



FAST TRACK COMMUNICATION • OPEN ACCESS

Quantum bootstrapping via compressed quantum Hamiltonian learning

To cite this article: Nathan Wiebe *et al* 2015 *New J. Phys.* **17** 022005

View the [article online](#) for updates and enhancements.

Related content

- [Robust online Hamiltonian learning](#)
Christopher E Granade, Christopher Ferrie, Nathan Wiebe *et al*.
- [Structured filtering](#)
Christopher Granade and Nathan Wiebe
- [Practical Bayesian tomography](#)
Christopher Granade, Joshua Combes and D G Cory

Recent citations

- [Quantum machine learning](#)
Jacob Biamonte *et al*
- [QInfer: Statistical inference software for quantum applications](#)
Christopher Granade *et al*
- [Direct certification of a class of quantum simulations](#)
D Hangleiter *et al*



FAST TRACK COMMUNICATION

OPEN ACCESS

RECEIVED

12 December 2014

REVISED

16 January 2015

ACCEPTED FOR PUBLICATION

23 January 2015

PUBLISHED

13 February 2015

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

Any further distribution of
this work must maintain
attribution to the author
(s) and the title of the
work, journal citation and
DOI.



Quantum bootstrapping via compressed quantum Hamiltonian learning

Nathan Wiebe¹, Christopher Granade^{2,3} and D G Cory^{3,4,5,6}¹ Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA² Department of Physics, University of Waterloo, Ontario N2L 3G1, Canada³ Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada⁴ Department of Chemistry, University of Waterloo, Ontario N2L 3G1, Canada⁵ Perimeter Institute, University of Waterloo, Ontario N2L 2Y5, Canada⁶ Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, CanadaE-mail: nawiebe@microsoft.com and cgranade@cgranade.com**Keywords:** quantum information, characterization, Lieb–Robinson bounds, machine learning, quantum simulation

Abstract

A major problem facing the development of quantum computers or large scale quantum simulators is that general methods for characterizing and controlling are intractable. We provide a new approach to this problem that uses small quantum simulators to efficiently characterize and learn control models for larger devices. Our protocol achieves this by using Bayesian inference in concert with Lieb–Robinson bounds and interactive quantum learning methods to achieve compressed simulations for characterization. We also show that the Lieb–Robinson velocity is epistemic for our protocol, meaning that information propagates at a rate that depends on the uncertainty in the system Hamiltonian. We illustrate the efficiency of our bootstrapping protocol by showing numerically that an 8 qubit Ising model simulator can be used to calibrate and control a 50 qubit Ising simulator while using only about 750 kilobits of experimental data. Finally, we provide upper bounds for the Fisher information that show that the number of experiments needed to characterize a system rapidly diverges as the duration of the experiments used in the characterization shrinks, which motivates the use of methods such as ours that do not require short evolution times.

1. Introduction

Rapid progress has been made within the last few years towards building computationally useful quantum simulators or computers, which promise to revolutionize the ways in which we solve problems in chemistry and material science, data analysis and cryptography [1–5]. Despite this, looming challenges involving calibrating and debugging quantum devices suggests another possible application for a small scale quantum computer: designing a larger quantum computer. This application is increasingly relevant as experiments push towards building fault-tolerant devices [6] and demonstrating large scale verifiable quantum computing protocols [7].

This task can be quite challenging classically. Simply characterizing the Hamiltonian dynamics of the system via tomography is inefficient [8] and existing efficient methods such as [9] require an amount of data that scales polynomially with the error tolerance, are not known to be error robust, and are only efficient for specific classes of Hamiltonians and measurements. This can render them impractical for problems like designing controls for quantum systems where exacting error tolerance and low fault sensitivity is required. Other methods are error robust and use a logarithmic amount of data, but also require performing quantum simulations that are intractable classically [10, 11]. The use of quantum simulators has been proposed as a solution to this problem [12, 13] but such schemes do not provide a means for characterizing and controlling a large quantum system because they require a simulator that is *at least as large* as the system being characterized. Other schemes have been introduced that allow a small quantum system to efficiently *certify* that a large quantum system implements a set of quantum gates or prepares a given state [7, 14–16]. These schemes are inspired by multi-prover systems

and output a certificate that states whether the errors in the gates are above or below a threshold. Such protocols are rigorous, require very weak assumptions about the errors in the larger system and allow the verifier to use a rudimentary quantum device. Unfortunately, they can also be computationally expensive and are difficult to apply to the case of certifying analog simulation. In particular, existing certification schemes that use small quantum verifiers do not *characterize* the larger system's Hamiltonian dynamics.

We provide a framework in this paper for overcoming the aforementioned obstacles to quantum device characterization. The idea behind our approach is to use a small simulator as a point of reference to compare the dynamics of the larger system against. We achieve this by using an interactive protocol wherein the dynamics of subsystems of the larger device are measured against the dynamics of the smaller system; thereby allowing us to infer a model for the larger system using the data collected about the relative dynamics of the two models on the individual subsystems. We call this process compressed quantum Hamiltonian learning (cQHL).

It is often insufficient to give a model of a system relative to an experimental device for which no mathematical model is known. To this end, we consider starting with a small quantum simulator for which a firm mathematical model is known. This system, which we call the *trusted simulator* [12, 13] is the key to allowing our method to provide an absolute, rather than a relative, model for the larger system's quantum dynamics.

Compressed Hamiltonian learning in turn leads to the ability to learn models of *control* dynamics, rather than simply the internal dynamics intrinsic to the untrusted device under study. This inferred model for control dynamics then allows cQHL to be performed again, with the previously-untrusted system now being used as the trusted simulator. That is, cQHL directly enables *quantum bootstrapping*, the process of iteratively building larger trusted simulators out of smaller trusted simulators.

To summarize, we provide two distinct applications:

Compressed QHL: Learning a Hamiltonian model for a large quantum system using a small quantum simulator.

Quantum bootstrapping: Designing and calibrating controls for a quantum system with rapidly decaying interactions using a smaller quantum simulator.

These applications are efficient if the unknown Hamiltonian belongs to an efficiently parameterized class of model Hamiltonians for which the interaction strengths between subsystems decay rapidly with distance, experiments are chosen such that the resultant distribution of likelihoods is far from uniform and that the control maps in the bootstrapping case are well conditioned. Efficiency also requires that the resampling algorithms used in the methods do not introduce substantial error in the system. This assumption empirically holds for non-degenerate learning problems, such as those we consider here and in prior work [12, 13].

The remainder of the paper is laid out as follows. We first provide a review of Bayesian methods for Hamiltonian characterization and quantum Hamiltonian learning (QHL) in section 2. We then discuss cQHL in section 3 and quantum bootstrapping in section 4. We then present in section 5 numerical results for bootstrapping and compressed quantum Hamiltonian in an important special case, characterizing and controlling 50 qubit Ising model simulators before concluding.

2. Review of Bayesian characterization and QHL

In developing cQHL, and hence quantum bootstrapping, we will use Bayesian particle filters as a subroutine. Here, we briefly review these methods in a classical context, as well as with the inclusion of quantum resources.

2.1. Bayesian characterization and sequential Monte Carlo

Bayesian methods have been used in a wide range of quantum information applications and experiments; for instance, to discriminate [17] or estimate states [18, 19], to incorporate models of noisy measurements [20], to characterize drifting frequencies [21], and to estimate Hamiltonians [10, 22–24]. They are particularly well-suited for quantum information, owing to their generality, robustness and the ease with which prior information can be incorporated into the algorithm. Moreover, Bayesian approaches have been shown to allow for near-optimal Hamiltonian learning in simple analytically tractable cases [10, 25].

Bayes' theorem provides the proper way to re-assess, or update, prior beliefs about the Hamiltonian for a system given an experimental outcome and a distribution describing prior beliefs. In particular,

$$\Pr(H|\text{data}) = \frac{\Pr(\text{data}|H)}{\Pr(\text{data})} \Pr(H), \quad (1)$$

where $\Pr(H|\text{data})$ is called the posterior distribution, $\Pr(H)$ is the prior distribution that encodes our initial beliefs about H and where $\Pr(\text{data}|H)$ is the likelihood function, which computes the probability that the observed data would occur if the Hamiltonian H correctly modeled the system. The likelihood function can be estimated by sampling from a quantum simulator for the Hamiltonian H and thus Bayesian inference causes Hamiltonian learning to reduce to a Hamiltonian simulation problem [10, 12, 13, 25].

Once the posterior distribution is found, an estimate of the Hamiltonian, \hat{H} , is given by the expectation over the posterior

$$\hat{H} := \mathbb{E}[H|\text{data}] = \int H \Pr(H|\text{data}) dH. \quad (2)$$

This integral is unlikely to be analytically tractable in practice, as it requires integrating the likelihood function $\Pr(\text{data}|H)$ over H . Monte Carlo integration, on the other hand, can be much more practical.

The sequential Monte Carlo algorithm (also known as a *particle filter*) provides a means of sampling from an inaccessible distribution using a transition kernel from some initial distribution [26]. We can sample from the posterior by using Bayes' rule as the SMC transition kernel, given samples from a *prior* distribution and evaluations of the likelihood function. Integrals over the posterior can then be *approximated* by using these samples, which allows quantities such as \hat{H} to be efficiently estimated. SMC has seen use in a range of quantum information tasks, including state estimation [19], frequency and Hamiltonian learning [10], benchmarking quantum operations [27], and in characterizing superconducting device environments [11]. Similar methods have also been applied to quantum error correction [28].

Hamiltonians are not usually represented explicitly as matrices when using SMC algorithms and are instead parameterized by a vector \mathbf{x} of model parameters such that $H = H(\mathbf{x})$. This representation allows for parameter reduction with prior information and can include effects outside of a purely quantum formalism, such as control distortions or stochastic fluctuations in measurement visibility. It also has the advantage that Hamiltonian learning is possible even in cases where matrix representations of individual terms in the Hamiltonian are not formally known.

Concretely, the SMC algorithm approximates prior and posterior distributions by weighted sums of delta-functions

$$\Pr(\mathbf{x}) \approx \sum_{i=1}^N w_i \delta(\mathbf{x} - \mathbf{x}_i), \quad (3)$$

such that the current state of knowledge can be tracked online using a classical computer to record a list of *particles*, each corresponding to a hypothesis \mathbf{x}_i , and having a relative weights $\{w_i\}$. These weights are then updated by substituting the SMC approximation (3) into Bayes' rule (1) to obtain

$$w_i \mapsto w_i \Pr(d|\mathbf{x}_i), \quad (4)$$

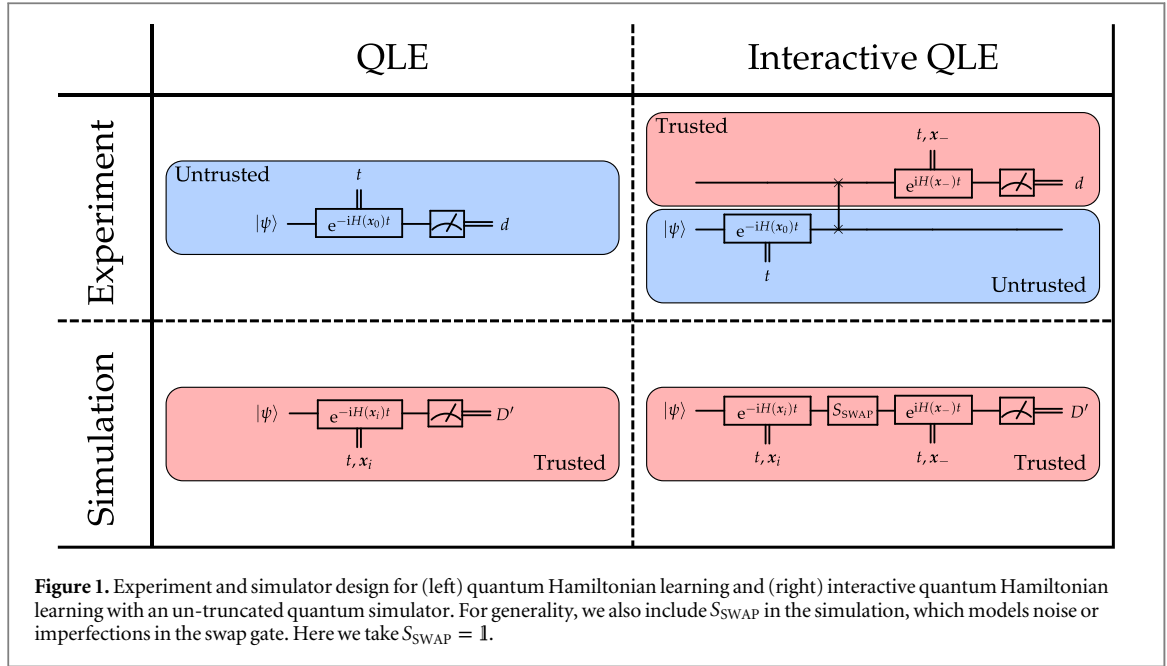
where d is an observation from the experimental system.

Over time, the particle weights for the majority of the particles will diminish as the SMC algorithm becomes more confident that certain hypotheses are wrong. This reduces the total effective number of particles in the posterior distribution and ultimately prevents learning. This issue is addressed by using a resampling algorithm, which draws a new set of uniformly weighted SMC particles that approximately maintain the mean and covariance matrix of the posterior distribution [29].

Although the remainder of the process of learning using the SMC process is rigorous and well understood, error estimates are not known for the resampling step. Furthermore, resampling methods such as the Liu and West algorithm [29] have known pathologies where they can fail if they are provided with a multi-modal distribution or if an improbable sequence of measurements leads to a resampling step that causes the SMC particle cloud to have no support over the true model. These issues are often addressed by varying the parameters used in the resampling algorithm, majority voting on the identity of the true model over multiple runs of SMC or adjusting the guess heuristic used to choose experiments. Moreover, these shortcomings are often heralded by effective sample size criteria built into SMC software [30], such that a more appropriate resampler or set of resampling parameters can be chosen. In spite of these theoretical shortcomings, resampling methods work exceptionally well in practice for problems in Hamiltonian learning [10–13], machine learning [31], computer vision [32], and artificial intelligence [26, 33].

2.2. Quantum Hamiltonian learning

QHL builds upon SMC by introducing weak simulation, in which the experimentalist has access to a 'black box' that produces data according to an input hypothesis \mathbf{x} . By repeatedly sampling this black box for each SMC hypothesis, the likelihood can be inferred from the frequencies of data output by the black-box simulator [34].



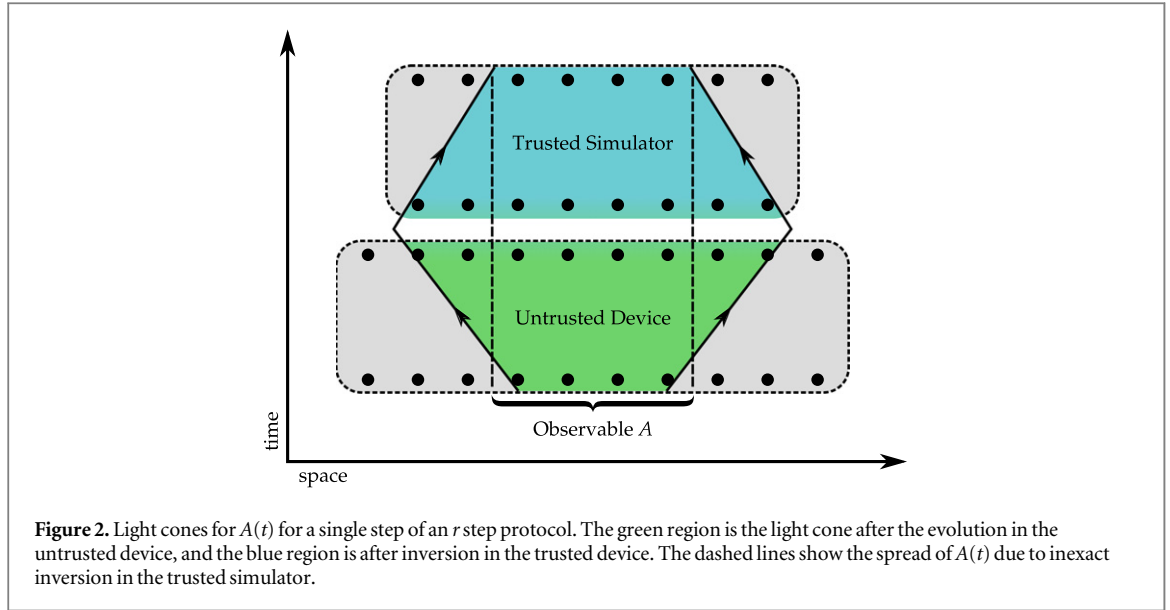
QHL is therefore a classical Bayesian inference algorithm that uses a fast quantum method for estimating the likelihood function via quantum simulation [12]. This augmented procedure is robust to errors in the likelihood function introduced by finite sampling of the black box and to approximation errors in the Hamiltonians used [13]. This latter property will be of particular importance in the development of cQHL, as it allows us to use a *truncation* of the complete system as an approximate simulator.

The simulators used in QHL can take many forms: they could be special purpose analog simulators such as an ion trap that implements a family of transverse Ising models [35]. On the other hand, the quantum simulation could be implemented by using a quantum computer to run a digital simulation algorithm that is capable of efficiently simulating a wide array of Hamiltonian models [36] (such as d -sparse row-computable Hamiltonians). We refer to all such devices as *quantum simulators* to reflect the fact that they need not be a fault-tolerant quantum computer. In our work we also require that the simulator be able to accept its initial state as input from another quantum system, but there are simpler QHL methods that do not need to be run in this fashion.

The simplest experimental design proposed for QHL is quantum likelihood evaluation (QLE), in which the experimenter prepares a state $|\psi\rangle$ on the untrusted system, evolve under the ‘true’ Hamiltonian $H(x_0)$ for some time t , and then measures $\{|\psi\rangle\langle\psi|, \mathbb{I} - |\psi\rangle\langle\psi|\}$ on the trusted simulator. This experiment is then repeated for each SMC hypothesis x_i until the variance in the estimated likelihood becomes sufficiently small. The experiment design is illustrated in figure 1. QLE can be effective for learning Hamiltonians, although it suffers from the fact that the evolution times used by the experiments must be small for most Hamiltonians. In the case of QLE, long evolution times for typical Hamiltonians (such as Gaussian random Hamiltonians [37, 38]) produce a distribution that is very close to uniform over measurement outcomes, such that experiments provide an exponentially small amount of information about the parameters.

The tendency of quantum systems to rapidly equilibrate also causes the update step in the SMC algorithm to become unstable [13]. Here by unstable we mean that small perturbations in the estimated likelihoods result in large deviations in the posterior distribution. This can be combatted by using short experiments, which lead to uncertainties about the true model (after one update) that scale at least as $O(t^{-1})$ (see appendix A). As a result, short-time evolutions necessitate processing exponentially more data than would be required if long experiments could be performed. Moreover, given that the time required for state preparation is independent of t , it is clear that the total experimental time used to learn the true model will be prohibitively large in such cases. Thus the ability to use long experiments can lead to substantial improvements for Hamiltonian learning.

To use the long evolution times requisite for expedient high-accuracy characterization the system of interest can be coupled to the simulator using SWAP gates, as shown in figure 1. This experiment design, interactive quantum likelihood evaluation (IQLE), uses the simulator to approximately *invert* the forward evolution under the unknown system, such that the measurement is approximately described by $H(x_0) - H(x_-)$ provided we choose an inversion hypothesis $H(x_-)$ that is close to $H(x_0)$. Intuitively, such experiments directly compare the dynamics of $H(x)$ and $H(x_-)$. Such experiments also reduce the norm of the effective system Hamiltonian, which typically allows the system to evolve for much longer before the quantum probability distribution becomes flat. These SWAP gates need not be perfect, as the learning protocol is known to be robust to such errors [13]. We further discuss the effects of faulty SWAP gates on bootstrapping in appendix D. In cases where a



SWAP gate cannot be implemented, such as when the trusted resource is implemented using an incompatible modality, non-interactive QHL can be used to perform the first iteration of bootstrapping, such that SWAP gates are available in all further iterations. That is, we can use QLE to initialize the bootstrapping procedure, and can proceed using IQLE.

In order to combat the exponentially diminishing likelihood of the system returning to its initial state after the inversion, we require that $\|H(x) - H(x_-)\|t$ is approximately constant [12]. We use the particle guess heuristic (PGH) to achieve this. The PGH involves drawing two hypotheses about H , x_- and x'_- , from the prior distribution and then choosing $t = 1/\|H(x_-) - H(x'_-)\|$. Since $\|H(x_-) - H(x'_-)\|$ is an estimate of the uncertainty in the Hamiltonian, we expect that at most a constant fraction of the prior distribution will satisfy $\|H(x_-) - H(x)\|t \geq 1$ assuming $H(x)$ is linear and the prior distribution has converged to a unimodal distribution centered about the true Hamiltonian. The heuristic therefore seldom leads to experiments for which $|\langle \psi | e^{iH(x_-)t} e^{-iH(x)t} | \psi \rangle|^2 \sim 1/2^n$ for most x . Moreover, since the PGH relies only on the current SMC approximation to the posterior, the heuristic incurs no additional simulation costs. Rather, the PGH provides adaptivity by depending on the current state of knowledge about the state of the quantum system through the particles x_- and x'_- . Experiment design via the particle guess heuristic has been shown to lead to efficient estimation of Hamiltonians using IQLE [12], and has since been usefully applied in other experimental contexts [11].

Previous work has analyzed the complexity of learning using IQLE [12]. In cases where the error in the characterized Hamiltonian scales as $e^{-\gamma N_{\text{exp}}}$ and N_{samp} samples are used to estimate the likelihood function, the protocol requires $O(N_{\text{samp}} N_{\text{particles}} \log(1/\epsilon)/\gamma)$ simulations to learn the vector of Hamiltonian parameters to within error ϵ as measured by the 2-norm. In practice, the decay constant γ depends on the number of parameters used to describe H and the properties of the experiments used. It does not directly depend on the dimension of H [12, 13]. The updating procedure used to combine these results is further known to be stable provided that the likelihoods of the observed experimental outcomes are not exponentially small for the majority of the SMC particles [12]. This occurs for well posed learning problems that use two outcome experiments.

3. Compressed QHL

Information locality is what enables cQHL and in turn quantum bootstrapping. This idea is made concrete via Lieb–Robinson bounds, which show that an analog of special relativity exists for local observables evolving under Hamiltonians that have rapidly decaying interactions [9, 39–41]. Lieb–Robinson bounds give an effective ‘light cone’, as illustrated in figure 2, in which the evolution of an observable A can be accurately simulated without needing to consider any subsystem outside of the light cone. Specifically, they imply that a local observable $A(t)$ provides at most an exponentially small amount of information about subsystems that are further than distance st away from the support of $A(0) \equiv A$, where s is the Lieb–Robinson velocity for the system and t is the evolution time. Here, s is analogous to the speed of light, and only depends on the geometry and strengths of the interactions in the system [39–41]. Thus, if st is bounded above by a constant and the initial

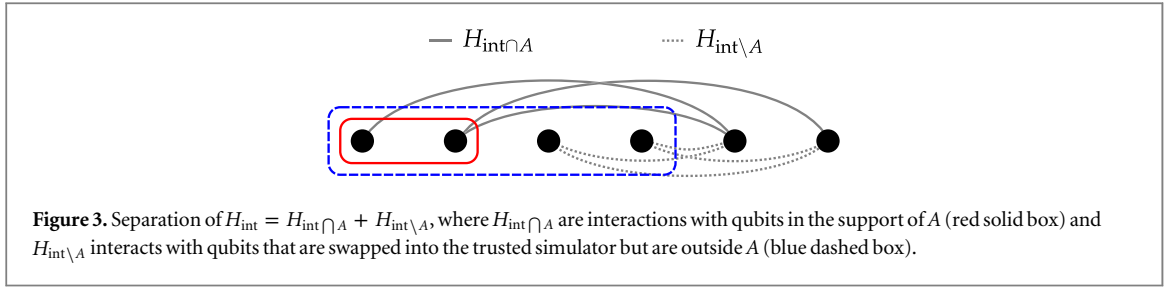


Figure 3. Separation of $H_{\text{int}} = H_{\text{int} \cap A} + H_{\text{int} \setminus A}$, where $H_{\text{int} \cap A}$ are interactions with qubits in the support of A (red solid box) and $H_{\text{int} \setminus A}$ interacts with qubits that are swapped into the trusted simulator but are outside A (blue dashed box).

support of A is small then the support of $A(t)$ is at most a constant. This shows that the dynamics of $A(t)$ can be simulated using a small quantum device, provided st is sufficiently small.

cQHL exploits this intuition by evolving an initial state under the untrusted quantum simulator, swapping the quantum state of a subsystem of the larger (uncharacterized) system into a quantum simulator, and then approximately inverting the evolution by guessing the Hamiltonian dynamics and simulating their inverse. It then measures the simulator to determine whether the inversion yielded the initial state or a state in its orthogonal complement. One step of this process is illustrated in figure 2.

The inversion process used in interactive QLE not only leads to more informative experiments, but we will show in section 3.2 that generalizing to include repeated applications of swaps and inverse simulations also delay the rate at which the light cone propagates from the observable. This in turn allows much longer evolution times to be used without the observable stretching beyond the confines of the trusted simulator.

In particular, this swapping procedure leads to characteristic Lieb–Robinson velocities that shrink as the experimentalist learns more about the system. *That is, the light cone represents an ‘epistemic’ speed of light in the coupled systems that arises from the speed of information propagation depending more strongly on the uncertainty in the Hamiltonian than the Hamiltonian itself.* Since the effective speed of light slows as more information is learned, long evolutions can be used when the uncertainty is small. This removes a major restriction of the method of Da Silva *et al* [9] since the variance of any unbiased estimator of the Hamiltonian parameters diverges as $\Omega(t^{-2})$ for Hamiltonian learning methods that use short time experiments (see appendix A for more details).

To analyse the error introduced by compressed simulation, we consider learning a Hamiltonian H by measurement of an observable A supported on a sites, using a simulator with support on $w > a$ sites. We then expand H into those terms H_{in} which we can access with our simulator, the terms H_{out} supported only outside of the simulator entirely, and the interaction H_{int} between these two partitions. We then further break down H_{int} into those terms $H_{\text{int} \cap A}$ which have non-trivial action between sites in A and the neglected sites, and those terms $H_{\text{int} \setminus A}$ which act upon sites in the simulator, but not the observable. That is, we expand H as

$$H = H_{\text{int}} + H_{\text{in}} + H_{\text{out}} = H_{\text{int} \cap A} + H_{\text{int} \setminus A} + H_{\text{in}} + H_{\text{out}}. \quad (5)$$

The decomposition of the interaction Hamiltonian H_{int} into couplings that include and exclude A is illustrated in figure 3.

cQHL neglects the terms included in $H_{\text{int} \cap A}$ when processing data collected from the system of interest. If H exhibits a finite Lieb–Robinson velocity, then we can bound the error introduced by this approximation. Moreover, we will show that by using interactivity as a resource, we can reduce this error as we become more certain about the dynamics of the untrusted system. We discuss these two considerations in more detail below.

3.1. Commuting Hamiltonians

It is helpful, however, to first build intuition by considering the special case that all terms in the unknown system’s Hamiltonian are local and mutually commute. This is true, for instance, in the Ising models (22) that we consider in numeric examples. In this case, the compressed interactive likelihood evaluation experiment described in figure 1 is particularly simple to analyze.

If we work in the Heisenberg picture then it is easy to see from the assumption that the Hamiltonian terms commute with each other (but not necessarily A) that $[H_{\text{int} \setminus A} + H_{\text{out}}, A(t)] = 0$. This implies that

$$\begin{aligned} A(t) &= e^{iH_{\text{in}}t} e^{iH_{\text{int} \cap A}t} A e^{-iH_{\text{in}}t} e^{-iH_{\text{int} \cap A}t} \\ \tilde{A}(t) &= e^{iH_{\text{in}}t} A e^{-iH_{\text{in}}t}, \end{aligned} \quad (6)$$

where $\tilde{A}(t)$ is the simulated observable within the trusted simulator.

Using Hadamard’s Lemma and the triangle inequality to bound the truncation error $\|\tilde{A}(t) - A(t)\|$, we obtain that

$$\|\bar{A}(t) - A(t)\| \leq \|A\| \left(e^{2\|H_{\text{int}} \cap A\|t} - 1 \right). \quad (7)$$

If the objective is to have error at most δ in the compressed simulation then it suffices to choose experiments with evolution time at most

$$t \leq \ln \left(\frac{\delta}{\|A\|} + 1 \right) \left(2\|H_{\text{int}} \cap A\| \right)^{-1}. \quad (8)$$

If the sum of the magnitudes of the interaction terms that are neglected in the simulation is a constant then (8) shows that t scales at most linearly in δ as $\delta \rightarrow 0$. This is potentially problematic because short experiments can provide much less information than longer experiments so it may be desirable to increase the size of the trusted simulator as δ shrinks to reduce the experimental time needed to bootstrap the system. QHL is robust to δ [12, 13] and $\delta \approx 0.01$ often suffices for the inference procedure to proceed without noticeable degradation.

Note that if $H_{\text{int}} \cap A = 0$ then infinite-time simulations are possible for commuting models (such as Ising models) because no truncation error is incurred. Non-trivial cases for QHL only occur in commuting models with long range interactions.

3.2. Non-commuting Hamiltonians

If the Hamiltonian contains non-commuting terms then the factorization of e^{-iHt} used in (7) no longer holds. This is because

$$e^{i(H_{\text{in}}+H_{\text{int}}+H_{\text{out}})} A e^{-i(H_{\text{in}}+H_{\text{int}}+H_{\text{out}})} \neq e^{i(H_{\text{in}}+H_{\text{int}})} A e^{-i(H_{\text{in}}+H_{\text{int}})}, \quad (9)$$

unlike in commuting models. Such dynamics can also lead to observables $A(t)$ that rapidly obtain non-negligible support near the boundary of the trusted simulator. The trusted system will not tend to simulate these evolutions accurately because significant interactions exist between $A(t)$ and the neglected portion of the system. This means a more careful argument will be needed to show that bootstrapping will also be successful here.

We address this issue by generalizing IQLE experiments. Typically, each IQLE experiment is of the form $e^{iH_-t} S e^{-iH_+t}$, where S is a swap operation and H_- is a Hamiltonian simulated by the trusted simulator. Instead of swapping the states of both devices once, we generalize such experiments to consist of r swaps, such that the observable evolves under the unitary $(e^{iH_-t/r} S e^{-iH_+t/r})^r$ before we perform a measurement. It is then easy to see from the Trotter formula that $e^{iH_-t/r} e^{-iH_+t/r} \approx e^{-i(H_-H_+)t/r}$. The inclusion of additional swap operations serves two purposes. Firstly, it causes the terms in the Hamiltonian to effectively commute with each other for r sufficiently large. Secondly, if t is small relative to $\|H_- - H_+\|^{-1}$ then r swaps of the two registers will not cause the $A(t)$ to have substantial support on the boundary of the trusted simulator at any step in the protocol.

If a large value of r is chosen, then the system effectively evolves under $e^{-i(H_-H_+)t} = e^{-i(H_{\text{out}}+H_{\text{int}}+\Lambda)t}$, where $\Lambda := H_{\text{in}} - H_-$. We expect that the dynamics of A will therefore be dictated by the properties of Λ for short evolutions. We make this intuition precise by showing in appendix B that the error from using a small trusted simulator obeys

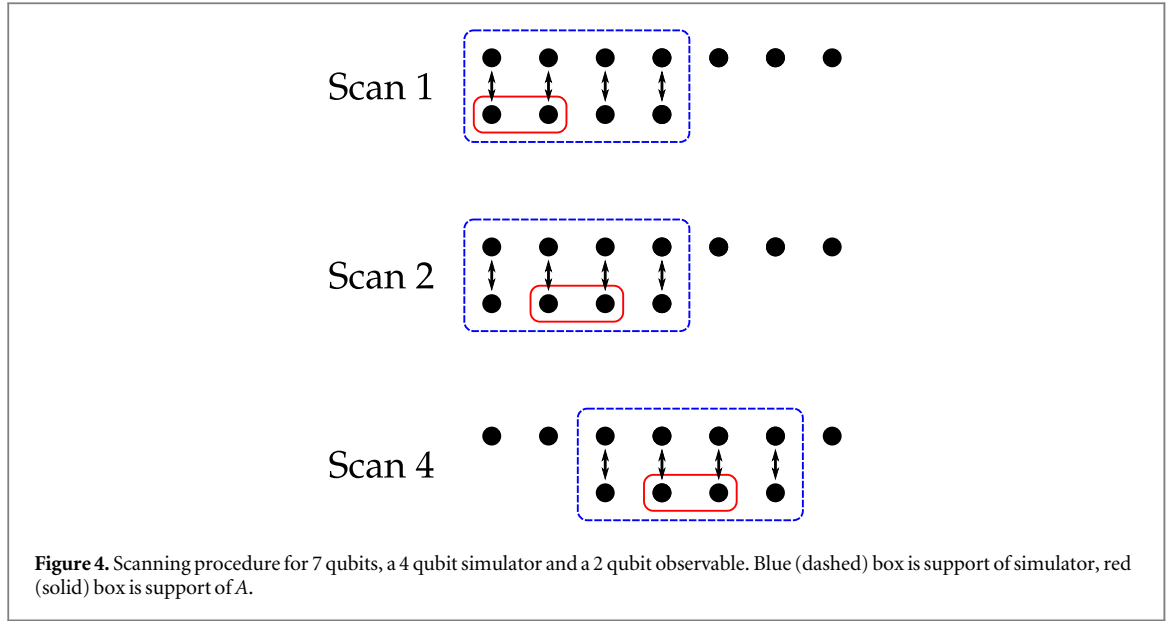
$$\begin{aligned} \|A(t) - \bar{A}(t)\| &\leq \left(\| [H_{\text{in}}, \Lambda] \| + \| [H_{\text{int}}, H_{\text{in}}] \| \right) \|A\| \frac{t^2}{r} + 2 \|H_{\text{int}} \cap A\| \|A\| t \\ &\quad + 2 \|H_{\text{int}} \setminus A\| \|A\| |\{A\}| t e^{-\mu \text{dist}(A, H_{\text{out}})} \left[e^{2s|t|} - 1 \right] e^{2\|H_{\text{out}}+H_{\text{int}} \setminus A\|t/r}, \end{aligned} \quad (10)$$

for cases of nearest-neighbor or exponentially decaying interactions between subsystems. $H_{\text{in}}, H_{\text{int}}, H_{\text{out}}$ and related terms are explained in (5) and the surrounding text. Here s is the Lieb–Robinson velocity for evolutions under Λ and μ is related to the rate at which interactions decay with the graph distance between subsystems. It is worth noting that (10) can be improved by using higher order Trotter–Suzuki formulas in place of the basic Trotter formula to reduce r [42] and also by using tighter Lieb–Robinson bounds for cases with nearest-neighbor Hamiltonians.

The variable Λ is related through the PGH to the uncertainty in the Hamiltonian, which implies that the speed of information propagation is also a function of the uncertainty in H [39, 40]. That is, longer evolutions can be taken as H becomes known with ever greater certainty. This means that the Lieb–Robinson velocity does not pose a fundamental restriction on the evolution times permitted because $s \rightarrow 0$ as $\Lambda \rightarrow 0$.

Of course, the error term $2\|H_{\text{int}} \cap A\|t$ in (10) places a limitation on the evolution time but that term can be suppressed exponentially by increasing the diameter of the set of qubits in the trusted simulator for systems with interactions that decay at least exponentially with distance. Thus the roadblocks facing cQHL can be addressed at modest cost by using our strategy of repeatedly swapping the subsystems in the trusted and untrusted devices.

As an example, if we assume (a) that the interactions are between qubits on a line (b) that $w - a$ is chosen such that $8st/\mu < w - a$ then in the limit as $r \rightarrow \infty$



$$t \leq \frac{\delta}{2\|A\| \left(\|H_{\text{int} \cap A}\| + 2\|H_{\text{int} \setminus A}\| |\{A\}| e^{-\mu(w-a)/4} \right)}, \quad (11)$$

suffices to guarantee simulation error of δ . This result is qualitatively similar to (8) and requires that $|w - a|$ scales at most logarithmically with the total evolution time desired.

3.3. Scanning

The previous methods provide a method for characterizing a subsystem of the global Hamiltonian. These results cannot be used directly to learn the full system Hamiltonian because the trusted simulator lacks couplings present in the full system. Instead the Hamiltonian must be inferred by patching together the results of many Hamiltonian inference steps. This process can be thought of as a scanning procedure wherein an observable is moved across the set of qubits collecting information about the couplings that strongly influence it. The scanning procedure is illustrated in figure 4.

In order to properly update the information about the system, we modify sequential Monte Carlo to use two particle clouds. The first is the global cloud, which keeps track of the prior distribution over all the parameters in the Hamiltonian model. The second is the local cloud, which keeps track of all of the parameters needed for the current cQHL experiment. The global cloud is constrained such that the weights of each particle in the cloud is constant (i.e. the probability density is represented by the density of particles rather than their weight), whereas the local cloud is not constrained in this fashion. This constraint on the global cloud is needed because resampling does not in general preserve the indices of each particle, so that there is no way to sensibly identify a global particle that corresponds to a particle in the local posterior.

Instead, by copying a subset of global parameters into the local cloud, our modified particle filter approximates the prior by a product distribution between the local and remaining parameters. Resampling the local posterior then makes this approximation again, ensuring that the local weights are uniform. Thus, we can copy the (newly resampled) local cloud into the global cloud, overwriting the corresponding parameters. Once the local cloud is merged back into the global cloud in this way, we begin the next step in the scan by selecting a different set of parameters for the local cloud, and continuing with the next cQHL experiment.

We implement this scanning procedure in our numerical experiments by using a local observable centered as far left on the spin chain as possible. We then infer the Hamiltonian for this location using a fixed number of experiments, swap the Hamiltonian parameters from each of the SMC particles to the global cloud and then move the observable one site to the right. This process is repeated until the observable has scanned over the entire chain of qubits, and then we begin again by scanning over the first $2a$ qubits in reverse, where a is the width of the observable. We do this to reduce the systematic bias that emerges from the fact that Hamiltonian parameters associated with couplings learned earlier in the procedure will have greater uncertainty.

3.4. Complexity of cQHL

Here investigate the circumstances under which cQHL is efficient. The number of experiments that are needed in cQHL is $N_{\text{exp}} = N_{\text{scans}} N_{\text{exp/scan}}$. If we then make the pessimistic assumption that $O(n)$ scans are needed and

assume that the error within each scan decays as $e^{-\gamma N_{\text{exp/scan}}}$ then the number of experiments needed to make the *combined error* in the inferred vector of Hamiltonian parameters, \mathbf{x} , at most ϵ it suffices to use a number of experiments that scales as

$$N_{\text{exp}} \in O\left(\frac{n \log(n/\epsilon)}{\gamma}\right). \quad (12)$$

Here we have used the fact that if there are n scans and the error in the Hamiltonian parameters is ϵ/n then the error in the reconstructed Hamiltonian parameters after n scans is at most $n \times \epsilon/n = \epsilon$. This bound is again pessimistic as in practice the information learned in subsequent scans actually reduces, rather than increases, the error in parameters learned from previous scans.

The number of calls to our trusted simulator needed to update the weights of the particles in the SMC cloud is $N_{\text{sim}} = N_{\text{exp}} N_{\text{samp}} N_{\text{part}}$. Firstly, if $N_{\text{samp}} \in O(1/\delta^2)$ samples are used to infer the likelihoods of the outcome for each of the N_{part} particles in the SMC cloud, then the total number of experiments required to learn a fixed Hamiltonian on n qubits the number of simulations that are needed to estimate the likelihoods within error δ is

$$N_{\text{sim}} \in O\left(N_{\text{part}} \times \frac{n \log(n/\epsilon)}{\gamma \delta^2}\right). \quad (13)$$

This quantifies the number of experiments that are needed to estimate the likelihoods within error $O(\delta)$ given a perfect simulator.

In cases with non-commuting Hamiltonians, error is also incurred by using a non-infinite value of r . If we also demand that the contribution from this source of error is $O(\delta)$ then it is straight forward to see from (10) that the necessary value of r scales at most as

$$r \in O\left(\frac{\left(\| [H_{\text{in}}, \Lambda] \| + \| [H_{\text{int}}, H_{\text{in}}] \| \right) \| A \| t^2}{\delta} + \left(\| H_{\text{out}} \| + \| H_{\text{int} \setminus A} \| \right) t\right). \quad (14)$$

The above relations set the complexity (as measured by the number of experiments and the number of swaps) of Hamiltonian characterization, but the space requirements also are problem dependent in cases that have non-commuting Hamiltonians. If we assume that all interactions between arbitrary qubits x and y decay at least as $e^{-\nu \text{dist}(x,y)}$, then the distance between A and the neglected part of the Hamiltonian $H_{\text{int} \setminus A}$ can be chosen as

$$\text{dist}(A, H_{\text{out}}) \in O\left(\frac{1}{\min\{\mu, \nu\}} \left(st + \log\left(\frac{\| H_{\text{int} \setminus A} \| \| A \| |\{A\}| t}{\delta}\right) \right)\right), \quad (15)$$

where $|\{A\}|$ is the number of sites on which A is supported, s is the Lieb–Robinson velocity for evolution under Λ , and where μ is the exponential clustering parameter used in (10).

From the PGH, we have that st is asymptotically constant and thus both the number of qubits and the number of experiments scale logarithmically with the desired accuracy. The number of inversions used r scales polynomially with t , which scales as $O(1/\epsilon)$ [12] and thus the number of swaps scales polynomially with the desired error tolerance. This scaling can be made to approach the Heisenberg limit of $O(1/\epsilon)$ scaling by replacing the Trotter formula used in the inversion step with increasingly high-order Trotter–Suzuki formulas as ϵ shrinks [42].

The above analysis can also be extended to systems with polynomial decay of interactions. In such cases, the cQHL algorithm is less efficient because the logarithmic scaling with $1/\epsilon$ is replaced with polynomial scaling with $1/\epsilon$ in most instances of polynomially-decaying interactions. Even in such cases, however, we expect the sequential Monte Carlo algorithm to remain robust to simulation errors such as truncation. Thus, we expect that cQHL can offer advantages when the Lieb–Robinson velocity of an untrusted system under inversion by a hypothesis is finite.

In some cases, such as dipolar coupling, existing Lieb–Robinson bounds diverge [39]. This does not imply, however, the lack of a finite information propagation velocity however; indeed, it is worth noting that finite speeds of information propagation are expected theoretically and observed experimentally for finite systems with dipolar couplings [43]. Moreover, in the presence of disorder, experimental evidence suggests that the information propagation velocity can be dramatically reduced [44].

An important remaining issue is the scaling of δ and the number of particles. We know from prior work that Bayesian inference is highly tolerant of errors in the likelihood function [13] and that $\delta \approx 0.01$ typically suffices. Furthermore, The number of particles in the SMC cloud, N_{part} , is a slowly increasing function of the number of model parameters for the Hamiltonian and does not explicitly depend on the Hilbert space dimension [45]. In the numerical experiments we have performed in this, and prior work, we observe that N_{part} tends to scale

roughly logarithmically with the number of model parameters [10, 12, 13]. This means that none of the above issues present a fundamental obstacle for cQHL.

4. Quantum bootstrapping

We will now turn our attention to quantum bootstrapping, which is an application wherein cQHL is used to infer control maps for uncharacterized devices. Control maps relate control settings of a device to its system Hamiltonian. Learning these maps is of particular importance if cross-talk or defects cause different parts of the system to respond differently to the same controls. In such cases, Hamiltonian characterization is a necessary part of the control design and calibration process.

To concretely show how QHL can address this challenge, we consider a model in which a row-vector of control settings C is related to the system Hamiltonian by an affine map $H(C)$,

$$H(C) = C \cdot [H_1, \dots, H_M] + H_0 \quad (16)$$

for some $M + 1$ unknown Hamiltonians $\{H_0, \dots, H_M\}$. By the same argument as before, let $H_j = H(x_j)$ be represented by a model parameter vector, such that this is an efficient representation of the control landscape.

The control learning process then proceeds as follows:

- (a) Set $C = 0$ and learn H_0 using cQHL.
- (b) For $k = 1, \dots, M$ set $C_j = \delta_{j,k}$ and learn $H_k + H_0$ using compressed QHL.
- (c) Subtract H_0 from these values to learn the vector v_k that describes the model for H_k .

This yields a vector of Hamiltonian parameters that describes each control term H_k . If we then imagine the matrix G such that $G_{k,j} = [v_k]_j$ then a model for $H(C)$ is given by GC^T , which allows the effect of control on the quantum system of interest to be predicted.

Nonlinear controls can be learned in a similar fashion by locally approximating the control function with a piecewise-linear function.

We complete our description of quantum bootstrapping by detailing how control learning can be used to calibrate an initially untrusted device. If $H(C)$ is an affine map then this can be accomplished using the following approach:

- (a) Learn $H(C)$ using the above method.
- (b) Choose a set of fundamental Hamiltonian terms, \mathcal{H}_j , from which all Hamiltonians in the class of interest can then be generated.
- (c) For each \mathcal{H}_j apply the Moore–Penrose pseudoinverse of G to $\mathcal{H}_j - H_0$ to find C_j such $H(C_j) \approx \mathcal{H}_j$.
- (d) Treat the system as a trusted simulator and repeat steps (a), (b) and (c) for a larger system.

In cases where $H_0 = 0$, $H(C)$ is linear and hence $H(aC_1 + bC_2) = a\mathcal{H}_1 + b\mathcal{H}_2$. This means that an arbitrary Hamiltonian formed from a linear combination of the \mathcal{H}_j can be implemented. If $H_0 \neq 0$, then this process is less straightforward. It can be solved by applying a pseudoinverse to find a control vector $C(a, b)$ that produces $a\mathcal{H}_1 + b\mathcal{H}_2$, but such controls will be specific to a and b . A simple way to construct a general control sequence is to use Trotter–Suzuki formulas to approximate the dynamics in terms of $\mathcal{H}_1 = H(C_1)$ and $\mathcal{H}_2 = H(C_2)$ as

$$\left(e^{-i\mathcal{H}_1 a \Delta t / R} e^{-i\mathcal{H}_2 b \Delta t / R} \right)^R = e^{-i(a\mathcal{H}_1 + b\mathcal{H}_2) \Delta t} + O\left(\frac{\Delta t^2}{R}\right). \quad (17)$$

Higher-order Trotter–Suzuki methods can be used to reduce the value of R if desired [42].

Errors accrue as the bootstrapping procedure progresses. However, since the error shrinks exponentially with the number of experiments for well-posed learning problems, the number of experiments needed per recursion will often scale linearly with the total number of intermediate untrusted devices needed to reach the final system of interest (see appendix D).

4.1. Error propagation in bootstrapping

Let G be the control map that is inferred via the inversion method discussed above and let $G + \mathcal{E}$ be the actual control map that the system performs. If we measure the error in a single step of bootstrapping to be the operator

norm of the difference between the bootstrapped and the target Hamiltonians then we have that the control error for the system after bootstrapping obeys

$$\|(G + \mathcal{E})G^+\mathcal{H}_k - \mathcal{H}_k\| \leq (\|GG^+ - \mathbb{1}\| + \|\mathcal{E}\|\|G^+\|)\|\mathcal{H}_k\|, \quad (18)$$

where G^+ is the pseudo-inverse of G . Equation (18) shows that the error after a single step is a multiple of the norm of the control Hamiltonian that depends not only on the error in the cQHL algorithm but also on $\|GG^+ - \mathbb{1}\|$ which measures the invertibility of the control map. Since the error is a multiplicative factor, it should not come as a surprise that the error after L bootstrapping steps grows at worst exponentially with L . In particular, the bootstrapping error is at most

$$L\Gamma_{\max}e^{(L-1)(\kappa_{\max}-1+\|\mathcal{E}_{\max}\|\|G_{\max}^+\|)}\max_k\|\mathcal{H}_k\|, \quad (19)$$

where Γ_{\max} is the maximum value of $(\|GG^+ - \mathbb{1}\| + \|\mathcal{E}\|\|G^+\|)$ over all the L bootstrapping steps, κ_{\max} is the maximum condition number for G , $\|\mathcal{E}_{\max}\|$ and $\|G_{\max}^+\|$ are the maximum values for the error operator and the pseudoinverse of G over all L steps. The proof of (19) is a straight forward application of the triangle inequality and is provided in appendix D.

Given that the error tolerance in the bootstrapping procedure is $\Delta \leq 1$, G is invertible and that w , a and t are chosen such that $\|\mathcal{E}_{\max}\| \leq e^{-\gamma N_{\text{exp}}}$ (i.e. a constant fraction of a bit is learned per experiment) it is easy to see that (19) is less than Δ if

$$N_{\text{exp}} \geq \frac{(L-1)\kappa_{\max} + \ln\left(\frac{L\|G_{\max}^+\|\max_k(\|\mathcal{H}_k\|, 1)}{\Delta}\right)}{\gamma}. \quad (20)$$

This process is efficient provided γ is at most polynomially small. If G is not invertible, then the error cannot generally be made less than Δ for all $\Delta > 0$. Although, if G is not invertible then the system is not fully controllable and so the task of calibrating a simulator will seldom be possible in cases where $\|GG^+ - \mathbb{1}\| \approx 0$ irrespective of the method used to control it.

It is difficult to say in general when the conditions underlying (20) will be met, as it is always possible for experiments to be chosen that provide virtually no information about the system. For example, the observable could be chosen to commute with the dynamics so that no information can be learned from the measurement statistics. Great experimental care must be taken in order to ensure that such pathological cases do not emerge [13]. These pathological experiments can be avoided for Ising models with exponential decaying interactions and we expect exponential decay of $\|\mathcal{E}_{\max}\|$ to be common for a wide range of models that also include noise and non-commuting terms based on previous studies [10, 12, 13].

The bootstrapped simulator also need not have as many controls as the simulator that is used to certify it. This does not necessarily mean that the controls in the bootstrapped device are less rich than that of the trusted simulator. If we assume, for example, that a general Ising simulator is used to bootstrap an Ising simulator with only nearest neighbor couplings (and universal single qubit control) then more general couplings terms can be simulated using two body interactions. For example, next-nearest neighbor interactions can be simulated using nearest neighbor couplings and single qubit control via:

$$e^{-2iZ \otimes \mathbb{1} \otimes Z \Delta t^2} |\phi\rangle + O(\Delta t^3) = e^{-iZ \otimes X \otimes \mathbb{1} \Delta t} e^{-i\mathbb{1} \otimes Y \otimes Z \Delta t} e^{iZ \otimes X \otimes \mathbb{1} \Delta t} e^{i\mathbb{1} \otimes Y \otimes Z \Delta t} |\phi\rangle, \quad (21)$$

where the middle qubit in $|\phi\rangle$ is set to $|0\rangle$. Higher-order and parallel methods also exist for engineering such interactions [46, 47].

5. Numerical results

In order to show that the cQHL and bootstrapping algorithms are scalable to large systems, we provide numerical evidence that 50 qubit Ising Hamiltonians with exponentially decaying interactions can be learned using an 8 qubit simulator. We further observe that only a few kilobits of experimental data are needed to infer an accurate model and that the observable, A , that is used for the inference only needs to be supported on a small number of qubits. Finally, we apply the cQHL algorithm to use the 8 qubit simulator to bootstrap a 50 qubit quantum simulator from an initially uncalibrated device with crosstalk on the controls. The bootstrapping procedure reduces the calibration errors in a 50 qubit simulator by two orders of magnitude using roughly 750 kilobits of experimental data. This calibrated 50 qubit simulator could then be used to bootstrap an even larger quantum device.

We perform numerical simulations using the open-source QInfer, SciPy and `fhf` libraries [30, 48, 49]. All numerical results simulate using IQLE (figure 1) with compressed simulation on an 8 qubit register.

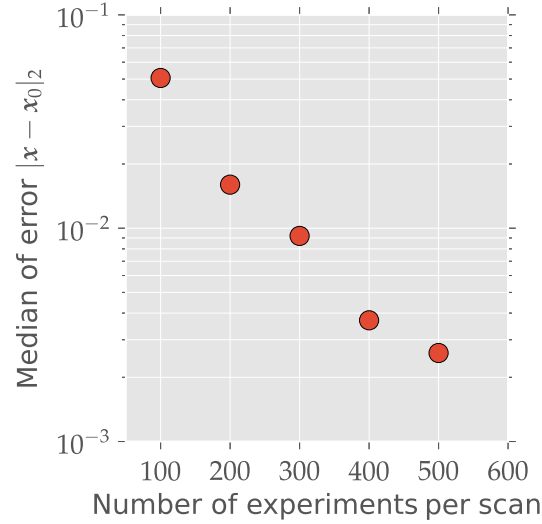


Figure 5. Error in QHL for $a = 4$ with varying N_{exp} per scan. Data consistent with $e^{-0.006N_{\text{exp}}}$ scaling.

Table 1. $|x - x_{\text{true}}|_2$ for QHL using different numbers of qubits for the observable at 500 experiments per scan.

a	75th percentile	Median error	25th percentile
6	0.0043	0.0029	0.0014
4	0.0029	0.0018	0.0014
2	0.0252	0.0234	0.0225

5.1. Compressed quantum Hamiltonian learning

Since quantum devices capable of implementing our bootstrapping protocol are not currently available, we examine systems that can be simulated efficiently using classical computers in order to demonstrate that our algorithm applies to large systems. Thus, we focus on the example of an Ising model on a linear chain of qubits, with exponentially decaying interactions

$$H(x) = \sum_{i \neq j} x_{i,j} \sigma_z^{(i)} \sigma_z^{(j)}, \quad (22)$$

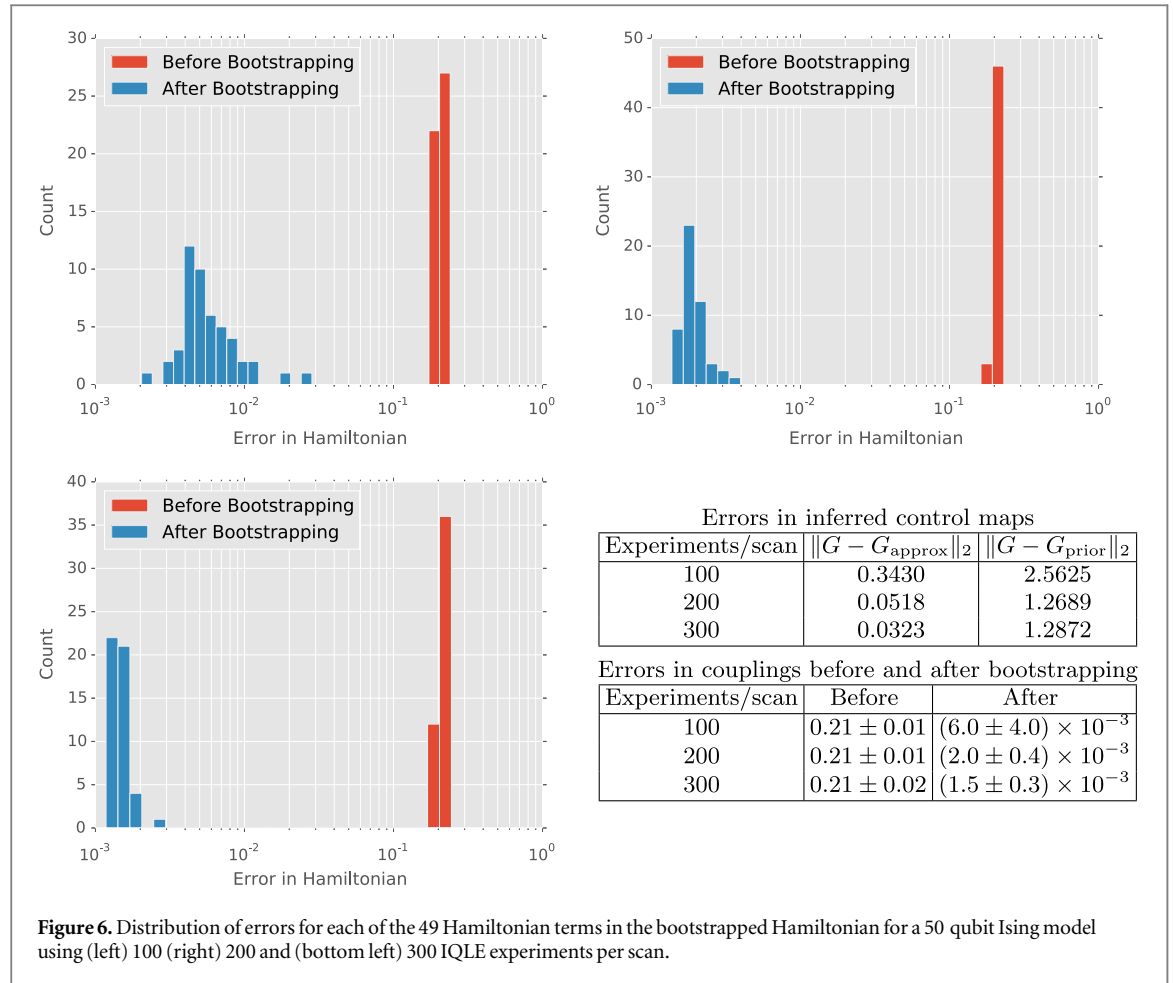
where the parameters $x_{i,j}$ are distributed according to the uniform distribution $x_{i,j} \sim \text{unif}(0, 1) 10^{-2(|i-j|-1)}$. In all cases, the observable used is $A = (|+\rangle\langle+|)^{\otimes a}$ for $a = \{2, 4, 6\}$, as this observable is maximally informative for Ising models. For more general models, a pseudorandom input state and observable can be used instead [13].

Figure 5 shows that a compressed quantum simulator using only 8 quantum bits is capable of learning a Hamiltonian model for a system with 50 qubits. The errors, as measured by the operator norm of difference between the actual Hamiltonian and the inferred Hamiltonian, are typically on the order of 10^{-2} after as few as 300 experiments per scan where 49 scans are used in total. This is especially impressive after noting that this constitutes roughly 750 kilobits of data and that this error of 10^{-2} is spread over 1225 terms. The data also shows evidence of exponential decay of the error, which is expected from prior studies [12, 13].

An important difference between this result and existing QHL schemes [12, 13] is that the observable will need to be, in some cases, substantially smaller than the simulator. Choosing a small observable is potentially problematic because it becomes more likely that an erroneous outcome will be indistinguishable from the initial state. Also, if a is too small then important long-range couplings can be overlooked because their effect becomes hard to distinguish from local interactions. We find in table 1 that the cases where $a = 4$ and $a = 6$ are virtually indistinguishable whereas the median errors are substantially larger for $a = 2$, but not substantially worse than $a = 4$ for 200 experiments per scan. This provides evidence that small a can suffice for Hamiltonian learning.

5.2. Quantum bootstrapping

The next set of results show that cQHL can be used to bootstrap a quantum simulator for a 50 qubit 1D Ising model. The bootstrapping problem that we consider can be thought of as correcting crosstalk in the large simulator. This crosstalk manifests itself when the experimentalist attempts to turn on only one of the Ising



couplings in the simulator but in fact all 1225 interactions are actually activated to some extent. We further assume that the 50 qubit simulator only has 49 controls corresponding to each of the nearest neighbor interactions. This means that a perfect control sequence will generally not exist because $49 < 1225$. The control Hamiltonians $[H_1, \dots, H_M]$ in (16) conform to (22) with $x_{i,j} \sim 10\delta_{p,i}\delta_{p,j-1} + \text{unif}(0, 1)10^{-2(|i-j|-1)}$. We also take $H_0 = 0$.

Figure 6 reveals that our bootstrapping procedure reduces control errors by two orders of magnitude in cases where 300 experiments per scan are used in the QHL step. Further reductions could be achieved by increasing the number of experiments per scan, but at 300 scans much of the error arises from $\|GG^+ - \mathbb{1}\| \neq 0$ so a richer set of controls in the 50 qubit system would be needed to substantially reduce the residual control errors. The errors are sufficiently small, however, that it is reasonable that the device could be used as a trusted simulator for nearest-neighbor Ising models. This means that it could be subsequently used to bootstrap another quantum simulator.

5.3. Scaling with n

All of the examples considered so far examine cQHL for 50 qubits. Although the fact that the protocol scales successfully up to 50 qubits already provides strong evidence for its scalability, we provide further evidence here that the errors in cQHL do not rapidly vary as a function of the number of qubits in the untrusted system, n . As per the previous numerical examples, we consider a 1D Ising model with $x_{i,j} \sim \text{unif}(0, 1)10^{-2(|i-j|-1)}$ and use a 4 qubit observable. Also 20 000 particles are used in the SMC approximation to the posterior and we take all data using 200 experiments per scan. Roughly 20 data points per value of n were considered.

We see in figure 7 that the error, as measured by the median L_2 distance between the inferred model and the true model, is a slowly increasing function of n . The data is consistent with a linear scaling in n , although the data does not preclude other scalings. This suggests that the error in cQHL does not rapidly increase for the class of Hamiltonians considered here and provides evidence that examples with far more than 50 qubits are not outside the realm of possibility for cQHL.

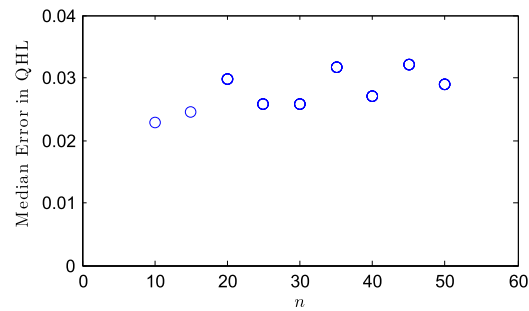


Figure 7. Median error in compressed QHL as a function of the number of qubits in the model.

6. Conclusions

We show that small quantum simulators can be used to characterize and calibrate larger devices, thus providing a way to bootstrap to capabilities beyond what can be implemented classically. In particular, we provide a cQHL algorithm that can infer Hamiltonians for systems with local or rapidly decaying interactions. The compressed algorithm is feasible because of the fact that local observables remain confined to light cones. Typically these light cones spread at a velocity that is dictated by the Hamiltonian. By contrast, in cQHL, light cones spread at a speed that depends on the uncertainty in the Hamiltonian in cQHL. This not only allows more informative experiments to be chosen but also shows that an epistemic speed of light can exist in systems that interact with an intelligent agent.

We then show that this algorithm provides the tools necessary to bootstrap a quantum system; wherein a small simulator to learn controls that correct Hamiltonian errors and uncertainties present in a larger quantum device. This protocol is useful, for instance, in calibrating control designs to deal with cross-talk, uncertainties in coupling strengths and other effects that cause the controls to act differently on the quantum system than the designed behavior.

Our approaches, being based on QHL, inherit the same noise and sample error robustness observed in that algorithm [12, 13]. We have provided numerical evidence that our techniques apply to systems with as many as 50 qubits, can further tolerate low precision observables, and are surprisingly efficient. Thus, quantum bootstrapping provides a potentially scalable technique for application in even large quantum devices, and in experimentally-reasonable contexts. Our work therefore provides a critical resource for building practical quantum information processing devices and computationally useful quantum simulators.

There are several natural extensions to our work. While we have focused on the case of time-independent quantum controls and Hamiltonians, our approaches can be generalized to the time dependent case using more general Lieb–Robinson bounds [50]. This is significant because techniques such as the method of Da Silva *et al* [9] do not apply for $H(t)$. Additionally, it would be very interesting to see if quantum simulation can be used to allow local optimization to design even better experiments than those yielded by the PGH. The introduction of cost effective local optimization strategies may lead to significant advantages for bootstrapping systems with non-commuting Hamiltonians.

As a final remark, our work provides an important step towards a practical general method for calibrating and controlling large quantum devices, by utilizing epistemic light cones to compress the simulation, thus enabling the application of small quantum devices as a resource. In doing so, our approach also provides a *platform* for building tractable solutions to more complicated design problems through the application of quantum simulation algorithms and characterization techniques.

Acknowledgments

We thank Troy Borneman and Chris Ferrie for suggestions and discussions. CG was supported by funding from Industry Canada, CERC, NSERC, and the Province of Ontario.

Appendix A. Fisher information for Hamiltonian learning

In the main body, we stated that short-time experiments typically do not lead to good estimates of the Hamiltonian parameters. Here, we justify this claim here by computing the Fisher information, which allows us to estimate the scaling of the Cramér–Rao bound [51], which lower bounds the expected variance of any

unbiased estimator of the Hamiltonian parameters. In particular, the Fisher information matrix can be written for a Hamiltonian $H = H(\mathbf{x})$ and measurement in a basis $\{|1\rangle, \dots, |D\rangle\}$ as

$$I(H) := \mathbb{E}_d \left[\nabla \ln \Pr(d|H) \nabla^T \ln \Pr(d|H) \right], \quad (\text{A1})$$

where d is a random variable representing the outcome of the measurement.

Applying the chain rule and writing out the expectation value gives that

$$I_{i,j}(H) = \sum_{d \in \{1, \dots, D\}} \frac{\partial_{x_i} \Pr(d|H) \partial_{x_j} \Pr(d|H)}{\Pr(d|H)}. \quad (\text{A2})$$

By Born's rule, if we let U be the time-evolution operator for the experiment and define the initial state to be $|0\rangle$ that

$$\Pr(d|H) = |\langle d|U|0\rangle|^2 = \langle 0|U^\dagger|d\rangle \langle d|U|0\rangle. \quad (\text{A3})$$

We then have that

$$\partial_{x_i} \Pr(d|H) = \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle d|U|0\rangle + \langle 0|U^\dagger|d\rangle \langle d|\partial_{x_i} U|0\rangle. \quad (\text{A4})$$

Upon substituting back into (A2), this yields

$$\begin{aligned} I_{i,j}(H) &= \sum_d \frac{\langle 0|\partial_{x_i} U^\dagger|d\rangle \langle d|U|0\rangle \langle 0|\partial_{x_j} U^\dagger|d\rangle \langle d|U|0\rangle + \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle d|U|0\rangle \langle 0|U^\dagger|d\rangle \langle d|U|0\rangle}{\langle 0|U^\dagger|d\rangle \langle d|U|0\rangle} \\ &+ \sum_d \frac{\langle 0|U^\dagger|d\rangle \langle d|\partial_{x_i} U|0\rangle \langle 0|\partial_{x_j} U^\dagger|d\rangle \langle d|U|0\rangle + \langle 0|U^\dagger|d\rangle \langle d|\partial_{x_i} U|0\rangle \langle 0|U^\dagger|d\rangle \langle d|\partial_{x_j} U|0\rangle}{\langle 0|U^\dagger|d\rangle \langle d|U|0\rangle} \\ &= \sum_d \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle 0|\partial_{x_j} U^\dagger|d\rangle \frac{\langle d|U|0\rangle}{\langle 0|U|d\rangle} + \langle 0|\partial_{x_i} U|d\rangle \langle 0|\partial_{x_j} U|d\rangle \frac{\langle 0|U|d\rangle}{\langle d|U|0\rangle} \\ &+ \sum_d \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle 0|\partial_{x_j} U|d\rangle + \langle 0|\partial_{x_i} U|d\rangle \langle 0|\partial_{x_j} U^\dagger|d\rangle. \end{aligned} \quad (\text{A5})$$

It is then straight forward to see that there exists ϕ_d such that for every d

$$\langle 0|U|d\rangle = e^{i\phi_d} \langle d|U^\dagger|0\rangle. \quad (\text{A6})$$

Furthermore from differentiating $UU^\dagger = \mathbb{1}$ and using the fact that U is unitary, it is clear that for $\|\cdot\|$ the induced 2-norm,

$$\|\partial_{x_i} U^\dagger\| = \|\partial_{x_i} U\|. \quad (\text{A7})$$

Seeking an upper bound on the Fisher information, we use the Cauchy–Schwarz inequality to show that

$$\begin{aligned} &\sum_d \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle 0|\partial_{x_j} U^\dagger e^{-i\phi_d}|d\rangle \\ &\leq \sqrt{\sum_d \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle d|\left(\partial_{x_i} U^\dagger\right)^\dagger|0\rangle} \sqrt{\sum_d \langle 0|\partial_{x_j} U^\dagger|d\rangle \langle d|\left(\partial_{x_j} U^\dagger\right)^\dagger|0\rangle}. \end{aligned} \quad (\text{A8})$$

Using the resolution of unity and the fact that $\|A^\dagger\| = \|A\|$ for the 2-norm, we find from (A8) and (A7) that

$$\sum_d \langle 0|\partial_{x_i} U^\dagger|d\rangle \langle 0|\partial_{x_j} U^\dagger e^{-i\phi_d}|d\rangle \leq \|\partial_{x_i} U\| \|\partial_{x_j} U\|. \quad (\text{A9})$$

The triangle inequality and equations (A9) and (A5) then imply that

$$I_{i,j}(H) \leq 4 \|\partial_{x_i} U\| \|\partial_{x_j} U\|. \quad (\text{A10})$$

An experiment for either the case where an inversion step is employed or the case where only forward evolution is used can be written using the unitary $U = e^{iH_- t} e^{-iHt}$, where $H_- = 0$ for the inversion-free case. Regardless, H_- is explicitly independent of the parameters $\{x_p\}$ of H ; therefore since U is unitary,

$$\|\partial_{x_p} U\| = \|e^{iH_- t} \partial_{x_p} e^{-iHt}\| = \|\partial_{x_p} e^{-iHt}\|. \quad (\text{A11})$$

Using the definition of the parametric derivative of an operator exponential, we find using the triangle inequality that

$$\|\partial_{x_p} e^{-iHt}\| = \left\| \int_0^1 e^{(1-\tau)(-iHt)} (-it\partial_{x_p} H) e^{\tau(-iHt)} d\tau \right\| \leq \|\partial_{x_p} H\| t. \quad (\text{A12})$$

This leads us to the conclusion that

$$I_{i,j}(H) \leq 4\|\partial_{x_i} H\| \|\partial_{x_j} H\| t^2. \quad (\text{A13})$$

The Cramér–Rao bound then states that, for any unbiased estimator \hat{x} of the Hamiltonian parameters [52],

$$\mathbb{E}_{\text{data}} \left[\text{Cov}(\hat{x}) \right] - I^{-1}(x) \geq 0, \quad (\text{A14})$$

where the expectation value is taken over all data records, here taken to be measurements of $d \in \{1, \dots, D\}$ [51]. Tracing both sides of the inequality immediately implies that the variance of any unbiased estimator of the Hamiltonian parameters scales with $\text{Tr}[I(H)^{-1}] \in \Omega(t^{-2})$.

We also consider the Bayesian Cramér–Rao bound [53–55], which bounds the performance of biased estimators by taking the expectation of the Cramér–Rao bound over a prior π ,

$$\mathbb{E}_{x \sim \pi} \left[\mathbb{E}_{\text{data}} \left[\text{Cov}(\hat{x}) \right] \right] - \mathbb{E}_{x \sim \pi} \left[I^{-1}(H(x)) \right] \geq 0. \quad (\text{A15})$$

Here, we note that the t^2 scaling obtained in (A13) is independent of x , it factors out of the expectation over Hamiltonian parameters, such that the Bayesian Cramér–Rao bound is also $\Omega(t^{-2})$ by the same argument, such that even biased estimators require evolution time that is the reciprocal of the desired standard deviation.

This implies that as $t \rightarrow 0$ the lower bound on the variance of the optimal estimator for x diverges, implying that the experiments become uninformative for the small values of time required for existing Hamiltonian identification methods to succeed. Also, since the cost of performing an experiment becomes dominated by the time required to prepare the initial state for small t , it is clear that the reduced cost of short-time experiments will not compensate for the exponentially diverging CRB and BCRB in such cases.

Appendix B. Lieb–Robinson bounds

We provide rigorous estimates for the truncation error in cases where the Hamiltonian is non-commuting in this section. The proof of the error bounds is elementary, with the exception that the results depend on the use of Lieb–Robinson bounds. To begin, let us first define some notation. Let us assume that r time reversals are used, that the Trotter formula (rather than higher order variants) is used, and then let us define the observable after n evolutions/inversions to be

$$A^{(n)} := e^{iHt/r} e^{-iH_-t/r} A^{(n-1)} e^{iH_-t/r} e^{-iHt/r}, \quad (\text{B1})$$

with $A^{(0)} = A$. Now, let $H_- = H_{\text{in}} - \Lambda$, where Λ is the discrepancy between the inversion Hamiltonian and the true Hamiltonian, supported on the region that can be simulated by the trusted device. Let us also define

$$\tilde{A}^{(n)} := e^{i\Lambda t/r} \tilde{A}^{(n-1)} e^{-i\Lambda t/r}, \quad (\text{B2})$$

where $\tilde{A}^{(0)} = A$; since we have assumed a Trotter formula, $[H_-, H_{\text{int}}]t/r \approx 0$, such that $\tilde{A}^{(n)}$ represents the observable as simulated by the trusted device alone. Thus, define the error operator $\delta^{(n)}$ such that

$$\delta^{(n)} := A^{(n)} - \tilde{A}^{(n)}. \quad (\text{B3})$$

The goal of this section will then be to provide upper bounds on $\|\delta^{(n)}\|$ which represents the error incurred from truncating the trusted simulator.

First, note that $\|\delta^{(0)}\| = 0$. This will serve as the base case in our inductive argument about the norm of $\delta^{(n)}$. The triangle inequality, together with the unitary invariance of $\|\cdot\|$, implies

$$\begin{aligned} \|\delta^{(n+1)}\| &= \|e^{iHt/r} e^{-iH_-t/r} A^{(n)} e^{iH_-t/r} e^{-iHt/r} - e^{i\Lambda t/r} \tilde{A}^{(n)} e^{-i\Lambda t/r}\| \\ &\leq \|\delta^{(n)}\| + \|e^{iHt/r} e^{-iH_-t/r} \tilde{A}^{(n)} e^{iH_-t/r} e^{-iHt/r} - e^{i\Lambda t/r} \tilde{A}^{(n)} e^{-i\Lambda t/r}\|. \end{aligned} \quad (\text{B4})$$

Equation (B4) provides a recursive expression for the error after $n + 1$ steps in terms of the error after n steps. Our bounds for the error in (10) follow from unfolding this recurrence relation after applications of the triangle inequality. The main challenge is that $e^{iHt/r} e^{-iH_-t/r} \tilde{A}^{(n)} e^{iH_-t/r} e^{-iHt/r} - e^{i\Lambda t/r} \tilde{A}^{(n)} e^{-i\Lambda t/r}$ is difficult to bound directly. So instead we introduce a telescoping series of terms such that the difference between any two consecutive terms in the series can be estimated. We then arrive at (10) by using the triangle inequality.

Our first such step considers the Trotter error involved in using the approximation $e^{-iH_-t/r} \approx e^{-iH_{\text{in}}t/r} e^{-i\Lambda t/r}$ in (B4). First, note that

$$\begin{aligned}
& \left\| e^{iHt/r} e^{-iH_-t/r} \tilde{A}^{(n)} e^{iH_-t/r} e^{-iHt/r} - e^{iHt/r} e^{-iH_{in}t/r} \tilde{A}^{(n+1)} e^{iH_{in}t/r} e^{-iHt/r} \right\| \\
&= \left\| e^{-iH_-t/r} \tilde{A}^{(n)} e^{iH_-t/r} - e^{-iH_{in}t/r} e^{i\Lambda t/r} \tilde{A}^{(n)} e^{-i\Lambda t/r} e^{iH_{in}t/r} \right\| \\
&= \left\| \left[e^{-i\Lambda t/r} e^{iH_{in}t/r} e^{-iH_-t/r}, \tilde{A}^{(n)} \right] \right\|.
\end{aligned} \tag{B5}$$

Using the result of Huyghebaert and De Raedt [56], we have that for any two operators A, B that have commutators of bounded norm

$$\|e^A e^B - e^{A+B}\| \leq \frac{1}{2} \| [A, B] \|. \tag{B6}$$

Then, using (B6) we see that there exists an operator C such that $\|C\| \leq 1$ and $e^A e^B e^{-(A+B)} = 1 + \frac{C}{2} \| [A, B] \|$.

Now noting that $H_- = H_{in} - \Lambda$, we see that

$$\left\| \left[e^{-i\Lambda t/r} e^{iH_{in}t/r} e^{-iH_-t/r}, \tilde{A}^{(n)} \right] \right\| = \left\| \left[e^{-i\Lambda t/r} e^{iH_{in}t/r} e^{-i(H_{in}-\Lambda)t/r}, \tilde{A}^{(n)} \right] \right\|. \tag{B7}$$

Therefore there exists an operator C with norm at most one such that

$$\begin{aligned}
\left\| \left[e^{-i\Lambda t/r} e^{iH_{in}t/r} e^{-i(H_{in}-\Lambda)t/r}, \tilde{A}^{(n)} \right] \right\| &= \left\| \left[1 + \frac{C \| [H_{in}, \Lambda] \| t^2}{2r^2}, \tilde{A}^{(n)} \right] \right\| \\
&\leq \| [H_{in}, \Lambda] \| \| \tilde{A}^{(n)} \| t^2 / r^2 = \| [H_{in}, \Lambda] \| \| A \| t^2 / r^2.
\end{aligned} \tag{B8}$$

Equation (B8) provides an upper bound for the error incurred by treating the evolution on the trusted simulator as if it were evolving separately under H_{in} and Λ during the inversion phase, rather than evolving under $H_- = H_{in} - \Lambda$.

Second, we have from similar reasoning and the facts that (a) $H = H_{out} + H_{int} + H_{in}$ and (b) H_{in} and H_{out} are disjoint in support and hence $[H_{in}, H_{out}] = 0$ that

$$\begin{aligned}
& \left\| e^{iHt/r} e^{-iH_{in}t/r} \tilde{A}^{(n+1)} e^{iH_{in}t/r} e^{-iHt/r} - e^{i(H_{out}+H_{int})t/r} \tilde{A}^{(n+1)} e^{-i(H_{out}+H_{int})t/r} \right\| \\
&\leq \| [H_{int}, H_{in}] \| \| A \| t^2 / r^2.
\end{aligned} \tag{B9}$$

It is then straightforward to see from adding and subtracting appropriate terms and then applying the triangle inequality that

$$\begin{aligned}
\| \delta^{(n+1)} \| &\leq \| \delta^{(n)} \| + \left(\| [H_{in}, \Lambda] \| + \| [H_{int}, H_{in}] \| \right) \| A \| \frac{t^2}{r^2} \\
&\quad + \left\| e^{i(H_{out}+H_{int})t/r} \tilde{A}^{(n+1)} e^{-i(H_{out}+H_{int})t/r} - \tilde{A}^{(n+1)} \right\|.
\end{aligned} \tag{B10}$$

Third, as illustrated in figure 3, there are two types of interaction terms: interactions between the neglected particles and those in the support of A and interactions between neglected qubits and those not in the support of A . The Hamiltonians composed of only these interactions are denoted $H_{int \cap A}$ and $H_{int \setminus A}$, such that

$$\begin{aligned}
& \left\| e^{i(H_{out}+H_{int})t/r} \tilde{A}^{(n+1)} e^{-i(H_{out}+H_{int})t/r} - e^{i(H_{out}+H_{int \setminus A})t/r} \tilde{A}^{(n+1)} e^{-i(H_{out}+H_{int \setminus A})t/r} \right\| \\
&= \left\| \left[e^{-i(H_{out}+H_{int \setminus A})t/r} e^{i(H_{out}+H_{int})t/r}, \tilde{A}^{(n+1)} \right] \right\|.
\end{aligned} \tag{B11}$$

Using the fact that $\left\| e^{-i(H_{out}+H_{int \setminus A})t/r} e^{i(H_{out}+H_{int})t/r} - 1 \right\| \leq \| H_{int \cap A} \| t/r$, the triangle inequality yields

$$\left\| \left[e^{-i(H_{out}+H_{int \setminus A})t/r} e^{i(H_{out}+H_{int})t/r}, \tilde{A}^{(n+1)} \right] \right\| \leq 2 \| H_{int \cap A} \| \| A \| t/r. \tag{B12}$$

This bound estimates the error incurred by neglecting direct interactions between the observable and the particles omitted from the trusted simulator. Thus (B10) can be simplified to

$$\begin{aligned}
\| \delta^{(n+1)} \| &\leq \| \delta^{(n)} \| + \left(\| [H_{in}, \Lambda] \| + \| [H_{int}, H_{in}] \| \right) \| A \| \frac{t^2}{r^2} + 2 \| H_{int \cap A} \| \| A \| t/r \\
&\quad + \left\| e^{i(H_{out}+H_{int \setminus A})t/r} \tilde{A}^{(n+1)} e^{-i(H_{out}+H_{int \setminus A})t/r} - \tilde{A}^{(n+1)} \right\|.
\end{aligned} \tag{B13}$$

Using Hadamard's lemma, this can be written as

$$\begin{aligned} \|\delta^{(n+1)}\| &\leq \|\delta^{(n)}\| + \left(\left\| [H_{\text{in}}, \Lambda] \right\| + \left\| [H_{\text{int}}, H_{\text{in}}] \right\| \right) \|A\| \frac{t^2}{r^2} + 2\|H_{\text{int}} \cap A\| \|A\| t/r \\ &\quad + \left\| i \left[H_{\text{out}} + H_{\text{int} \setminus A}, \tilde{A}^{(n+1)} \right] t/r - \frac{1}{2!} \left[H_{\text{out}} + H_{\text{int} \setminus A}, \right. \right. \\ &\quad \left. \left. \left[H_{\text{out}} + H_{\text{int} \setminus A}, \tilde{A}^{(n+1)} \right] \right] t^2/r^2 + \dots \right\|. \end{aligned} \quad (\text{B14})$$

Applying the triangle inequality, factoring and recombining terms as an exponential yields

$$\begin{aligned} \|\delta^{(n+1)}\| &\leq \|\delta^{(n)}\| + \left(\left\| [H_{\text{in}}, \Lambda] \right\| + \left\| [H_{\text{int}}, H_{\text{in}}] \right\| \right) \|A\| \frac{t^2}{r^2} + 2\|H_{\text{int}} \cap A\| \|A\| t/r \\ &\quad + \left\| \left[H_{\text{out}} + H_{\text{int} \setminus A}, \tilde{A}^{(n+1)} \right] \right\| e^{2\|H_{\text{out}} + H_{\text{int} \setminus A}\| t/r} t/r. \end{aligned} \quad (\text{B15})$$

Fourth, and finally, we apply the Lieb–Robinson bound to upper bound $\| [H_{\text{out}} + H_{\text{int} \setminus A}, \tilde{A}^{(n+1)}] \|$. Assuming that the interactions that comprise H_{int} are nearest-neighbor or exponentially decay with the graph distance between the qubits in question, the Lieb–Robinson bound states that there exist constants s and μ that are only dependent on the properties of Λ [39] such that

$$\left\| \left[H_{\text{out}} + H_{\text{int} \setminus A}, \tilde{A}^{(n+1)} \right] \right\| \leq 2\|H_{\text{out}} + H_{\text{int} \setminus A}\| \|A\| \{A\} e^{-\mu \text{dist}(A, H_{\text{out}})} \left[e^{2s|t|(n+1)/r} - 1 \right]. \quad (\text{B16})$$

Substituting (B16) into (B15) and noting that $[H_{\text{out}}, \tilde{A}^{(n+1)}] = 0$ (because Λ and H_{out} have disjoint support) yields

$$\begin{aligned} \|\delta^{(n+1)}\| &\leq \|\delta^{(n)}\| + \left(\left\| [H_{\text{in}}, \Lambda] \right\| + \left\| [H_{\text{int}}, H_{\text{in}}] \right\| \right) \|A\| \frac{t^2}{r^2} + 2\|H_{\text{int}} \cap A\| \|A\| t/r \\ &\quad + 2\|H_{\text{int} \setminus A}\| \|A\| \{A\} e^{-\mu \text{dist}(A, H_{\text{out}})} \left[e^{2s|t|(n+1)/r} - 1 \right] e^{2\|H_{\text{out}} + H_{\text{int} \setminus A}\| t/r} t/r. \end{aligned} \quad (\text{B17})$$

Applying (B17) recursively, it is then clear that

$$\begin{aligned} \|A(t) - \tilde{A}(t)\| &= \|\delta^{(r)}\| \leq \left(\left\| [H_{\text{in}}, \Lambda] \right\| + \left\| [H_{\text{int}}, H_{\text{in}}] \right\| \right) \|A\| \frac{t^2}{r^2} + 2\|H_{\text{int}} \cap A\| \|A\| t \\ &\quad + 2\|H_{\text{int} \setminus A}\| \|A\| \{A\} e^{-\mu \text{dist}(A, H_{\text{out}})} \left[e^{2s|t|} - 1 \right] e^{2\|H_{\text{out}} + H_{\text{int} \setminus A}\| t/r}. \end{aligned} \quad (\text{B18})$$

There are a few interesting points to note about (B18). Firstly, the Lieb–Robinson velocity that appears in the equation is that of Λ not H . If the PGH is used to select experiments, then the speed at which the commutator depends on the *uncertainty* in the Hamiltonian rather than the actual Hamiltonian. A consequence of this is that the Lieb–Robinson velocity here is an epistemic, rather than a physical, property of the system. This means that, even though the experimental times increase under the PGH as more information is learned about H , the Lieb–Robinson velocity relevant to this problem will shrink. Very long experiments can therefore be used without requiring that the distance between A and the neglected qubits (i.e. the volume of the trusted simulator) grows linearly with the evolution time. In particular, $\text{dist}(A, H_{\text{out}})$ must grow at most logarithmically with the evolution time rather than linearly.

It should also be noted that in cases where nearest-neighbor couplings are present, rather than exponentially decaying couplings, that these bounds are known to be loose. More sophisticated treatments of the Lieb–Robinson bounds show that the error shrinks as $e^{-\text{const} \times \text{dist}(A, H_{\text{out}})^2}$ for such systems [41]. Taken together with the observation that $H_{\text{int}} \cap A = 0$ for nearest neighbor couplings, since A is not supported on the boundary of the trusted simulator, this tighter scaling implies that the volume of the trusted simulator can be quadratically smaller in such cases.

Appendix C. Bounds for 1D Ising models

In the main body, we showed that for commuting Hamiltonians (such as the Ising model) that the maximum evolution time allowed by compressed simulation is dictated by the strength of the interactions between the observable and the neglected subsystems. Here we provide bounds that are useful for estimating the maximum value of this interaction strength, which is useful for knowing when the simulations fail. We used these bounds in our numerical simulations to set limits on the allowable evolution time, thereby allowing us to simulate dynamics on the full 50 qubit system on a desktop.

As particular examples, if we assume that the Hamiltonian is an Ising model on a line of length ℓ with non-nearest neighbor couplings between sites i and j that scale at most as $be^{-\alpha|i-j|}$, A is supported on a sites and the trusted simulator can simulate w sites then

$$\|H_{\text{int} \cap A}\|^{-1} \geq (1 - e^{-\alpha}) e^{\lfloor \frac{w-a}{2} \rfloor} / ab. \quad (\text{C1})$$

It therefore suffices to take $w - a$ logarithmic in t to guarantee error of δ for any fixed t . Similarly, if we assume the interaction strength between sites i and j is at most $b/|i - j|^\alpha$ for $\alpha > 1$ then

$$\|H_{\text{int} \cap A}\|^{-1} \geq \frac{\left(\left\lfloor \frac{w-a}{2} \right\rfloor + 1\right)^\alpha (\alpha - 1)}{aba\alpha}. \quad (\text{C2})$$

Picking $w - a \in O(t^{1/\alpha})$ guarantees fixed error δ for experimental time t . These scalings are justified below.

Assume that the Hamiltonian is an Ising model on a line of length ℓ with a trusted simulator that can simulate at most w sites and an observable that has support on a sites. We then can write the norm of the Hamiltonian terms that are neglected by the trusted simulator as $\|H_{\text{int} \cap A}\| \leq a \sum_{j=\lfloor (w-a)/2 \rfloor + 1}^{\ell-a} f(j)$ for some function $f(j)$ that describes how quickly the interactions decay with distance from the observable. Here we take the lower limit of the sum to be $\lfloor (w - a)/2 \rfloor + 1$ because this is the closest possible site within the support of the observable A to the un-modeled portion of the spin chain. Note that in cases of non-periodic boundary conditions this minimum distance may be farther for simulations that occur near the end of the chain. Similarly, the furthest any site can be in the chain from A is $\ell - a$ which justifies the upper bound for the sum. Again this upper limit may not be tight for periodic boundary conditions.

The two most interesting cases, experimentally, are cases with exponential decay and polynomial decay with j . If we assume that $f(j) \leq be^{-(j-1)\alpha}$ then

$$\sum_{j=\lfloor (w-a)/2 \rfloor + 1}^{\ell-a} af(j) \leq \sum_{j=\lfloor (w-a)/2 \rfloor + 1}^{\infty} abe^{-(j-1)\alpha} = \frac{abe^{-\lfloor \frac{w-a}{2} \rfloor \alpha}}{1 - e^{-\alpha}}. \quad (\text{C3})$$

This justifies the claim made in the main body.

Polynomial decay is similar. Assume $f(j) \leq b/j^\alpha$ then

$$\begin{aligned} \sum_{j=\lfloor \frac{w-a}{2} \rfloor + 1}^{\ell-a} \frac{ab}{j^\alpha} &= \frac{ab}{\left(\left\lfloor \frac{w-a}{2} \right\rfloor + 1\right)^\alpha} \sum_{k=0}^{\ell-a-\lfloor \frac{w-a}{2} \rfloor - 1} \frac{1}{\left(1 + \frac{k}{(w-j)/2}\right)^\alpha} \\ &\leq \frac{ab}{\left(\left\lfloor \frac{w-a}{2} \right\rfloor + 1\right)^\alpha} \left(1 + \int_0^{\ell-a-\lfloor \frac{w-a}{2} \rfloor - 1} \frac{1}{\left(1 + \frac{k}{(w-j)/2}\right)^\alpha} dk\right). \end{aligned} \quad (\text{C4})$$

This bound can be evaluated for cases where $\alpha = 1$, and logarithmic divergence with ℓ will be observed in those cases. Given the assumption that $\ell > 1$, the integral is convergent so for simplicity we can take the limit of this equation as $\ell \rightarrow \infty$. Evaluating the integral and some elementary simplifications leads to

$$\sum_{j=\lfloor \frac{w-a}{2} \rfloor + 1}^{\ell-a} \frac{ab}{j^\alpha} \leq \frac{aba\alpha}{\left(\left\lfloor \frac{w-a}{2} \right\rfloor + 1\right)^\alpha (\alpha - 1)}. \quad (\text{C5})$$

This justifies the claim in the main body about polynomial scaling and shows that increasing w to increases the maximum value of t allowable in the experiment design step.

Appendix D. Bounds for errors in bootstrapping

To begin let us consider the error incurred by trying to find a control sequence that produces a Hamiltonian \mathcal{H}_k on an initially untrusted quantum device. If the inferred control map is G_1 and the actual control map is $G_1 + \mathcal{E}_1$ then the error in the implemented Hamiltonian, after one bootstrapping step, is

$$\left\| \left((G_1 + \mathcal{E}_1) G_1^\dagger - \mathbb{1} \right) \mathcal{H}_k \right\| \leq \left(\|G_1 G_1^\dagger - \mathbb{1}\| + \|\mathcal{E}_1\| \|G_1^\dagger\| \right) \|\mathcal{H}_k\|. \quad (\text{D1})$$

Now let us consider the error incurred after bootstrapping L times. Or in other words, consider the error that arises from using a trusted simulator that was calibrated via $L - 1$ steps of bootstrapping. If we define G_j and \mathcal{E}_j

to be the control maps and error operators that arise after j steps (where each \mathcal{E}_j is the error with respect to the ‘trusted simulator’ calibrated via $j - 1$ bootstrapping steps) then the error is

$$\left\| \left((G_L + \mathcal{E}_L) G_L^+ \right) (G_{L-1} + \mathcal{E}_{L-1}) G_{L-1}^+ \cdots (G_1 + \mathcal{E}_1) G_1^+ - \mathbb{1} \right\| \mathcal{H}_k \right\|. \quad (\text{D2})$$

By adding and subtracting $(G_1 + \mathcal{E}_1) G_1^+, (G_2 + \mathcal{E}_2) G_2^+ (G_1 + \mathcal{E}_1) G_1^+$ and so forth from (D2) we obtain from the triangle inequality that the error is at most

$$\sum_{j=1}^L \left(\|G_j G_j^+ - \mathbb{1}\| + \|\mathcal{E}_j\| \|G_j^+\| \right) \prod_{k=1}^{j-1} \left(\|G_k\| \|G_k^+\| + \|\mathcal{E}_k\| \|G_k^+\| \right). \quad (\text{D3})$$

Noting that the condition number, κ_k , for G_k is $\|G_k\| \|G_k^+\|$, (D3) can be upper bounded by the maximum values of each of the terms involved. If we specifically define Γ_{\max} to be the maximum value of $\|G_j G_j^+ - \mathbb{1}\| + \|\mathcal{E}_j\| \|G_j^+\|$ and κ_{\max} to be the maximum condition number then

$$\begin{aligned} & \sum_{j=1}^L \left(\|G_j G_j^+ - \mathbb{1}\| + \|\mathcal{E}_j\| \|G_j^+\| \right) \prod_{k=1}^{j-1} \left(\|G_k\| \|G_k^+\| + \|\mathcal{E}_k\| \|G_k^+\| \right) \\ & \leq L \Gamma_{\max} \left(1 + (\kappa_{\max} - 1) + \|\mathcal{E}_{\max}\| \|G_{\max}^+\| \right)^{L-1}. \end{aligned} \quad (\text{D4})$$

The result in (19) then follows from the fact that $(1 + x) \leq e^x$ for all $x \in \mathbb{R}$.

Note that this bound is expected to be quite pessimistic for bootstrapping in general. The analysis makes liberal use of the triangle inequality and uses worst case estimates on top of that. Additionally, the user in the bootstrapping protocol has some knowledge of the error from the fact that $G_j G_j^+ - \mathbb{1}$ can be computed for these problems since the matrices are of polynomial size. We avoid including this knowledge in the argument since the user does not necessarily know what \mathcal{E}_j is and hence it is conceivable in extremely rare cases that the errors from the approximate inversion could counteract the errors in the Hamiltonian inference. A more specialized argument may be useful for predicting better bounds for the error in specific applications.

Finally, if the swap gates also have miscalibration errors of Δ then there is a maximum value of r that can be used before the contributions of such errors become dominant. A simple inductive argument shows that

$$r \leq \frac{1}{2} \left(\frac{\delta}{\Delta} + 1 \right), \quad (\text{D5})$$

suffices to guarantee that such errors sum to at most δ . This shows that the protocol is only modestly sensitive to such errors and that if quantum bootstrapping is used to calibrate the swap gates then it is reasonable to expect that δ can often be made sufficiently small using a logarithmic number of experiments.

References

- [1] Wiebe N, Braun D and Lloyd S 2012 *Phys. Rev. Lett.* **109** 050505
- [2] Kassal I, Whitfield J D, Perdomo-Ortiz A, Yung M-H and Aspuru-Guzik A 2011 *Annu. Rev. Phys. Chem.* **62** 185
- [3] Hastings M B, Wecker D, Bauer B and Troyer M 2015 *Quant. Inf. Comput.* **15** 1–21
- [4] Shor P W 1997 *SIAM J. Comput.* **26** 1484
- [5] Amento B, Rötteler M and Steinwandt R 2013 *Quantum Inf. Comput.* **13** 631
- [6] Chow J M *et al* 2014 *Nat. Commun.* **5** 4015
- [7] Barz S, Kashefi E, Broadbent A, Fitzsimons J F, Zeilinger A and Walther P 2012 *Science* **335** 303
- [8] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
- [9] da Silva M P, Landon-Cardinal O and Poulin D 2011 *Phys. Rev. Lett.* **107** 210404
- [10] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 *New J. Phys.* **14** 103013
- [11] Stenberg M P V, Sanders Y R and Wilhelm F K 2014 *Phys. Rev. Lett.* **113** 210404
- [12] Wiebe N, Granade C, Ferrie C and Cory D 2014 *Phys. Rev. Lett.* **112** 190501
- [13] Wiebe N, Granade C, Ferrie C and Cory D 2014 *Phys. Rev. A* **89** 042314
- [14] Aharonov D, Ben-Or M and Eban E 2008 arXiv:0810.5375
- [15] Kapourniotis T, Kashefi E and Datta A 2014 *Proc. 9th Conf. on Theory of Quantum Computation, Communication and Cryptography* pp 176–204
- [16] Aolita L, Gogolin C, Kliesch M and Eisert J 2014 arXiv:1407.4817
- [17] Becerra F E, Fan J, Baumgartner G, Goldhar J, Kosloski J T and Migdall A 2013 *Nat. Photonics* **7** 147
- [18] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
- [19] Huszar F and Houlsby N M T 2012 *Phys. Rev. A* **85** 052120
- [20] Ferrie C and Blume-Kohout R 2012 Estimating the bias of a noisy coin *AIP Conf. Proc.* **1443** 14–21
- [21] Shulman M D, Harvey S P, Nichol J M, Bartlett S D, Doherty A C, Umansky V and Yacoby A 2014 *Nat. Commun.* **5** 5156
- [22] Sergeevich A, Chandran A, Combes J, Bartlett S D and Wiseman H M 2011 *Phys. Rev. A* **84** 052315
- [23] Schirmer S G and Langbein F C 2010 *4th Int. Symp. on Communications, Control and Signal Processing (ISCCSP)* pp 1–5
- [24] Schirmer S G and Oi D K L 2009 *Phys. Rev. A* **80** 022333
- [25] Ferrie C, Granade C E and Cory D G 2013 *Quantum Inf. Process.* **12** 611

- [26] Doucet A, Godsill S and Andrieu C 2000 *Stat. Comput.* **10** 197
- [27] Granade C, Ferrie C and Cory D G 2015 *New J. Phys.* **17** 013042
- [28] Bravyi S and Vargo A 2013 *Phys. Rev. A* **88** 062308
- [29] Liu J and West M 2001 *Sequential Monte Carlo Methods in Practice* ed D Freitas and N Gordon (New York: Springer)
- [30] Granade C, Ferrie C et al 2012 *QInfer: Library for Statistical Inference in Quantum Information* (available at <https://github.com/csferrie/python-qinfer>)
- [31] Shan C, Tan T and Wei Y 2007 *Pattern Recognit.* **40** 1958
- [32] Isard M and Blake A 1998 *Int. J. Comput. Vis.* **29** 5
- [33] Arulampalam M S, Maskell S, Gordon N and Clapp T 2002 *IEEE Trans. Signal Process.* **50** 174
- [34] Ferrie C and Granade C E 2014 *Phys. Rev. Lett.* **112** 130402
- [35] Korenblit S et al 2012 *New J. Phys.* **14** 095024
- [36] Berry D W, Childs A, Cleve R and Kothari R 2014 *Proc. 46th ACM Symposium on Theory of Computing (STOC 2014)* pp 283–92
- [37] Ududec C, Wiebe N and Emerson J 2013 *Phys. Rev. Lett.* **111** 080403
- [38] Gogolin C, Kliesch M, Aolita L and Eisert J 2013 arXiv:1306.3995
- [39] Hastings M B and Koma T 2006 *Commun. Math. Phys.* **265** 781
- [40] Nachtergaele B and Sims R 2006 *Commun. Math. Phys.* **265** 119
- [41] Hastings M B 2010 arXiv:1008.5137
- [42] Suzuki M 1990 *Phys. Lett. A* **146** 319
- [43] Richerme P, Gong Z-X, Lee A, Senko C, Smith J, Foss-Feig M, Michalakakis S, Gorshkov A V and Monroe C 2014 *Nature* **511** 198
- [44] Alvarez G A, Kaiser R and Suter D 2013 *Ann. Phys., Lpz.* **525** 833
- [45] Beskos A, Crisan D and Jasra A 2014 *Ann. Appl. Probab.* **24** 4
- [46] Borneman T W, Granade C E and Cory D G 2012 *Phys. Rev. Lett.* **108** 140502
- [47] Childs A M and Wiebe N 2013 *J. Math. Phys.* **54** 062202
- [48] Jones E et al 2001 SciPy: Open source scientific tools for python (available at www.scipy.org)
- [49] Barbey N 2010 [github:nbarbey/fht](https://github.com/nbarbey/fht)
- [50] Kliesch M, Gogolin C and Eisert J 2014 *Many-Electron Approaches in Physics Chemistry and Mathematics, Mathematical Physics Studies* ed V Bach and L D Site (Berlin: Springer) pp 301–18
- [51] van den Bos A 2007 *Parameter Estimation for Scientists and Engineers* (New York: Wiley)
- [52] Cover T M and Thomas J A 2006 *Elements of Information Theory* (Hoboken, N.J.: Wiley-Interscience)
- [53] van Trees H L and Bell K L 1968 *Detection Estimation and Modulation Theory Part 1* 2nd edn (Oxford: Wiley)
- [54] Gill R D and Levit B Y 1995 *Bernoulli* **1** 59
- [55] Dauwels J 2005 *Proc. Int. Symp. on Information Theory* (IEEE) pp 425–9
- [56] Huyghebaert J and Raedt H D 1990 *J. Phys. A: Math. Gen.* **23** 5777