\boldsymbol{x}')\|$ is relatively large compared to t^{-1}, then the Loschmidt echo guarantees that $|\langle\psi| e^{iH(\boldsymbol{x}_-)t} e^{-iH(\boldsymbol{x}_{\text{true}})t} |\psi\rangle|^2 \in O(\frac{1}{2^n})$ for almost all Hamiltonians (chosen, for example, uniformly over the Gaussian unitary ensemble of random Hamiltonians [8]).

The key message from these results is that the PGH exploits the unitary (or approximately unitary) nature of time evolution to provide experiments that are likely to be informative. In particular, the Loschmidt echo is exploited by IQLE experiments through the PGH to provide a test to determine with high probability whether the inferred Hamiltonian is close to the "true" Hamiltonian. If the learning problem is well posed, then this allows the inference algorithm to learn that a constant fraction of model Hamiltonians is closer to the true Hamiltonian than the other models. This leads to a constant number of bits of information to be learned, on average, about the Hamiltonian per experiment, which leads to the uncertainty in the inference scaling like

$$\delta \sim Ae^{-\gamma N}, \tag{5}$$

where N is the number of experiments performed and γ is some constant which is independent of N.

The scaling in Eq. (5) implies that the total computational time used scales as $t_{\text{total}} \propto e^{\gamma N}$, which at first glance seems to suggest inefficiency but since the uncertainty also drops exponentially $N \propto \log(1/\delta)$ and hence $t \propto \delta^{-\gamma}$. This scaling is comparable to that expected for Heisenberg-limited metrology in phase estimation protocols if $\gamma \approx 1$. In contrast, if the learning problem is less well posed and $\delta \sim N^{-\Gamma}$, it means that the total time required scales as $1/\epsilon^{\frac{\Gamma+1}{\Gamma}}$. This yields similar scaling to shot-noise-limited metrology when $\Gamma = 1$.

Putting everything together, we obtain an algorithm for performing an IQLE experiments, detailed in Algorithm 1.

For more details on the resampling step and the SMC approximation, see [10]. The method for performing QLE experiments is identical, except $H_- = 0$ in those cases.

How do these methods compare to conventional approaches? Techniques such as MPS tomography can be efficient for states with low bond dimension and are error robust [13]. However, they are inefficient for systems that have high

Algorithm 1. Quantum Hamiltonian learning algorithm.

Input: Particle weights w_i, $i \in \{1, \ldots, M\}$, Particle locations x_i, $i \in \{1, \ldots, M\}$, number of samples used to estimate probabilities N_{samp}, total number of experiments used N_{exp}, state preparation protocol for $|\psi_0\rangle$, protocol for implementing POVM P such that $|\psi_0\rangle\langle\psi_0|$ is an element, resampling algorithm R.

Output: Hamiltonian parameters x such that $H(x) \approx H(x_{\text{true}})$.

---

function QHL ($\{w_i(D)\}$, $\{x_i\}$, N_{samp}, N_{exp}, P, R)
  for $i \in 1 \rightarrow N_{\text{exp}}$ **do**
    Draw x_- and x' from $\text{Pr}(x) := w_i / \sum_i w_i$. $\triangleright$ Choose x_- according to PGH
    $t \leftarrow 1/\|H(x) - H(x')\|$. $\triangleright$ Choose t according to PGH
    $D \leftarrow$ measurement of $e^{iH(x_-)t} e^{-iH(x_{\text{true}})t}$ using P. $\triangleright$ Perform IQLE experiment on untrusted system.
    for $j \in 1 \rightarrow M$ **do** $\triangleright$ Compute likelihoods using trusted simulator
      $p_j \leftarrow 0$.
      for $k \in 1 \rightarrow N_{\text{samp}}$ **do**
        $D' \leftarrow$ measurement of $e^{iH(x_-)t} e^{-iH(x_j)t} |\psi_0\rangle$ using P.
        if $D' = D$ **then**
          $p_j \leftarrow p_j + 1/N_{\text{samp}}$.
        end if
      end for
    end for
    $Z \leftarrow \sum_{m=1,M} w_m p_m$.
    $w_i \leftarrow w_i p_i / Z$. $\triangleright$ Perform update.
    if $1/(\sum_m w_i^2) < M/2$ **then**
      $(\{w_i\}, \{x_i\}) \leftarrow R(\{w_i\}, \{x_i\})$. $\triangleright$ Resample if weights are too small
    end if
  end for
  return $\sum_m p_m x_m$ $\triangleright$ Return Bayes estimate of x_{true}.
end function

---

bond dimension, use potentially expensive (although efficient) tomographic processes, and can have exponentially worse scaling with the error than our approach. Other approaches such as the direct characterization method of [17] apply to cases where the form of the true model is not known and require very little classical postprocessing, and very few ensemble measurements are available. On the other hand, these methods are inefficient, give exponentially worse scaling with the desired error if ensemble measurements are not available, and do not exploit any prior information about the system. Our previous work [10] is inefficient, but scales exponentially better with the error tolerance than the aforementioned methods, can exploit prior information, and is suitable in cases where single shot readout is used. QHL therefore can combine the desirable properties of all of these methods at the price of using more sophisticated quantum resources.

IV. QHL EXPERIMENTS WITHOUT NOISE

Here we examine QHL in the absence of noise to provide a basis of comparison for the performance of the algorithm when physically realistic noise is considered. We ignore the effects of sampling noise as it has already been studied in Ref. [9], wherein the algorithm is shown to be robust against sampling errors given that a resampling algorithm is used and that experiments are chosen such that the majority of the outcomes do not have exponentially small likelihoods for typical particles in the SMC approximation to the prior distribution.

We consider a Hamiltonian of the form of Eq. (3) with unknown coefficients chosen to be uniformly distributed in the range $[-1/\pi, 1/\pi]$. We then apply the learning algorithm to a case where 20 000 particles are used to describe the prior distribution and examine the error, given by the quadratic loss,

$$L(x, x_{\text{true}}) = \sum_{j=1}^{d} (x_j - x_{\text{true},j})^2, \qquad (6)$$

where $d = n - 1$ is the number of model parameters. We choose the Liu and West resampling algorithm with $a = 0.9$ (see [10] for more details) and, unless otherwise specified, the initial state $|\psi\rangle$ is chosen to be $|+^{\otimes n}\rangle$. We examined the performance of the algorithm for 320 different randomly chosen Ising models. Although the model is not frustrated, previous work suggests that the absence of frustration does not qualitatively change the learning problem [9].

We see from Fig. 2 that the quadratic loss shrinks exponentially with the number of experiments. This is in agreement with prior results from [9]. The learning rate, γ, which is found by fitting individual samples to $Ae^{-\gamma}$ agrees with an $O(1/d)$ scaling as shown in Ref. [9]. Since $d = n - 1$ for this model, such Hamiltonians should be easy to learn for IQLE experiments even in the limit of large n.

A. Noncommuting models

QHL is not limited to models with commuting Hamiltonians. However, in the general noncommuting case, it may
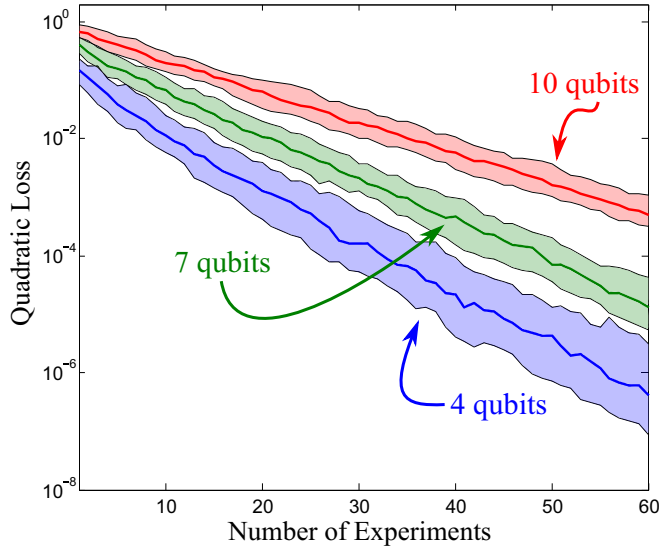
FIG. 2. (Color online) Quadratic loss as a function of number of experiments for learning unknown Ising couplings. The shaded regions represent a 50% confidence interval for the quadratic loss and the solid line gives the median loss.
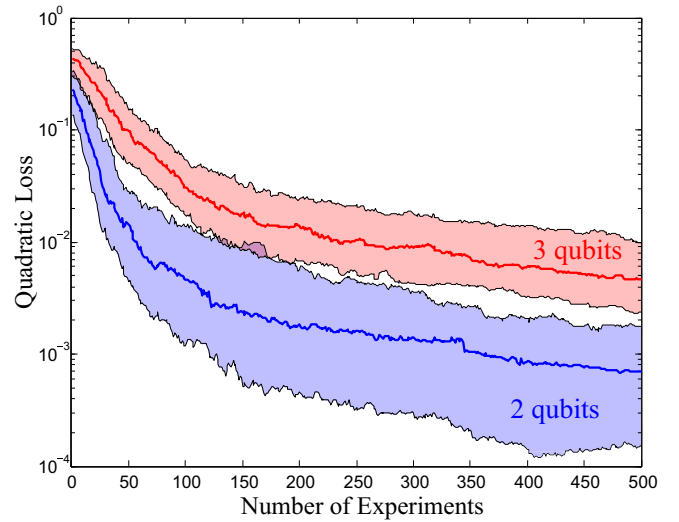

FIG. 3. (Color online) Median quadratic loss as a function of the number of experiments for two- and three-qubit transverse Ising Hamiltonians chosen with $\boldsymbol{x}_k$ uniform in $[0,1]$. The shaded regions give a 50% confidence interval for the data.

be much more difficult to find appropriate initial states that maximize the information yielded by each experiment. We illustrate this by applying QHL to a transverse Ising model of the form

$$H(\boldsymbol{x}) = \sum_{k=1}^{n} \boldsymbol{x}_k \sigma_x^{(k)} + \sum_{k=1}^{n-1} \boldsymbol{x}_{k+n} \sigma_z^{(k)} \otimes \sigma_z^{(k+1)}. \tag{7}$$

The dynamics of the transverse Ising model are clearly much more rich than that of the Ising model and a naive guess for an appropriate initial state or measurement operator for an IQLE experiment is unlikely to yield as much information as the choice of $|\psi_0\rangle = |+\rangle^{\otimes n}$ that was made in the prior example because that choice was motivated by the dynamics of the Ising interaction.

For example, a natural approach to solve such problems in NMR would be to use refocusing to suppress the $\sigma_z \sigma_z$ couplings while leaving the transverse field terms proportional to σ_x by periodically applying $\pi)_x$ pulses to the system being studied. Such pulses can be designed in a broadband manner such that only rough knowledge of $H(\boldsymbol{x})$ is required to implement $\pi)_x$ pulses (see Sec. VI for more details on neglecting terms in the Hamiltonian). After learning the these terms accurately, the interaction terms can be learned to much greater accuracy without suffering loss of contrast. Nonetheless, we show that, in principle, the QHL algorithm can be used directly to learn these couplings using maximally naive experiments; specifically, we generate our initial states randomly in each experiment by applying a random series of local Clifford operations to each qubit, similar to [35,36].

Figure 3 shows that the QHL algorithm can continue to learn Hamiltonian parameters despite the fact that the model is noncommuting. The data were collected using 5000 particles and 160 samples and using the Liu-West resampling algorithm with $a = 0.98$ for all numerical results in this section. It is also worth noting that we are restricted to two or

three qubit systems because (approximately) exponentiating noncommuting Hamiltonians is much more expensive classically than it is for commuting Hamiltonians. The learning rate for short experiments is very rapid, whereas for later times the learning rate substantially slows. This is expected because for short times $\exp(-iHt) \approx \mathbb{1} - i(\boldsymbol{x_1}\sigma_x^{(1)} + \boldsymbol{x_2}\sigma_x^{(2)} + \boldsymbol{x_3}\sigma_z^{(1)} \otimes \sigma_z^{(2)})t$ and hence the single-qubit and multiqubit terms have a clear and separable effect on the experimental outcomes. This results in rapid learning of these Hamiltonian parameters. At later times, progress is substantially slower because the way that the Hamiltonian parameters affect the probabilities of different outcomes becomes less distinct. Specifically, the scaling of the quadratic loss reduces to $\delta \propto 1/N$. This results in a total simulation time that scales as $1/\epsilon^2$, which is comparable to shot-noise-limited metrology. We therefore see that in such cases the PGH alone is insufficient to find highly informative experiments and intelligent choices of experiments and perhaps even local optimization (such as gradient descent optimization as per [10]) become important for optimizing the performance of QHL.

In essence, the slow learning rate of IQLE experiments in Fig. 3 is a consequence of an approximate degeneracy that arises between the on-site terms and the interaction terms wherein the precise effect of a single on-site term becomes hard to resolve. This raises the question of whether exponential scaling of the estimation accuracy can be restored if we break this degeneracy. Figure 4 shows that the answer to this question is "yes." We break the degeneracy there by assuming that the interaction and on-site terms in the Hamiltonian are translationally invariant:

$$H(\boldsymbol{x}) = \boldsymbol{x_1} \sum_{k=1}^{n} \sigma_x^{(k)} + \boldsymbol{x_2} \sum_{k=1}^{n-1} \sigma_z^{(k)} \otimes \sigma_z^{(k+1)}. \tag{8}$$

In the case where $n = 2$, this translationally invariant version has only one fewer parameter than the original Hamiltonian, and yet the performance differences in the learning algorithm
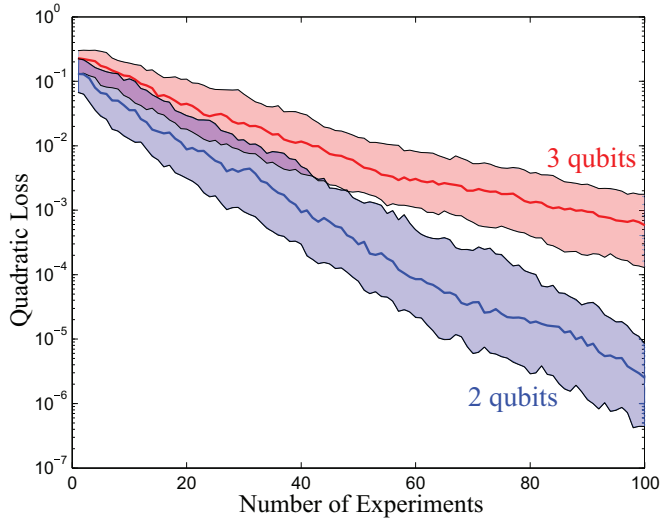
FIG. 4. (Color online) Median quadratic loss as a function of the number of experiments for two- and three-qubit translationally invariant transverse Ising models chosen with x_k uniform in [0,1]. The shaded regions give a 50% confidence interval for the data.
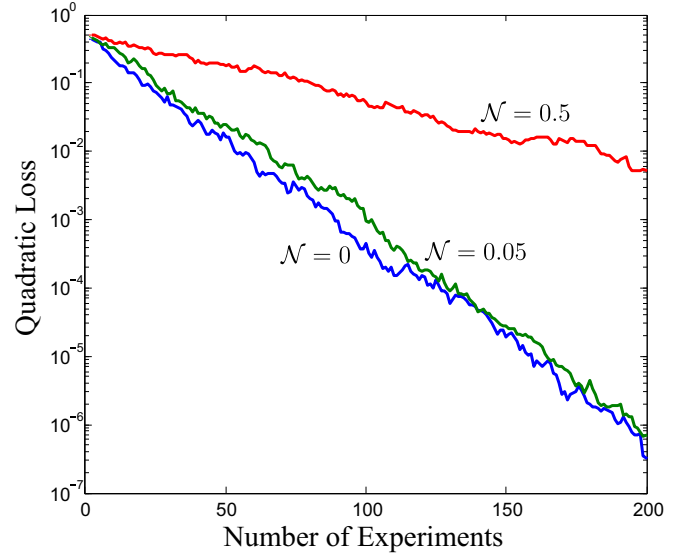

FIG. 5. (Color online) The median quadratic loss plotted for IQLE experiments on the Ising Hamiltonian on a line with four qubits for varying levels of depolarizing noise.

are striking. This emphasizes that finding good experiments that provide high contrast on the model parameters that we are trying to learn is crucial and that QHL, in principle, faces no difficulties in learning Hamiltonians with noncommuting terms. We expect these qualitative features to remain the same even when $n > 3$ because no new symmetries are introduced or broken as we scale up the system.

V. ROBUSTNESS OF QUANTUM HAMILTONIAN LEARNING TO NOISE

In practice, swapping the quantum state out of the uncharacterized quantum system and into the trusted simulator will often be the most error-prone step of an IQLE experiment. The noise may be relatively small in some cases. For example, in superconducting systems or scalable ion traps the trusted simulator could be part of the chip and the untrusted system could be another region that has not been adequately characterized. The noise introduced by transferring the quantum state can be minimal since such architectures naturally permit state swapping. On the other hand, noise in the SWAP operation could also be catastrophic in cases where the trusted simulator is not naturally coupled to the system, such as cases where flying qubits in the form of superconducting resonators or photons must be used as auxiliary resources to couple two simulators. The inevitable question is as follows: "What level of noise can the learning algorithm sustain before it fails?"

We address the question by examining the performance of QHL using IQLE experiments for systems where the noise is known and the trusted simulator is capable of simulating the noise [37,38]. We examine the performance of QHL theoretically and numerically for depolarizing noise, as well as physically realistic models of noise for quantum dots and superconducting circuits. We see that substantial depolarizing noise can be tolerated by the QHL algorithm and realistic noise models for existing swap gates similarly do not substantially impede learning. We do not provide examples for noncom-

muting models here because such numerical experiments are computationally expensive.

A. Depolarizing noise

The robustness of the learning algorithm to depolarizing noise arises similarly from the fact that a large number of particles are used in the SMC approximation, but also because of the fact that we assume that the strength of the depolarizing noise is known. This robustness can be seen quite clearly in Fig. 5, where we show that 50% depolarizing noise only slows the learning process down by a constant factor for random four-qubit Hamiltonians of the form of Eq. (3). In contrast, 5% depolarizing noise led to a negligible change in the scaling of the quadratic loss.

This is surprising at first glance because depolarizing noise implies that the final state of the system is

$$\rho_{\text{final}} = (1 - \mathcal{N}) |\psi(t)\rangle \langle \psi(t)| + \mathcal{N} \mathbb{1}/2^n \qquad (9)$$

for a known value of $\mathcal{N}$. However, this is not a substantial problem if $\mathcal{N}$ is known because we can calculate likelihoods for quantum experiments with a fixed amount of depolarizing noise. This, in effect, means that the worst thing that well-characterized depolarizing noise can do is reduce the visibility of an experiment, which in turn simply slows the learning process by a factor no worse than $\mathcal{N}^2$ [39]. In fact, the rate at which IQLE experiments learn the unknown parameters will typically only be slowed down by a rate proportional to $\frac{1}{1-\mathcal{N}}$ compared to the noise-free case.

It is worth noting that the depolarizing channel commutes with all the operations in an IQLE experiment. This means that it does not matter when the depolarizing noise is introduced to the system, unlike the other noise models that we consider. For convience, then, in our numerical simulations, we have applied the depolarizing noise at the end, represented as an effective visibility as given by Eq. (9).

Let us consider an IQLE experiment where the measurement results are coarse grained into two outcomes, ψ_0 and its orthogonal compliment $\psi_0^\perp$. Then Bayes' theorem states that the expected update to the prior distribution of a given experiment is

$$\mathbb{E}_{y \in \{\psi_0, \psi_0^\perp\}} \left[\frac{\Pr(\boldsymbol{x}|y)}{\Pr(\boldsymbol{x})} \right] = \frac{\Pr^2(\psi_0|\boldsymbol{x})}{\sum_j \Pr(\psi_0|\boldsymbol{x}_j)\Pr(\boldsymbol{x}_j)}$$
$$+ \frac{\Pr^2(\psi_0^\perp|\boldsymbol{x})}{\sum_j \Pr(\psi_0^\perp|\boldsymbol{x}_j)\Pr(\boldsymbol{x}_j)}. \quad (10)$$

If in the absence of noise, $\Pr(\psi_0|\boldsymbol{x}) = A$ and $\Pr(\psi_0^\perp|\boldsymbol{x}) = 1 - A$, then Eq. (9) gives us that for any $\boldsymbol{x}$

$$\Pr(\psi_0|\boldsymbol{x}) = A(1 - \mathcal{N}) + \frac{\mathcal{N}}{2^n}, \quad (11)$$

$$\Pr(\psi_0^\perp|\boldsymbol{x}) = (1 - A)(1 - \mathcal{N}) + \frac{\mathcal{N}(2^n - 1)}{2^n}, \quad (12)$$

in the presence of depolarizing noise.

We then find the following by substituting Eq. (11) and Eq. (12) into Eq. (10) and assuming that $\min_j(1 - A_j) \gg \mathcal{N}/(1 - \mathcal{N})$ along with $\mathcal{N} \ll 2^n$:

$$\mathbb{E}_{D \in \{\psi_0, \psi_0^\perp\}} \left[\frac{\Pr(\boldsymbol{x}|D)}{\Pr(\boldsymbol{x})} \right]$$
$$= \frac{[A(1 - \mathcal{N}) + \mathcal{N}/2^n]^2}{\sum_j [A_j(1 - \mathcal{N}) + \mathcal{N}/2^n]P(\boldsymbol{x}_j)}$$
$$+ \frac{\left[(1 - A)(1 - \mathcal{N}) + \frac{\mathcal{N}(2^n - 1)}{2^n} \right]^2}{\sum_j [(1 - A_j)(1 - \mathcal{N}) + \mathcal{N}(2^n - 1)/2^n]P(\boldsymbol{x}_j)}$$
$$\sim (1 - \mathcal{N}) \left[\frac{A^2}{\sum_j A_j \Pr(\boldsymbol{x}_j)} + \frac{(1 - A)^2}{\sum_j (1 - A_j)\Pr(\boldsymbol{x}_j)} \right]. \quad (13)$$

Therefore, under these assumptions, the expected relative impact of an observation on the posterior probability $\Pr(\boldsymbol{x}|D)$ is a factor of $1 - \mathcal{N}$ smaller than would be expected in the absence of noise. This in turn suggests that the learning rate, as parameterized by γ, scales like $1 - \mathcal{N}$ in the presence of depolarizing noise. This shows theoretically that small amounts of depolarizing noise will not be sufficient to destabilize QHL.

Figure 6 shows that the median value of γ found by fitting the quadratic loss for 200 random Ising models of the form of Eq. (3), for varying d, to $Ae^{-\gamma N}$ scales as $1 - \mathcal{N}$. Twenty thousand particles were used for these experiments. The constant γ clearly represents a characteristic time scale for the learning problem, and hence it is clear that the learning rate is slowed by a factor of $(1 - \mathcal{N})$ for these problems. This agrees with our prior theoretical expectations in the limit where $\Pr(D|\boldsymbol{x}_j) \gg \mathcal{N}/(1 - \mathcal{N})$. It is further worth noting that the learning rate scales roughly as d^{-1}, which suggests that the cost of Hamiltonian learning scales efficiently with the number of qubits in the chain, as also noted in Ref. [9].

This shows that our method is robust to the presence of a well-characterized source of depolarizing noise. It is worth mentioning, however, that the reduced visibility imposed by the depolarizing noise may be especially problematic for QLE experiments since the distribution of outcomes tends to be
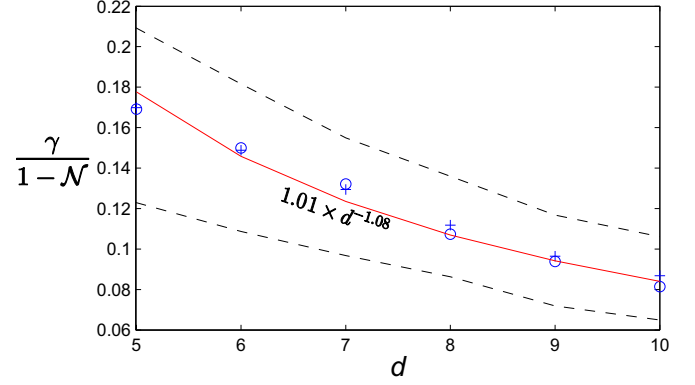

FIG. 6. (Color online) The median value of γ found by fitting the quadratic loss of 200 random local Hamiltonians to $Ae^{-\gamma N}$. The dashed lines represent a 50% confidence interval for the data, and the crosses and circles correspond to $\mathcal{N} = 0.75/d$ and $\mathcal{N} = 0.5/d$, respectively, where d is the number of model parameters.

much flatter in such cases than the corresponding outcome distributions for IQLE experiments. This further underscores the utility of using the trusted simulator in an interactive fashion in such learning protocols.

B. Realistic models for SWAP gate errors

While the above argument lets us reason analytically about the effects of depolarizing noise on inference, in practice the implementation of a SWAP gate need not admit as simple a description as that. To remedy this, a more complete model of the errors in a SWAP implementation can be incorporated into our IQLE simulations. In particular, starting from the cumulant expansion [40], we can simulate the effects of stochastic processes in the environment, open quantum dynamics, and the limited fidelity of a particular shaped pulse sequence derived by optimal control theory [41]. The cumulant expansion generalizes the Magnus expansion [42] to incorporate the effects of stochastic operators and has been used in quantum information to design control methods that are robust to stochastic fields [43]. Numerically evaluating a truncation of the cumulant expansion then gives us a *superoperator* that describes the action of the SWAP gate, so that we can reproduce its effect on the trusted simulation alone by engineering noise on that system [44].

Concretely, the cumulant expansion provides a solution to the ensemble-average time-ordered exponential

$$\hat{\hat{S}}(t) = \left\langle T \exp\left(\int_0^t \hat{\hat{G}}(t)dt \right) \right\rangle, \quad (14)$$

where $\hat{\hat{G}}(t)$ is a stochastic and time-dependent operator in $\mathcal{L}(\mathcal{L}(\mathcal{H}))$ (commonly denoted as a superoperator), such that

$$\hat{\hat{G}}(t)[\rho] = -i[H(t), \rho] + \hat{\hat{D}}[\rho] \quad (15)$$

for a Hamiltonian operator H and a dissipative map $\hat{\hat{D}}$. That is, $\hat{\hat{G}}(t)$ is a superoperator implementing the adjoint map ad $H(t)$ together with the generator of a quantum dynamical semigroup.

Given that $\hat{\hat{G}}$ is a linear operator, we can represent it as a matrix acting on $\mathcal{L}(\mathcal{H})$, the elements of which are *vectors* representing *operators* in the original vector space $\mathcal{H}$. A convenient choice for such vectorizations $|\rho\rangle\!\rangle$ is to stack the columns of ρ to make a vector; for example,

$$\left|\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right\rangle\!\!\right\rangle = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}. \tag{16}$$

More generally, $||i\rangle\langle j|\rangle\!\rangle = \langle j|^{\mathrm{T}} \otimes |i\rangle$ in this convention.

Using this formalism, $\hat{\hat{S}}(t)$ is seen to be a propagator acting on $\mathcal{L}(\mathcal{H})$ that represents the effect of the stochastic process described by $\hat{\hat{G}}(t)$ on vectorizations of mixed states $|\rho\rangle\!\rangle$. Truncating the cumulant expansion at the second order,

$$\hat{\hat{S}}(t) = \exp(\hat{\hat{K}}_1 + \hat{\hat{K}}_2), \quad \text{where} \tag{17}$$

$$\hat{\hat{K}}_1 = \frac{1}{t} \int_0^t dt_1 \langle \hat{\hat{G}}(t_1) \rangle,$$

$$\hat{\hat{K}}_2 = \frac{1}{t^2} T \int_0^t dt_1 \int_0^t dt_2 \langle \hat{\hat{G}}(t_1) \hat{\hat{G}}(t_2) \rangle - \hat{\hat{K}}_1^2.$$

These integrals can readily be numerically computed by characterizing the stochastic process $\hat{\hat{G}}$ in terms of a correlation function, as is discussed at greater length in Ref. [45].

Applying this expansion to the problem of simulating realistic errors in coupling the trusted and untrusted simulators, we start with the models of a superconducting and quantum dot systems described by Puzzuoli *et al.* [45], using the parameters described in the Appendix. Next, we use a gradient ascent optimization method known as the GRAPE algorithm [41] to design a SWAP implementation using the controls admitted by each of these systems. We also consider two superconducting models whose noise strength has been substantially increased, resulting in lower-fidelity implementations for comparison. The quantum dots SWAP implementation uses an XY4 sequence [46] to decouple from the environment. In the superconducting model, we consider both an XY4 and a "primitive" (that is, a single pulse found

using optimal control theory via GRAPE) implementation for the lowest-noise case and the primitive implementation only for the other two models.

We then find the noise map Λ_{noise} for the cumulant-simulated superoperator $\hat{\hat{S}}_{\mathrm{SWAP}}$ for each swap gate used the IQLE experiment (see Fig. 1). In particular, we note that the action of the SWAP gate on the input state $\rho = |\psi\rangle\langle\psi|$ is given by

$$\rho \mapsto \mathrm{Tr}_{\mathrm{untrusted}}(\hat{\hat{S}}_{\mathrm{SWAP}}[\rho \otimes |0\rangle\langle 0|]), \tag{18}$$

where $|0\rangle\langle 0|$ is the initial state of the trusted quantum simulator. By representing the state preparation and partial trace as nonrectangular superoperators, we have that in the supermatrix representation

$$\Lambda_{\mathrm{noise}} = \hat{\hat{S}}_{\mathrm{Tr_{untrusted}}} \circ \hat{\hat{S}}_{\mathrm{SWAP}} \circ \hat{\hat{S}}_{\mathrm{prep}}. \tag{19}$$

Note that even though there is no noise in either the trace or the preparation, it is convenient to keep with superoperators so that composition of maps is represented by simple matrix multiplication.

For a single qubit, we can easily express these superoperators in the column-stacking basis of $\mathcal{L}(\mathcal{L}(\mathcal{H}))$ as

$$\hat{\hat{S}}_{\mathrm{prep}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{20}$$

and

$$\hat{\hat{S}}_{\mathrm{Tr_{untrusted}}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{21}$$

The multiple-qubit superoperators are found from these single-qubit operators using the techniques described in Ref. [47].

While simulating or characterizing a SWAP gate in this manner is not, in general, tractable, recent work demonstrates that we can obtain an *honest approximation* to gates such as the SWAP gate that are restricted to a subclass of efficiently simulatable channels, but which only exaggerate the error [45]. In the case of SMC, this exaggerated error manifests as an additional source of sampling error, such that we can make a trade-off between the error within the SMC procedure and the

accuracy with which we model quantum couplings between the trusted and untrusted registers. Nonetheless, the cost of these simulations limits our numerics to only two qubits.

In addition to allowing for reduction in the resources required for this secondary characterization task, honest approximation allows us to reduce the simulation resources needed to model an IQLE experiment entirely in the trusted register, such that a full open-system simulation need not be necessary. For instance, if we wish to represent the SWAP gate by a Pauli channel immediately following the SWAP, we can use
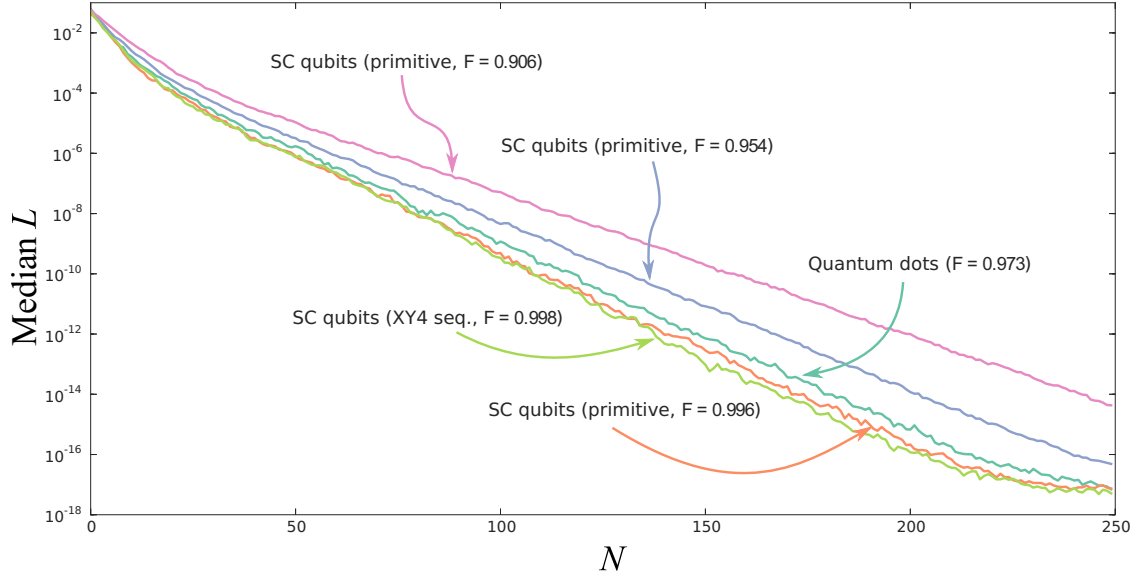
FIG. 7. (Color online) The median value of the loss L in estimating the J coupling between two qubits as a function of the number of measurements N performed for each of the five physical models of the SWAP gate considered. The fidelity F of each implemented SWAP is shown in the inset.

the PUZZUOLI algorithm to find the probabilities with which we should insert perfect single-qubit gates into the trusted register so as to honestly approximate the dynamics of the actual coupling.

The resultant density operator can be computed for an IQLE inference procedure via

$$\rho(t) = e^{iH_-t} \Lambda_{\text{noise}} [e^{-iH(x)t}|+^{\otimes 2}\rangle\langle+^{\otimes 2}|e^{iH(x)t}] e^{-iH_-t}, \quad (22)$$

or equivalently, using the vectorization $|\rho\rangle\rangle$

$$|\rho(t)\rangle\rangle = e^{i\hat{\hat{G}}_-t} \circ \hat{\hat{S}}_{\text{Tr}_{\text{untrusted}}} \circ \hat{\hat{S}}_{\text{SWAP}} \circ \hat{\hat{S}}_{\text{prep}}$$

$$\circ\, e^{-i\hat{\hat{G}}(x)t}||+^{\otimes 2}\rangle\langle+^{\otimes 2}|\rangle\rangle, \quad (23)$$

where Λ_{noise} is the result of a cumulant simulation and where the Hamiltonian used is the J coupling between a pair of qubits,

$$H = J\sigma_z^1\sigma_z^2, \quad (24)$$

equivalent to the $n = 2$ line model above, and where the $\hat{\hat{G}}(x)|\rho\rangle\rangle = |[H(x),\rho]\rangle\rangle$, representing that no noise acts on the system during free evolution before and after the imperfect SWAP gate. We let $\hat{\hat{G}}_- = \hat{\hat{G}}(x_-)$ in analogy to our notation for H.

Figure 7 shows that IQLE experiments continue to gain information at an exponential rate for these realistic levels of noise. The learning is so rapid that after approximately 200 experiments the unknown value of J will be known to six to seven digits of accuracy despite the fact that the SWAP gate infidelities are 4 or 5 orders of magnitude greater than these uncertainties. This robustness largely arises because the model knows the noise model for the SWAP gate. If it did not, then we would not see such large separations between the magnitudes of the uncertainties in J and the gate infidelities. We explore this point in more detail in the next section. It is worth noting before proceeding that, in principle, we do not need to know

the precise noise model of the SWAP gate before performing the experiment: It can be learned simultaneously with J using the exact same approach (see [10] for such an example).

VI. ROBUSTNESS OF ALGORITHM TO ERRORS IN MODEL

When modeling physical systems, it is usually not convenient to include every possible interaction that could exist in the system. For example, in spin systems with dipolar coupling it is common to neglect interactions between distant particles in the system because such interactions decay rapidly with distance. This raises a problem for QHL: It is quite likely that the untrusted quantum system contains couplings that are not modeled by the trusted simulator. It is therefore important to show that QHL will remain stable and continue to infer the best possible model in spite of the fact that the set of allowed models does not contain the true Hamiltonian. We show here that small discrepancies between the model used in the trusted simulator and the true model for the untrusted system are not catastrophic for QHL; in fact, QHL continues to learn H until saturating at a level of uncertainty that scales at most linearly with the number of neglected terms.

It is shown in Ref. [48] that for any two Hamiltonians H and $\tilde{H}$

$$\|e^{-iHt} - e^{-i\tilde{H}t}\| \leqslant \|H - \tilde{H}\|t. \quad (25)$$

This implies that if the Hamiltonian H is used to model the Hamiltonian $\tilde{H}$, then the error in the likelihood function obeys

$$\Delta\Pr(D) := |\langle D| e^{-iHt} |\psi_0\rangle|^2 - |\langle D| e^{-i\tilde{H}t} |\psi_0\rangle|^2$$

$$\leqslant |\langle D| e^{-iHt} - e^{-i\tilde{H}t} |\psi_0\rangle|^2 \quad (26)$$

$$\leqslant \|H - \tilde{H}\|^2 t^2.$$

Equation Eq. (26) implies that the error due to using an approximate Hamiltonian model is negligible provided
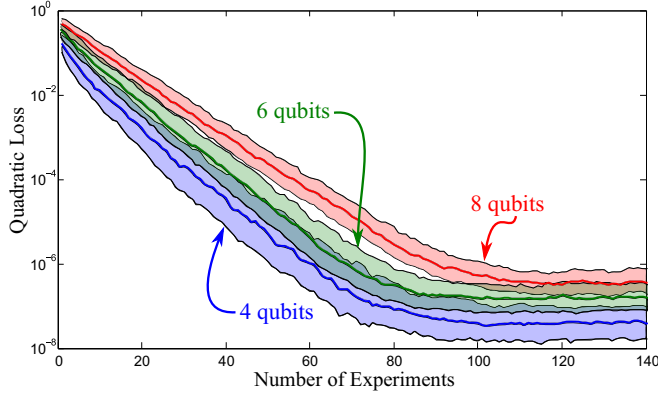
FIG. 8. (Color online) The performance of QHL for the case where the trusted simulator uses an Ising model on the line given that the true Hamiltonian is an Ising model on the complete graph with non-nearest-neighbor interactions on the order of 10^{-4} and nearest-neighbor interactions on the order of 0.5.

$\|H - \tilde{H}\| \ll t^{-1}$. Our use of the PGH implies that the time chosen is (with high probability) approximately the reciprocal of the uncertainty of the uncertainty in the Hamiltonian (i.e., $t \propto \Delta H^{-1}$). The use of an inexact model therefore is not problematic for the inference algorithm unless

$$\|H - \tilde{H}\| \approx \Delta H. \tag{27}$$

In particular, if we parametrize the Hamiltonians via $\boldsymbol{x} \in \mathbb{R}^d$ as $H(\boldsymbol{x})$, then it is sufficient to assert that

$$\min_{\boldsymbol{x}} \|H(\boldsymbol{x}) - \tilde{H}(\boldsymbol{x})\| \ll \Delta H. \tag{28}$$

It is, however, often sufficient in practice to assert that the expectation value over all particles is sufficiently small compared to ΔH.

Also note that if the terms $\mathcal{H}_1, \ldots, \mathcal{H}_R$ are neglected from the Hamiltonian model, then $\min_{\boldsymbol{x}} \|H(\boldsymbol{x}) - \tilde{H}(\boldsymbol{x})\| \leqslant R \max_{j=1,\ldots,R} \|\mathcal{H}_j\|$, which implies that the use of an inexact model will not be problematic if

$$\frac{\Delta H}{R} \gg \max_{j=1,\ldots,R} \|\mathcal{H}_j\|. \tag{29}$$

This implies that the point at which the algorithm ceases to learn varies at most linearly with R (assuming $\|\mathcal{H}_j\|$ is independent of R). Since R will typically vary polynomially with the number of interacting particles in a system, our algorithm remains tractable for physically motivated high-dimensional systems.

We see this behavior clearly illustrated in Fig. 8, where we examine the performance of QHL given that an inexact model is used for the unknown Hamiltonian. In particular, we take

$$H = \sum_{i=1}^{n-1} a_i \sigma_i^z \sigma_{i+1}^z,$$

$$\tilde{H} = \sum_{i=1}^{n-1} a_i \sigma_i^z \sigma_{i+1}^z + \sum_{i=1}^{n-1} \sum_{j=i+2}^{n} b_{i,j} \sigma_i^z \sigma_j^z. \tag{30}$$

The coupling constants a_i are each chosen uniformly from the interval $[-1/2, 1/2]$ and the coupling constants $b_{i,j}$ are
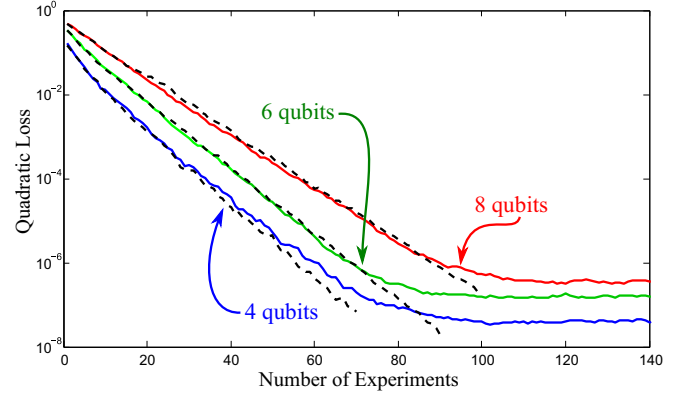

FIG. 9. (Color online) As Fig. 8, with the addition of data (dotted lines) for the performance when the true Hamiltonian only has nearest-neighbor couplings on a line. In contrast, the true Hamiltonian in Fig. 8 contains small non-nearest-neighbor couplings.

each chosen according to a Gaussian distribution with mean zero and standard deviation 10^{-4}. In this case, the models have different dimension, so we compute the quadratic loss by square error in the inferred values of the a_i only.

Note that in practical cases, such as those based on dipolar Hamiltonians, next-nearest-neighbor interactions are often on the order of 10^{-1}. We use 10^{-4} to illustrate the qualitative difference between the regime in which the algorithm is learning the Hamiltonian and the region where learning ceases by creating a stronger separation between included and neglected terms.

We note that for the data considered in Fig. 8 learning proceeds at an exponential rate until saturating at a value that is approximately on the order of $\|H - \tilde{H}\|^2$. This shows that QHL is still valuable in cases where an inexact model is used by the simulator, which further underscores the utility of this procedure in finding Hamiltonian models for unknown quantum systems. Figure 9 shows us that, before saturation, the differences in the performance of QHL are negligible relative to the experimental uncertainties in the performance seen in Fig. 8. The use of an approximate Hamiltonian model does not substantially degrade the performance of the learning algorithm until the uncertainty in the inference is comparable to the sum of the magnitudes of the neglected couplings. Such plateaus do not represent a failure of QHL; on the contrary, they point to failures in our modeling of the system and that new physics may be required to understand the system in question.

Conversely, one could also consider the problem of what happens when Bayesian inference is used when there are too many parameters. It is conceivable in such cases that, rather than outputting the simplest possible model for the Hamiltonian, QHL outputs an unnecessarily complicated model that nonetheless predicts the experimental results with high probability. Such examples are known to not be typical of Bayesian inference [49]. In fact, Bayesian inference includes Occam's razor by implicitly penalizing unnecessarily complicated models for the data. We discuss this next.

A. Learning the best Hamiltonian model

Our results so far have shown that small imperfections do not typically prevent the QHL algorithm from learning

the correct Hamiltonian model for a system, but a problem remains: Can we use the QHL algorithm to find an accurate and concise model for a unknown quantum system where the form of the Hamiltonian is not even known? Consider the case that the true model is $\tilde{H}$, but that we posit the model H. Then we find that in some sense, the algorithm still learns the "best" set of parameters within the set of allowed parameters of the model H. However, since the "true" parameters lie outside the set of those allowed by H, the distance (as measured by the quadratic loss) of the estimated parameters in H to the true parameters in $\tilde{H}$ is bounded, as we show in Figs. 8 and 9.

This behavior is, in fact, desirable. Since modeling physical systems always necessitates some approximation, the optimal estimation procedure ought to find the parameters within the allowed set that is closest to those true parameters outside it. We can do more, however. In addition to behaving near optimally within each model, our algorithm naturally accommodates *model selection*, whereby it ranks models according to their relative plausibility. That is, the algorithm simultaneously solves the parameter estimation problem and the metaproblem of finding best model while minimizing the *effective* number of model parameters used. We illustrate this in the case where our hypothetical model H is tested against the true model $\tilde{H}$.

To this end, we compare the probabilities, given the data, that either H or $\tilde{H}$ is true: $\Pr(H|D)$ versus $\Pr(\tilde{H}|D)$. Using the Bayes rule we have

$$\Pr(H|D) = \frac{\Pr(D|H)\Pr(H)}{\Pr(D)}. \qquad (31)$$

Taking the ratio is then convenient as the normalization factor cancels,

$$\frac{\Pr(\tilde{H}|D)}{\Pr(H|D)} = \frac{\Pr(D|\tilde{H})}{\Pr(D|H)}\frac{\Pr(\tilde{H})}{\Pr(H)}, \qquad (32)$$

which is called the *posterior odds ratio* and forms the basis for comparing models [50]. If the posterior odds ratio is larger than 1, the evidence favors $\tilde{H}$ and vice versa if the value is less than 1. The last fraction is called the *prior odds*, and the unbiased choice favoring neither model is to set this term equal to 1. Doing so leaves us with

$$\frac{\Pr(\tilde{H}|D)}{\Pr(H|D)} = \frac{\Pr(D|\tilde{H})}{\Pr(D|H)}, \qquad (33)$$

which is called the *Bayes factor* [51].

The use of the Bayes factor for model selection is well motivated by other model selection criteria. The most commonly used model selection technique is the Akaike information criterion (AIC) as it assumes the simple form in $\mathrm{AIC} = \max_H \Pr(D|H) - d$, where d is the number of parameters in the model [52]. The preferred model is the one with largest value of AIC. Thus, it is clear how models with more parameters are penalized. The Bayesian approach we advocate above is more generally applicable. However, it is less obvious how the Bayes factor includes an Occam's razor to penalize more complex models. The simplest way to see the effect is to consider the asymptotics of the Bayes factor terms. Ignoring terms constant in N, the asymptotic marginal

likelihood is (see, for example, [49])

$$\Pr(D) = \max_H \Pr(D|H) - \frac{d}{2}\ln N, \qquad (34)$$

which is the well-known Bayesian information criterion or BIC [52]. Noticing the striking similarity to the AIC mentioned above, it is now clear that the Bayesian approach also penalizes extra free parameters. This asymptotic form clarifies how additional parameters are penalized; our SMC algorithm approximates the exact (to within numerical accuracy), nonasymptotic distribution.

For an arbitrary Hamiltonian, H, $\Pr(D|H)$ is called the *marginalized likelihood* since we can obtain its value via marginalizing the likelihood function over the model parameters of H:

$$\Pr(D|H) = \mathbb{E}_{\boldsymbol{x}|H}(\Pr(D|\boldsymbol{x};H)). \qquad (35)$$

This value can be computed online using the likelihood values that are computed in QHL (or more generally a SMC algorithm). To show this, consider two pieces of data, $D = \{d_2, d_1\}$ and

$$\begin{aligned}\Pr(d_2, d_1|H) &= \Pr(d_2|d_1; H)\Pr(d_1|H), \\ &= \mathbb{E}_{\boldsymbol{x}|d_1;H}(\Pr(d_2|\boldsymbol{x};H))\mathbb{E}_{\boldsymbol{x}|H}(\Pr(d_1|\boldsymbol{x};H)).\end{aligned} \qquad (36)$$

These are expectations over the current distribution, which is exactly what the SMC algorithm is designed to efficiently approximate. One might suspect that being expectations over the likelihood, such calculations would require more costly simulations of the model. However, note that, under the SMC approximation,

$$\Pr(D|H) = \mathbb{E}_{\boldsymbol{x}|H}(\Pr(D|\boldsymbol{x};H)) \approx \sum_{j=1}^{|\{\boldsymbol{x}_i\}|} \Pr(D|\boldsymbol{x}_j)w_j, \qquad (37)$$

which is exactly the normalization of the weights after the update rule. That is, the marginal likelihood is already explicitly calculated as part of the SMC algorithm used in QHL.

A natural way to use the Bayes factor to perform model selection is to simultaneously run two copies of QHL: one using model H and the other using model $\tilde{H}$. We refer to the model output by using QHL with H as the null model and the model output by QHL with $\tilde{H}$ the alternate model. The PGH is used for a guess experiment using data from the null model at each step of the QHL algorithm by default. The same experimental parameters are also used in when QHL is applied to the alternate model (even if H and $\tilde{H}$ have fundamentally different forms and/or parametrizations). The Bayes factor is then computed by taking the expectation values of the likelihoods computed (using data from the quantum simulator) in both cases and dividing the two results. If this ratio is greater than 1, then the roles of the null and alternate models are reversed: The alternate model now dictates the choice of experimental parameters in QHL. These steps are repeated until the uncertainty in the Hamiltonian favored by the posterior odds ratio is sufficiently low.

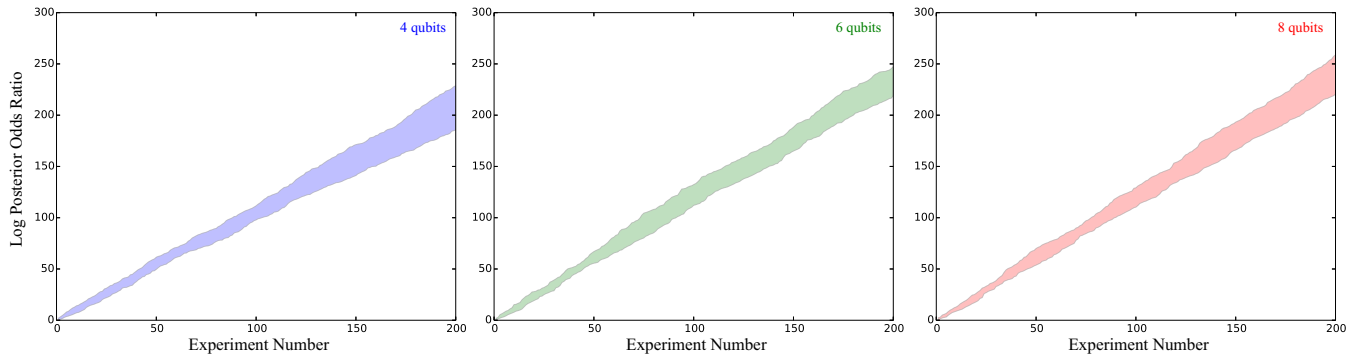To illustrate this, consider the example of the previous section (and presented in Figs. 8 and 9). The incorrect

FIG. 10. (Color online) The logarithm of the posterior odds ratio of the true model to the reduced model. Here the reduced model is a Hamiltonian with only nearest-neighbor couplings on a line while the true Hamiltonian contains small non-nearest-neighbor couplings (as described also in Fig. 8). The bands encompass *all* data.

model learns at an exponential rate but then saturates as the true parameters lie outside the set allowed by the posited model. Suppose, however, that we use competing models, which could be realized as competing simulators or the same simulator with restricted controls. To be clear, we take the true Hamiltonian to be an Ising model on the complete graph, $\tilde{H}$ from Eq. (30), and take the model Hamiltonian to be an Ising model on the line, H from Eq. (30). The initial prior is taken to be uniform over both a_i and the true Hamiltonian is drawn from a similar distribution over a_i and $b_{i,j}$ for each sample. We then use the Bayesian model selection approach outlined above to decide which model is best. Figure 10 shows the logarithm of the posterior odds ratio Eq. (32) of the true (non nearest-neighbor Hamiltonian) model to the reduced model (with only nearest-neighbor couplings). By the 200th measurements, the odds are at least an astounding $10^{120} : 1$ against the reduced model suggesting we can also rapidly distinguish good models from bad. The data for the dual problem—when the true model contains *fewer* parameters—is presented in Fig. 11. This corresponds to switching the roles of the true and model Hamiltonians in the previous example. Again, the algorithm rapidly learns the true model, which in this case is also hedging against overfitting (Occam's razor).

VII. CONCLUSION

We show in this paper, both numerically and theoretically, that even imperfect quantum simulators are powerful resources for quantum computation and the characterization of quantum systems. We show that QHL using interactive likelihood estimation can tolerate substantial amounts of depolarizing before failing to provide useful information about the Hamiltonian. We also show that realistic errors in the SWAP gate do not pose a problem and that the learning algorithm also can be applied in cases where the model does not commute. The algorithm is also shown to be robust even in the presence of small unmodeled terms in the actual Hamiltonian; and we see in a numerical example that the algorithm succeeds in finding approximate Hamiltonians that are maximally close to the true Hamiltonian, which has interactions that are not present in the model. Such cases are particularly intriguing since they can point to failures in the physical models used to describe systems. The particular way in which the model fails can also be learned by incorporating model selection to distinguish good models from bad.

These results provide a proof of principle that realistic quantum simulators can be used in concert with Bayesian inference to address certain seemingly intractable problems in Hamiltonian estimation, illustrating that quantum resources
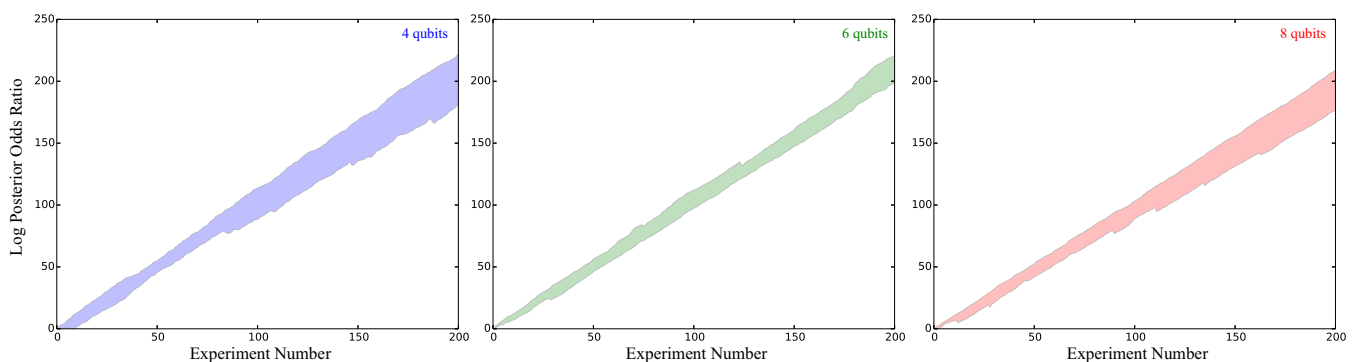

FIG. 11. (Color online) The logarithm of the posterior odds ratio for the problem dual to that in Fig. 10. Here, the true model contains only nearest-neighbor couplings. The compared model contains all coupling terms and is hence overfit. The plot shows that QHL rapidly detects overfitting and urges us to select the reduced model.

can be very useful for characterizing and controlling quantum information processors. This work by no means gives the final answer on Hamiltonian inference; nor does it provide a replacement for strong experimental intuition. On the contrary, we see that strong understanding of the physics of the system is essential for optimizing the learning rate for the QHL algorithm. From this perspective, our work raises the possibility of a future where classical machine learning algorithms are employed according to our best knowledge of physics and in tandem with quantum devices in order to learn properties of unknown systems, certify untrusted quantum devices, and perhaps even to discover new physics.

There are a number of natural extensions of this work. First of all, although the PGH often yields very good experiments, it does not necessarily pick ones that are locally optimal. Locally optimal experiments could be found by minimizing the Bayes' risk using algorithms such as conjugate gradient optimization or differential evolution similar to [10,21]. Second, many of the steps in the QHL algorithm could be substantially sped up by using a quantum computer. A specialized version of the algorithm that incorporates techniques such as amplitude estimation [53] and gradient estimation [54] may show that quantum resources can be leveraged to provide even greater advantages than those considered here. Finally, although the median quadratic loss tends to behave very well for our algorithm, in relatively rare cases the algorithm can stop learning altogether. Finding new ways to detect and recover from these errors would be invaluable for reducing the number of times the algorithm must be run in order to have confidence that the resultant Hamiltonian can actually be trusted.

Our work thus establishes a promising avenue of research in quantum information processing. In particular, our work demonstrates that quantum information processing devices will be useful in the development of further advances in quantum information processing by making possible the use of quantum simulation as a resource. This capability is especially important now, as the scale of quantum information processing devices grows beyond our classical simulation capacity; hence, the ability to use quantum resources to inexpensively characterize large quantum information processors may prove vital for the development of the next generation of quantum computers and quantum simulators.

APPENDIX: PARAMETERS USED FOR PHYSICAL SIMULATIONS

We use the qubit model of [45] to obtain realistic SWAP gates for superconducting systems. This model expresses the Hamiltonian as the sum of two component terms given by the single-qubit Hamiltonian

$$H^{(i)}(t) = \frac{1}{2}\{B(t)[1 + \beta_1(t)] + \beta_2(t)\}\,\sigma_z^{(i)}$$
$$+ \frac{1}{2}[1 + \alpha(t)]\big\{\cos[\phi(t)]\sigma_x^{(i)} + \sin[\phi(t)]\sigma_y^{(t)}\big\}, \quad \text{(A1)}$$

and by the two-qubit interaction Hamiltonian

$$H^{(ij)} = -\frac{1}{2}C(t)[1 + \gamma(t)]\sigma_z^{(i)}\sigma_z^{(j)}. \quad \text{(A2)}$$

In this model, A, B, C, and ϕ are time-varying controls, while α, β, and γ are taken to be zero-mean Gaussian processes with $1/f$ power spectral densities having amplitudes denoted by Γ and cutoffs $\Lambda^{(l)}, \Lambda^{(u)}$.

While we mainly consider examples of primitive gates, consisting of a single shaped pulse derived from optimal

TABLE I. Noise parameters used for superconducting gates.

	XY4	Primitive		
Fidelity with SWAP	0.998	0.996	0.954	0.906
Discretization time step (s)	2.5×10^{-10}	1×10^{-10}	1×10^{-10}	1×10^{-10}
T_1 (s)	10^{-4}	10^{-5}	10^{-5}	10^{-5}
Γ_α (Hz)	3×10^4	10^4	10^6	10^6
Γ_{β_1} (Hz)	3×10^4	0	10^6	10^6
Γ_{β_2} (Hz)	$10^6/2\pi$	10^4	10^6	1.5×10^6
$\Lambda_\alpha^{(l)}$ (Hz)	$1/2\pi$	$1/2\pi$	$1/2\pi$	$1/2\pi$
$\Lambda_\alpha^{(u)}$ (Hz)	10^9	10^9	10^9	10^9
$\Lambda_{\beta_1}^{(l)}$ (Hz)	$1/2\pi$	$1/2\pi$	$1/2\pi$	$1/2\pi$
$\Lambda_{\beta_1}^{(u)}$ (Hz)	10^9	10^9	10^9	10^9
$\Lambda_{\beta_2}^{(l)}$ (Hz)	$1/2\pi$	$1/2\pi$	$1/2\pi$	$1/2\pi$
$\Lambda_{\beta_2}^{(u)}$ (Hz)	10^9	10^9	10^9	10^9
Γ_γ (Hz)	$1.2 \times 10^3/2\pi$	$1.2 \times 10^3/2\pi$	$1.2 \times 10^5/2\pi$	$1.2 \times 10^5/2\pi$
$\Lambda_\gamma^{(l)}$ (Hz)	$1/2\pi$	$1/2\pi$	$1/2\pi$	$1/2\pi$
$\Lambda_\gamma^{(u)}$ (Hz)	10^9	10^9	10^9	10^9

control theory [41], we also include a higher-fidelity gate obtained by interlacing with an XY4 decoupling sequence [46] for comparison. This allows us to reason separately about the impact of the Lindblad generators $L^{(i)} = \frac{1}{2\sqrt{T_1}}(\sigma_x^{(i)} + i\sigma_y^{(i)})$ and the impact of stochastically varying control fields, given that interlacing with the XY4 sequence refocuses away much of the stochastic contributions.

We also include an example drawn from the quantum dots model of [45]. The primary source of noise in this model is the inclusion of stochasticity in the voltage detuning and Zeeman splitting processes, giving the single-qubit Hamiltonian

$$H^{(i)}(t) = \frac{1}{2}\frac{A(t) + \alpha(t)}{\sqrt{1 + \exp\left(\frac{B(t)}{B_1} - B_2\right)}}\sigma_x^{(i)}$$
$$+ \frac{1}{2}\frac{B(t) - B_0 + \beta(t)}{1 + \exp\left(-\left[\frac{B(t)}{B_1} - B_2\right]\right)}\sigma_z^{(i)}, \quad \text{(A3)}$$

where A and B are control parameters for detuning and splitting, respectively, and where α and β are again stochastically varying noise sources. We then simulate a SWAP gate for the quantum dot model using the parameters for the two qubit gates in Ref. [45]. Apart from providing an example with a relatively low-fidelity SWAP gate, this example also illustrates that our results are not predicated on a specific model being used.

In order to generate a range of different gate qualities, and hence demonstrate the effectiveness of our algorithm in a variety of different physically realistic scenarios, we increase the noise from those gates used in Ref. [45]. In particular, we shorten the relaxation time T_1, and increase the $1/f$ noise amplitudes Γ. The former causes the dissipative process acting on our system to become stronger, while the latter increases the stochasticity of the control fields. Varying noise parameters in this way, we show gates with fidelities ranging from $F \approx 0.9$ to nearly ideal. In Table I, we list the noise parameters used in the numerical experiments.

[1] 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).

[2] S. Korenblit, D. Kafri, W. C. Campbell, R. Islam, E. E. Edwards, Z.-X. Gong, G.-D. Lin, L. M. Duan, J. Kim, K. Kim *et al.*, New J. Phys. **14**, 095024 (2012).

[3] C.-C. J. Wang, A. C. Keith, and J. K. Freericks, Phys. Rev. A **87**, 013422 (2013).

[4] R. Islam, C. Senko, W. C. Campbell, S. Korenblit, J. Smith, A. Lee, E. E. Edwards, C.-C. J. Wang, J. K. Freericks, and C. Monroe, Science **340**, 583 (2013).

[5] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, Nat. Photon. **7**, 540 (2013).

[6] P. Hauke, F. M. Cucchietti, L. Tagliacozzo, I. Deutsch, and M. Lewenstein, Rep. Prog. Phys. **75**, 082401 (2012).

[7] C. Gogolin, M. Kliesch, L. Aolita, and J. Eisert, arXiv:1306.3995.

[8] C. Ududec, N. Wiebe, and J. Emerson, Phys. Rev. Lett. **111**, 080403 (2013).

[9] N. Wiebe, C. Granade, C. Ferrie, and D. G Cory, arXiv:1309.0876.

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

[11] D. Gottesman, http://pirsa.org/08080052/.

[12] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

[13] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nat Commun **1**, 149 (2010).

[14] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010).

[15] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, New J. Phys. **14**, 095022 (2012).

[16] O. Landon-Cardinal and D. Poulin, New J. Phys. **14**, 085004 (2012).

[17] M. Mohseni and D. A. Lidar, Phys. Rev. Lett. **97**, 170501 (2006).

[18] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **104**, 063603 (2010).

[19] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **107**, 233601 (2011).

[20] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Phys. Rev. A **84**, 052315 (2011).

[21] N. B. Lovett, C. Crosnier, M. Perarnau-Llobet, and B. C. Sanders, Phys. Rev. Lett. **110**, 220501 (2013).

[22] M. West, J. R. Stat. Soc. Ser. B (Methodological) **55**, 409 (1993).

[23] J. Liu and M. West, *Combined Parameter and State Estimation in Simulation-Based Filtering* (Springer-Verlag, Berlin, 2000).

[24] C. Ferrie, New J. Phys. **16**, 023006 (2014).

[25] 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).

[26] P. Richerme, C. Senko, S. Korenblit, J. Smith, A. Lee, W. C. Campbell, and C. Monroe, Phys. Rev. Lett. **111**, 100506 (2013).

[27] B. P. Lanyon, C. Hempel, D. Nigg, M. Mller, 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).

[28] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science **309**, 1704 (2005).

[29] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri *et al.*, Nat. Chem. **2**, 106 (2010).

[30] R. Gerritsma, G. Kirchmair, F. Zähringer, E. Solano, R. Blatt, and C. F. Roos, Nature (London) **463**, 68 (2010).

[31] 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).

[32] N. Wiebe, D. W. Berry, P. Høyer, and B. C. Sanders, J. Phys. A: Math. Theor. **44**, 445308 (2011).

[33] D. W. Berry and A. M. Childs, Quantum Inf. Comput. **12**, 29 (2012).

[34] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

[35] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, Science **302**, 2098 (2003).

[36] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009).

[37] S. Lloyd and L. Viola, Phys. Rev. A **65**, 010101 (2001).

[38] D.-S. Wang, M. C de Oliveira, D. W. Berry, and B. C. Sanders, Phys. Rev. Lett. **111**, 130504 (2013).

[39] C. Ferrie, C. E. Granade, and D. G. Cory, Quant. Info. Proc. **12**, 611 (2013).

[40] R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).

[41] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).

[42] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).

[43] P. Cappellaro, J. S. Hodges, T. F. Havel, and D. G. Cory, J. Chem. Phys. **125**, 044514 (2006).

[44] E. M. Fortunato, L. Viola, J. S. Hodges, G. Teklemariam, and D. G. Cory, New J. Phys. **4**, 5 (2002).

[45] D. Puzzuoli, C. Granade, H. Haas, B. Criger, E. Magesan, and D. G. Cory, Phys. Rev. A **89**, 022306 (2014).

[46] T. Gullion, J. Magn. Reson. **85**, 614 (1989).

[47] C. J. Wood, J. D. Biamonte, and D. G. Cory, arXiv:1111.6950.

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

[49] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*, 1st ed. (Cambridge University Press, cambridge, UK, 2003).

[50] H. Jeffreys, *Theory of Probability*, 2nd ed. (Oxford University Press, Oxford, UK, 1939).

[51] R. E. Kass and A. E. Raftery, J. Am. Stat. Assoc. **90**, 773 (1995).

[52] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York, 2002).

[53] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, arXiv:quant-ph/0005055.

[54] S. P. Jordan, Phys. Rev. Lett. **95**, 050501 (2005).

[55] E. Jones, T. Oliphant, and P. Peterson *et al.*, SCIPY: Open source scientific tools for PYTHON, 2001, http://scipy.org/

[56] P. Peterson, Int. J. Comput. Sci. Eng. **4**, 296 (2009).

[57] C. Ferrie and C. Granade *et al.*, QINFER: Library for statistical inference in quantum information, 2012, http://github.com/csferrie/python-qinfer