Quantum Bootstrapping via Compressed Quantum Hamiltonian Learning

Nathan Wiebe,[1,*] Christopher Granade,[2,3,†] and D .G. Cory[4,3,5,6]

[1]*Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA*
[2]*Department of Physics, University of Waterloo, Ontario N2L 3G1, Canada*
[3]*Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada*
[4]*Department of Chemistry, University of Waterloo, Ontario N2L 3G1, Canada*
[5]*Perimeter Institute, University of Waterloo, Ontario N2L 2Y5, Canada*
[6]*Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8*

Recent work has shown that quantum simulation is a valuable tool for learning empirical models for quantum systems. We build upon these results by showing that a small quantum simulators can be used to characterize and learn control models for larger devices for wide classes of physically realistic Hamiltonians. This leads to a new application for small quantum computers: characterizing and controlling larger quantum computers. 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. Whereas Fisher information analysis shows that current methods which employ short-time evolution are suboptimal, interactive quantum learning allows us to overcome this limitation. 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.

Rapid progress has been made within the last few years towards building computationally useful devices that 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: building 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].

We develop this application by building on recent work on quantum Hamiltonian learning (QHL) which shows that a quantum simulator or quantum computer can be used to exponentially reduce the cost of learning a Hamiltonian model for the system [8, 9] relative to state of the art methods such as classical particle filters [10, 11]. Quantum Hamiltonian learning also incorporates prior information to reduce the number of parameters, avoiding the exponential growth encountered by tomography-based approaches to characterization [12]. Importantly, quantum Hamiltonian learning has been shown to be robust to a wide variety of errors [9], such that even in the presence of large-scale quantum simulators, it remains a potentially practical algorithm for characterizing large quantum systems.

A major limitation of the QHL approach suggested in [8] is that the quantum simulator used to characterize the system is taken to be as large, or larger, than the system of interest. This means that its utility is limited in cases where we would like to use it to characterize and control a larger quantum system. Alternative classical methods, such as the method of Da Silva et al [13] do not suffer from this problem. However, the assumptions that make classical learning practical in such methods may render them impractical for learning control maps for poorly calibrated quantum devices, which need not satisfy these assumptions. Furthermore, the short time evolutions used in [13] to infer the Hamiltonian is shown in Appendix A to be suboptimal for devices in which ensemble measurements are expensive. A new method that overcomes the drawbacks of both QHL and existing classical alternatives is therefore needed before quantum control and characterization of large quantum systems becomes a reality.

RESULTS

Our approach combines ideas from both quantum Hamiltonian learning [8] and information locality [13] to circumvent the limitations of each approach alone. Our results therefore naturally lead to two distinct applications:

Compressed QHL: Learning a Hamiltonian model for a large quantum system with rapidly decaying interactions using a small quantum simulator.

Quantum bootstrapping: Designing controls for a larger quantum system with rapidly decaying interactions using a small quantum simulator.

Information locality is what enables compressed QHL 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 [13–16]. Lieb–Robinson bounds give

* nawiebe@microsoft.com
† cgranade@cgranade.com

an effective "light cone", as illustrated in Figure 1, in which the evolution of an observable A can be accurately simulated without needing to consider qubits outside 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 [14–16]. Thus, if st is bounded above by a constant and the support of A is small then the measurement can be efficiently simulated.

We also use ideas from quantum Hamiltonian learning by swapping the quantum state of a subsystem of the larger (uncharacterized) system into a quantum simulator and that approximately inverts the evolution by simulating the inverse of a guess for the Hamiltonian dynamics. One step of this process is illustrated in Figure 1. This not only leads to more informative experiments but an expansion of the light cone, as repeated applications of swaps and inverse simulations delay the rate at which the light cone propagates from the observable. This 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 [13].

We show several important analytical results involving compressed QHL. The first, and perhaps most significant, is an upper bound on the error involved in simulating $A(t)$ that reveals that the simulation error can be made negligibly small at low costs in both space and experimental data for systems with exponentially decaying interactions. Secondly, we provide bounds that are potentially tighter in cases where the Hamiltonian terms mutually commute. Finally, we show that the algorithm is efficient if the Hamiltonian has interactions that decay exponentially with distance and each experiment yields a fraction of a bit of information about H.

Compressed QHL is a natural stepping stone towards quantum bootstrapping, which uses a small quantum simulator to design controls for a larger quantum device. This design process can then be repeated by using the newly certified simulator to design controls for an even larger simulator. Bootstrapping uses compressed QHL to learn the effect of each control in isolation and then inverts the resulting control map to find optimized control settings for the target dynamics. We provide upper bounds for the error in bootstrapping that show that the process is efficient if compressed QHL is efficient and the control map is invertible.

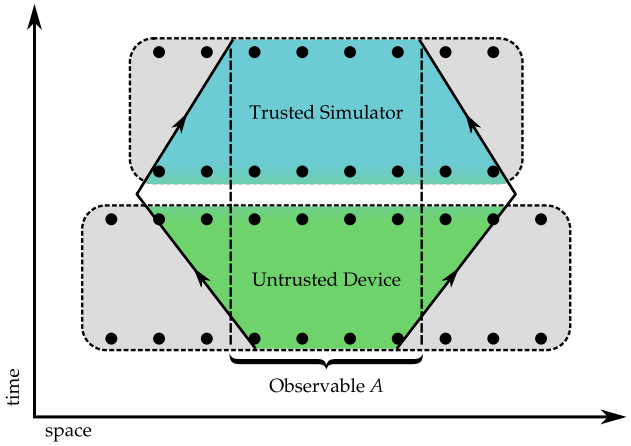


Figure 1: Light cones for $A(t)$ for a single step of an r step protocol. The green region is the light cone after the evolution in the untrusted device, and the blue region is after inversion in the trusted device. The dashed lines show the spread of $A(t)$ due to inexact inversion in the trusted simulator.

In order to show that the compressed quantum Hamiltonian learning 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 compressed quantum Hamiltonian learning 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.

METHODS

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]. Bayesian approaches to characterization are well-suited for quantum information, owing to their generality and the ease with which prior

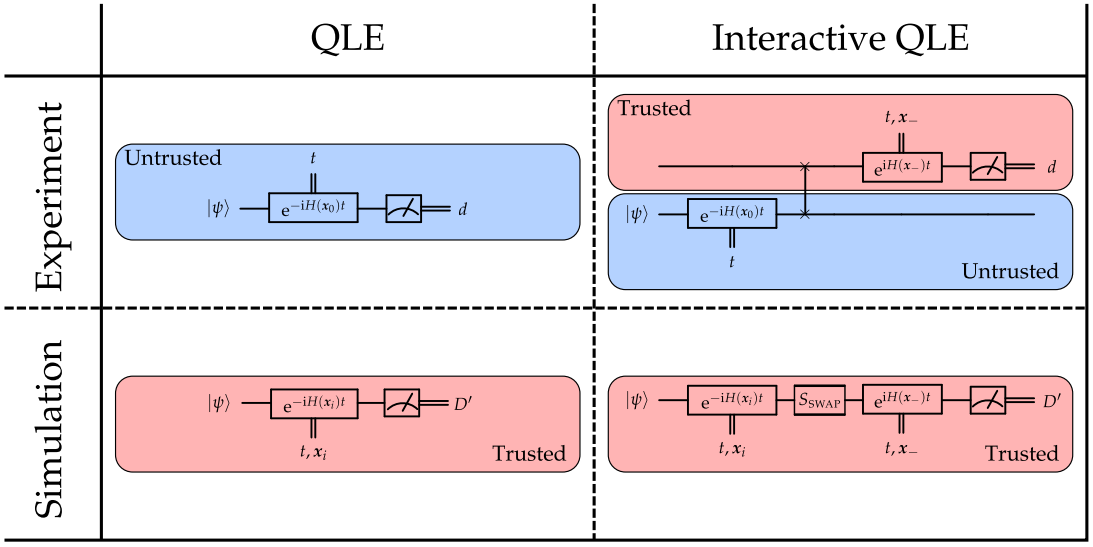


Figure 2: Experiment and simulator design for (left) quantum Hamiltonian learning and (right) interactive quantum Hamiltonian learning with an un-truncated quantum simulator. For generality, we also include S_{SWAP} in the simulation, which models noise or imperfections in the swap gate. Here we take $S_{\mathrm{SWAP}} = \mathbb{1}$.

information can be incorporated into the algorithm. Moveover, 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,

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

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 [8–10, 25].

Once the posterior distribution is found, estimates $\hat{H}$ of the Hamiltonian can be produced by considering the expectation over the posterior,

$$\hat{H} := \mathbb{E}[H|\text{data}] = \int H \Pr(H|\text{data}) \, \mathrm{d}H. \tag{2}$$

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. These samples then allow us to *approximate* integrals over the posterior, such as $\hat{H}$. 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].

Hamiltonians are not usually represented explicitly as matrices when using SMC algorithms, but are instead often parameterized by a vector $\boldsymbol{x}$ of model parameters such that $H = H(\boldsymbol{x})$. 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,

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

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 $\boldsymbol{x}_i$, and having a relative weights $\{w_i\}$. These weights are then updated by substituting the SMC approximation (3) into Bayes' rule (1) to obtain

$$w_i \mapsto w_i \Pr(d|\boldsymbol{x}_i), \tag{4}$$

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 [28].

Quantum Hamiltonian Learning

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 $\boldsymbol{x}$. 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 [29]. In this way, QHL is a semi–quantum algorithm when augmented with strong measurement of a quantum simulation device [8]. 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 [9]. This latter property is of particular importance, as it allows us to use as an approximate simulator a *truncation* of the complete system.

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 $H(\boldsymbol{x}_0)$ for some time t, and then measures $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$ on the trusted simulator. This experiment is then repeated for each SMC hypothesis $\boldsymbol{x}_i$ until the variance in the estimated likelihood becomes sufficiently small. The experiment design is illustrated in Figure 2. QLE can be effective for learning Hamiltonians, though 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 result for typical Hamiltonians (such as Gaussian random Hamiltonians [30, 31]) 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 update step in the SMC algorithm becomes unstable under in such cases [9] which necessitates using short, and in turn uninformative, experiments (see Appendix A).

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 2. 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 $H(\boldsymbol{x}_0) - H(\boldsymbol{x}_-)$ provided we choose an inversion hypothesis $H(\boldsymbol{x}_-)$ that is close to $H(\boldsymbol{x}_0)$. Intuitively, such experiments directly compare the dynamics of $H(\boldsymbol{x})$ and $H(\boldsymbol{x}_-)$. 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.

In order to combat the exponentially diminishing likelihood of the system returning to its initial state after the inversion, we require that $\|H(\boldsymbol{x}) - H(\boldsymbol{x}_-)\|t$ is approximately constant [8]. We use the particle guess heuristic (PGH) to achieve this. The PGH involves drawing two hypotheses about H, $\boldsymbol{x}_-$ and $\boldsymbol{x}'_-$, from the prior distribution and then choosing $t = 1/\|H(\boldsymbol{x}_-) - H(\boldsymbol{x}'_-)\|$. Since $\|H(\boldsymbol{x}_-) - H(\boldsymbol{x}'_-)\|$ is an estimate of the uncertainty in the Hamiltonian, we expect that at most a constant fraction of the prior distribution will satisfy $\|H(\boldsymbol{x}_-) - H(\boldsymbol{x})\|t \geq 1$ assuming $H(\boldsymbol{x})$ 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 $|\langle\psi| e^{iH(\boldsymbol{x}_-)t} e^{-iH(\boldsymbol{x})t} |\psi\rangle|^2 \sim 1/2^n$ for most $\boldsymbol{x}$. Also since the PGH relies only on the current SMC approximation to the posterior, and thus incurs no additional simulation costs. It is furthermore an adaptive method since the distribution for $\boldsymbol{x}_-$ and $\boldsymbol{x}'_-$ depends on the current state of knowledge about the quantum system. Experiment design via the particle guess heuristic has been shown to enable the efficient estimation of Hamiltonians using IQLE [8], and has since been usefully applied in other experimental contexts [11].

Previous work has analyzed the complexity of learning using IQLE [8]. 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 $O(N_{\text{samp}} N_{\text{particles}} \log(1/\delta)/\gamma)$ simulations to learn the Hamiltonian to within error δ. 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 [8, 9]. 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 [8]. This occurs for well posed learning problems that use two outcome experiments.

Learning Commuting Hamiltonians

We begin 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 (21) that we consider in numeric examples. In this case, compressed IQLE is particularly simple to analyze; we discuss the more general case later.

Consider an observable A that has support on a sites in a graph to analyze the error introduced by the com-

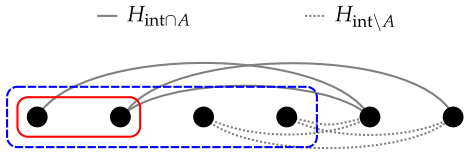


Figure 3: Separation of $H_{\text{int}} = H_{\text{int}\cap A} + H_{\text{int}\setminus A}$ where $H_{\text{int}\cap A}$ are interactions with qubits in the support of A (red solid box) and $H_{\text{int}\setminus A}$ interacts with qubits that are swapped into the trusted simulator but are outside A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

pressed simulation, and expand the Hamiltonian of the system as

$$\begin{aligned} H &= H_{\text{int}} + H_{\text{in}} + H_{\text{out}} \\ &= H_{\text{int}\cap A} + H_{\text{int}\setminus A} + H_{\text{in}} + H_{\text{out}}, \end{aligned} \tag{5}$$

where H_{out} represents all terms that do not interact with the subsystems in the trusted simulator, $H_{\text{int}} = H_{\text{int}\cap A} + H_{\text{int}\setminus A}$ is the sum of all interactions between these subsystems and the simulated subsystems, and where H_{in} represents the internal dynamics common to both the simulator and the larger system. The decomposition of the interaction Hamiltonian H_{int} into couplings that include and exclude A is illustrated in Figure 3.

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

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

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

$$\|\tilde{A}(t) - A(t)\| \leq \|A\|(e^{2\|H_{\text{int}\cap A}\|t} - 1) \tag{7}$$

Thus, if we can tolerate an error of δ in the simulation then it suffices to choose experiments with evolution time at most

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

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 $\delta \to 0$. 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 [8, 9] 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 bootstrapping therefore only occur in commuting models with long range interactions.

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

$$\|H_{\text{int}\cap A}\|^{-1} \geq (1 - e^{-\alpha})e^{\lfloor \frac{w-a}{2} \rfloor}/ab. \tag{9}$$

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

$$\|H_{\text{int}\cap A}\|^{-1} \geq \frac{(\lfloor \frac{w-a}{2} \rfloor + 1)^{\alpha}(\alpha - 1)}{ab\alpha}. \tag{10}$$

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

Learning Non–commuting Hamiltonians

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

$$\begin{aligned} &e^{i(H_{\text{in}}+H_{\text{int}}+H_{\text{out}})} A e^{-i(H_{\text{in}}+H_{\text{int}}+H_{\text{out}})} \\ &\qquad \neq e^{i(H_{\text{in}}+H_{\text{int}})} A e^{-i(H_{\text{in}}+H_{\text{int}})}, \end{aligned} \tag{11}$$

unlike in commuting models. A further issue arises from the fact that such dynamics can rapidly lead to observables, $A(t)$, that have non–trivial support on the edge 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 $e^{iH_-t} S e^{-iHt}$, 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: $(e^{iH_-t/r} S e^{-iHt/r})^r$ before measuring A. It is then easy to see from the Trotter formula that $e^{iH_-t/r} e^{-iHt/r} \approx e^{-i(H-H_-)t/r}$. This 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 that the error from using a small trusted simulator obeys

$$\begin{aligned}\|A(t) - \tilde{A}(t)\| \leq (\|[H_{\text{in}}, \Lambda]\| + \|[H_{\text{int}}, H_{\text{in}}]\|)\|A\|\frac{t^2}{r} + 2\|H_{\text{int}\cap A}\|\|A\|t \\ + 2\|H_{\text{int}\setminus A}\|\|A\|\|\{A\}|te^{-\mu\text{dist}(A,H_{\text{out}})}\left[e^{2s|t|}-1\right]e^{2\|H_{\text{out}}+H_{\text{int}\setminus A}\|t/r},\end{aligned} \tag{12}$$

for cases of nearest–neighbor or exponentially decaying interactions between subsystems. 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 (12) can be improved by using higher order Trotter–Suzuki formulas in place of the basic Trotter formula to reduce r [32] and also by using tighter bounds for cases with nearest–neighbor Hamiltonians.

The variable Λ is related through the particle guess heuristic to the uncertainty in the Hamiltonian, which implies that the speed of information propagation is also a function of the uncertainty in H [14, 15]. 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 $s \to 0$ as $\Lambda \to 0$.

Of course, the error term $2\|H_{\text{int}\cap A}\|t$ in (12) does place 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 compressed QHL 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 $r \to \infty$

$$t \leq \frac{\delta}{2\|A\|(\|H_{\text{int}\cap A}\| + 2\|H_{\text{int}\setminus A}\|\|\{A\}|e^{-\mu(w-a)/4})}, \tag{13}$$

suffices to guarantee simulation error of δ. This result is qualitatively similar to the commuting case in (8).

Note that 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

$$r \leq \frac{1}{2}\left(\frac{\delta}{\Delta} + 1\right), \tag{14}$$

suffices to guarantee that such errors sum to at most δ.

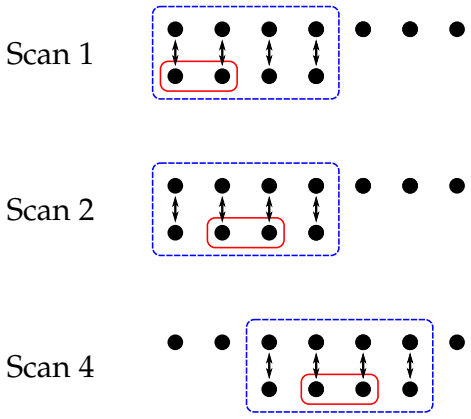


Figure 4: Scanning procedure for 7 qubits, a 4 qubit simulator and a 2 qubit observable. Blue (dashed) box is support of simulator, red (solid) box is support of A.

This shows that the protocol is only modestly sensitive to such errors.

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 inferred Hamiltonians. 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 use two SMC 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 compressed QHL 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, we approximate 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 compressed QHL 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.

Learning Controls

Having shown a procedure for learning Hamiltonians using compressed quantum simulators, we now apply our algorithm to infer the map from control settings to applied control Hamiltonians. This is of particular importance, for instance, 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 process.

To show how quantum Hamiltonian learning can address this challenge, we consider a model in which a row–vector of control settings $\boldsymbol{C}$ is related to the system Hamiltonian by an affine map $H(\boldsymbol{C})$,

$$H(\boldsymbol{C}) = \boldsymbol{C} \cdot [H_1, \ldots, H_M] + H_0 \tag{15}$$

for some $M+1$ unknown Hamiltonians $\{H_0, \ldots, H_M\}$. By the same argument as before, let $H_j = H(\boldsymbol{x}_j)$ 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 $\boldsymbol{C} = \boldsymbol{0}$ and learn H_0 using compressed QHL.

(b) For $k = 1, \ldots, M$ set $C_j = \delta_{j,k}$ and learn $H_k + H_0$ using the compressed quantum Hamiltonian learning procedure above.

(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 $H(\boldsymbol{C})$ is given by $G\boldsymbol{C}^T$, which allows the effect of control on the quantum system of interest to be predicted.

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

Bootstrapping

We complete our description of quantum bootstrapping by detailing how control learning can be used to calibrate an initially untrusted device. That is, we use knowledge of the control landscape to find control settings that allow the system to act as a larger trusted simulator. This larger device is then used in subsequent rounds of bootstrapping to calibrate an even larger quantum device.

If $H(\boldsymbol{C})$ is an affine map then this can be accomplished using the following approach:

(a) Learn $H(\boldsymbol{C})$ 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 $\boldsymbol{C}_j$ such $H(\boldsymbol{C}_j) \approx \mathcal{H}_j$.

(d) Treat the system as a trusted simulator and repeat steps (a), (b) and (c) for a larger system.

In cases where $H_0 = 0$, $H(\boldsymbol{C})$ is linear and hence $H(a\boldsymbol{C_1} + b\boldsymbol{C_2}) = a\mathcal{H}_1 + b\mathcal{H}_2$. 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 straight forward. It can be solved by applying a pseudoinverse to find $\boldsymbol{C}$ 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 $\mathcal{H}_1 = H(\boldsymbol{C}_1)$ and $\mathcal{H}_2 = H(\boldsymbol{C}_2)$ as

$$\left(e^{-i\mathcal{H}_1 a\Delta t/R} e^{-i\mathcal{H}_2 b\Delta t/R}\right)^R = e^{-i(a\mathcal{H}_1 + b\mathcal{H}_2)\Delta t} + O\left(\frac{\Delta t^2}{R}\right). \tag{16}$$

Higher–order Trotter–Suzuki methods can be used to reduce the value of R if desired [32].

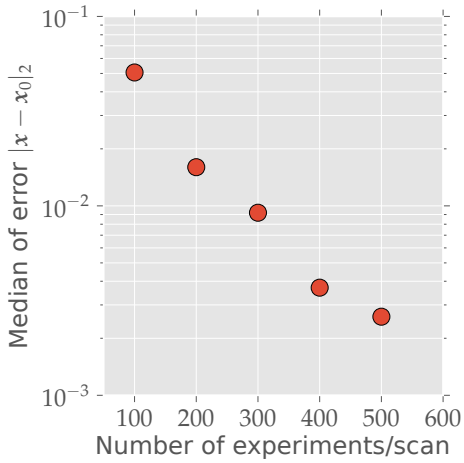


Figure 5: Error in QHL for $a = 4$ with varying N_{exp} per scan. Data consistent with $e^{-0.006 N_{\text{exp}}}$ scaling.

It is also important to note that the bootstrapped simulator 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 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:

$$\begin{aligned} &e^{-2iZ\otimes\mathbb{1}\otimes Z\Delta t^2}|\phi\rangle + O(\Delta t^3) \\ &= e^{-iZ\otimes X\otimes\mathbb{1}\Delta t}e^{-i\mathbb{1}\otimes Y\otimes Z\Delta t}e^{iZ\otimes X\otimes\mathbb{1}\Delta t}e^{i\mathbb{1}\otimes Y\otimes Z\Delta t}|\phi\rangle, \end{aligned} \tag{17}$$

where the middle qubit in $|\phi\rangle$ is set to $|0\rangle$. Higher–order and parallel methods for engineering such interactions are given in [33, 34].

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

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

where G^{+} is the pseudo–inverse of G. Eq. (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 compressed QHL 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

$$L\Gamma_{\max}e^{(L-1)(\kappa_{\max}-1+\|\mathcal{E}_{\max}\|\|G^{+}_{\max}\|)}\max_k\|H_k\|, \tag{19}$$

where $\Gamma_{\max}$ is the maximum value of $(\|GG^{+} - \mathbb{1}\| + \|\mathcal{E}\|\|G^{+}\|)$ over all the L bootstrapping steps, $\kappa_{\max}$ 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 the appendix for completeness.

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

$$N_{\text{exp}} \geq \frac{L - 1 + \ln\left(\frac{L\|G^{+}_{\max}\|\max(\max_k\|H_k\|,1)}{\Delta}\right)}{\gamma}. \tag{20}$$

This process is then clearly 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 $\Delta > 0$.

It is difficult to say in general when the conditions underlying (20) will be met. This is because 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. No information can be learned from the measurement statistics in such cases. Great experimental care must be taken in order to ensure that such pathological cases do not emerge [9]. However, we will see that these pathological experiments can be avoided Ising models with exponential decaying interactions in the following section and expect exponential decay of $\|\mathcal{E}_{\max}\|$ to be common for a wide range of models based on previous studies [8–10].

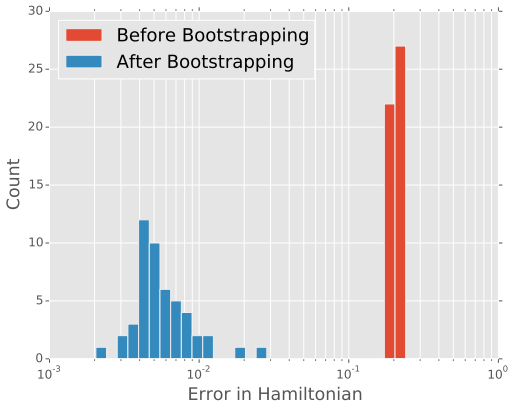
a	75$^{\text{th}}$ percentile	Median error	25$^{\text{th}}$ percentile
6	0.0043	0.0029	0.0014
4	0.0029	0.0018	0.0014
2	0.0252	0.0234	0.0225

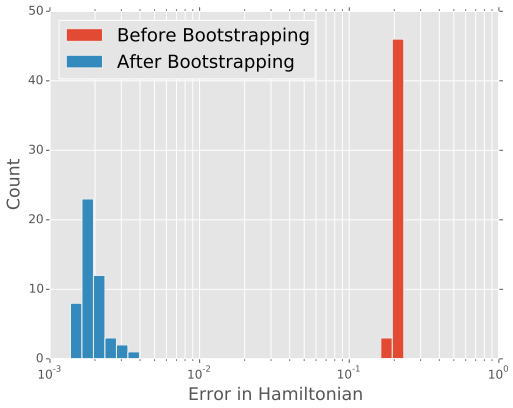
Table I: $|x - x_{\text{true}}|_2$ for QHL using different number of qubits for the observable at 500 experiments/scan.

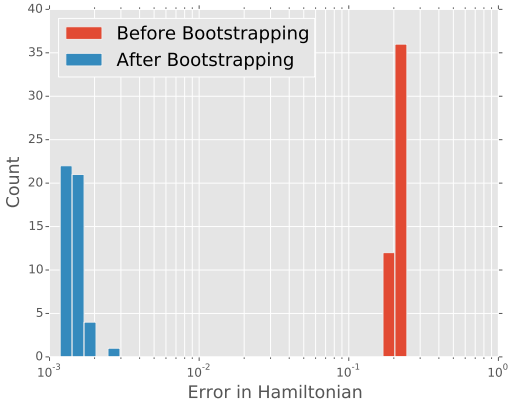
NUMERICAL RESULTS

Compressed Quantum Hamiltonian Learning

Since quantum devices capable of implementing our bootstrapping protocol do not yet exist, we examine







Errors in inferred control maps

Experiments/scan	$\|G - G_{\text{approx}}\|_2$	$\|G - G_{\text{prior}}\|_2$
100	0.3430	2.5625
200	0.0518	1.2689
300	0.0323	1.2872

Errors in couplings before and after bootstrapping

Experiments/scan	Before	After
100	0.21 ± 0.01	$(6.0 \pm 4.0) \times 10^{-3}$
200	0.21 ± 0.01	$(2.0 \pm 0.4) \times 10^{-3}$
300	0.21 ± 0.02	$(1.5 \pm 0.3) \times 10^{-3}$

Figure 6: Distribution of errors for each of the 49 Hamiltonian terms in the bootstrapped Hamiltonian for a 50 qubit Ising model using (left) 100 (right) 200 and (bottom left) 300 IQLE experiments per scan.

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,

$$H(\boldsymbol{x}) = \sum_{i \neq j} x_{i,j} \sigma_z^{(i)} \sigma_z^{(j)}, \tag{21}$$

where the parameters $x_{i,j}$ are distributed according to $x_{i,j} \sim \text{unif}(0,1) 10^{-2(|i-j|-1)}$. 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 [9].

Compressed QHL on the full 50-qubit system then gives a $\binom{50}{2} = 1225$ parameter estimation problem. In order to ensure classical simulatability, we cap the maximum value of t allowed to satisfy (8) with $\delta = 0.01$. Previous studies, show that errors on this scale lead to a negligible shift in the speed with which the algorithm learns the Hamiltonian [9]. Also, all experiments are performed with a 20,000 particle SMC approximation to the prior distribution, and the Liu-West resampling algorithm is used [28], also described in [10]. We also use the open-source implementation of SMC provided by the QInfer, SciPy and `fht` libraries [35–37].

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 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 [8, 9].

An important difference between this result and [8, 9] 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 I 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/scan. This provides evidence that small a can suffice for Hamiltonian learning.

Quantum Bootstrapping

The next set of results build upon the previous results by showing that compressed QHL 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 addressing crosstalk in the large simulator. This crosstalk manifests itself in the fact that when the experimentalist attempts to turn on only one of the Ising couplings in the simulator, all 1225 interactions are also activated. 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 $[H_1, \ldots, H_M]$ in (15) conform to (21) 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/scan are used in the QHL step. Further reductions could be achieved by increasing the number of experiments/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.

CONCLUSION

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 compressed quantum Hamiltonian learning 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; whereas it spreads at a speed that depends on the uncertainty in the Hamiltonian in compressed QHL. 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 quantum Hamiltonian learning, inherit the same noise and sample error robustness observed in that algorithm [8, 9]. 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, our quantum bootstrapping algorithm 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. Firstly, we have not investigated how to calibrate the SWAP used to interact the trusted simulator with the untrusted device. Although previous work showed that QHL is robust to such errors [9], they can be substantially larger in bootstrapping experiments because multiple rounds of SWAP gates are used to mitigate truncation errors and also such gates might not be implemented with a single Hamiltonian. Thus calibrating those controls may fall outside the purview of our Hamiltonian learning methods. Finally, 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 [38]. This is significant because techniques such [13] do not apply for $H(t)$.

As a final remark, our work partially provides an important step towards the calibration and control of 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 by 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.

[1] N. Wiebe, D. Braun, and S. Lloyd, Physical Review Letters **109**, 050505 (2012).
[2] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, Annual Review of Physical Chemistry **62**, 185 (2011).
[3] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, arXiv:1403.1539 [quant-ph] (2014).
[4] P. W. Shor, SIAM journal on computing **26**, 1484 (1997).
[5] B. Amento, M. Rötteler, and R. Steinwandt, Quantum Info. Comput. **13**, 631 (2013).
[6] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, and M. Steffen, Nature Communications **5** (2014).
[7] S. Barz, E. Kashefi, A. Broadbent, J. F. Fitzsimons, A. Zeilinger, and P. Walther, Science **335**, 303 (2012).
[8] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Physical Review Letters **112**, 190501 (2014).
[9] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Physical Review A **89**, 042314 (2014).
[10] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New Journal of Physics **14**, 103013 (2012).
[11] M. P. V. Stenberg, Y. R. Sanders, and F. K. Wilhelm, arXiv:1407.5631 (2014).
[12] G. M. D'Ariano, L. Maccone, and M. G. A. Paris, Physics Letters A **276**, 25 (2000).
[13] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Physical Review Letters **107**, 210404 (2011).
[14] M. B. Hastings and T. Koma, Communications in Mathematical Physics **265**, 781 (2006).
[15] B. Nachtergaele and R. Sims, Communications in Mathematical Physics **265**, 119 (2006).
[16] M. B. Hastings, arXiv preprint arXiv:1008.5137 (2010).
[17] F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdall, Nature Photonics **7**, 147 (2013).
[18] **12**, 10.1088/1367-2630/12/4/043034.
[19] F. Huszr and N. M. T. Houlsby, Physical Review A **85**, 052120 (2012).
[20] C. Ferrie and R. Blume-Kohout, *Estimating the bias of a noisy coin*, arXiv e-print 1201.1493 (2012) AIP Conf. Proc. 1443, pp. 14-21 (2012).
[21] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby, arXiv:1405.0485 (2014).
[22] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Physical Review A **84**, 052315 (2011).
[23] S. Schirmer and F. Langbein, in *2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)* (2010) pp. 1–5.
[24] S. G. Schirmer and D. K. L. Oi, Physical Review A **80**, 022333 (2009).
[25] C. Ferrie, C. E. Granade, and D. G. Cory, Quantum Information Processing **12**, 611 (2013).
[26] A. Doucet, S. Godsill, and C. Andrieu, STATISTICS AND COMPUTING **10**, 197 (2000).
[27] C. Granade, C. Ferrie, and D. G. Cory, arXiv:1404.5275 [quant-ph] (2014).
[28] J. Liu and M. West, in *Sequential Monte Carlo Methods in Practice*, edited by D. Freitas and N. Gordon (Springer-Verlag, New York, 2001).
[29] C. Ferrie and C. E. Granade, Physical Review Letters **112**, 130402 (2014).
[30] C. Ududec, N. Wiebe, and J. Emerson, Physical Review Letters **111**, 080403 (2013).
[31] C. Gogolin, M. Kliesch, L. Aolita, and J. Eisert, arXiv:1306.3995 [quant-ph] (2013).
[32] M. Suzuki, Physics Letters A **146**, 319 (1990).
[33] T. W. Borneman, C. E. Granade, and D. G. Cory, Physical Review Letters **108**, 140502 (2012).
[34] A. M. Childs and N. Wiebe, Journal of Mathematical Physics **54**, 062202 (2013).
[35] C. Granade, C. Ferrie, *et al.*, *QInfer: Library for Statistical Inference in Quantum Information* (2012).
[36] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for python," (2001).
[37] N. Barbey, "github:nbarbey/fht," (2010).
[38] M. Kliesch, C. Gogolin, and J. Eisert, in *Many-Electron Approaches in Physics, Chemistry and Mathematics*, Mathematical Physics Studies, edited by V. Bach and L. D. Site (Springer International Publishing, 2014) pp. 301–318.
[39] H. L. Van Trees and K. L. Bell, *Detection estimation and modulation theory. Part 1*, 2nd ed. (Wiley-Blackwell, Oxford, 1968).
[40] R. D. Gill and B. Y. Levit, Bernoulli **1**, 59 (1995).
[41] J. Dauwels, in *International Symposium on Information Theory, 2005. ISIT 2005. Proceedings* (IEEE, 2005) pp. 425–429.
[42] J. Huyghebaert and H. D. Raedt, Journal of Physics A: Mathematical and General **23**, 5777 (1990).

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, 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 $H = H(\boldsymbol{x})$ and measurement in a basis $\{|1\rangle, \ldots, |D\rangle\}$ as

$$\boldsymbol{I}(H) := \mathbb{E}_d[\boldsymbol{\nabla} \ln \Pr(d|H) \boldsymbol{\nabla}^{\mathrm{T}} \ln \Pr(d|H)], \tag{A1}$$

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

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

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

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

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

We then have that

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

Upon substituting back into (A2), this yields

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

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

$$\langle 0|U|d\rangle = e^{i\phi_d}\langle d|U^\dagger|0\rangle. \tag{A6}$$

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,

$$\|\partial_{x_i}U^\dagger\| = \|\partial_{x_i}U\|. \tag{A7}$$

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

$$\sum_d \langle 0|\partial_{x_i}U^\dagger|d\rangle\langle 0|\partial_{x_j}U^\dagger e^{-i\phi_d}|d\rangle \le \sqrt{\sum_d \langle 0|\partial_{x_i}U^\dagger|d\rangle\langle d|(\partial_{x_i}U^\dagger)^\dagger|0\rangle \sum_d \langle 0|\partial_{x_j}U^\dagger|d\rangle\langle d|\left(\partial_{x_j}U^\dagger\right)^\dagger|0\rangle}. \tag{A8}$$

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

$$\sum_d \langle 0|\partial_{x_i}U^\dagger|d\rangle\langle 0|\partial_{x_j}U^\dagger e^{-i\phi_d}|d\rangle \le \|\partial_{x_i}U\|\|\partial_{x_j}U\|. \tag{A9}$$

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

$$I_{i,j}(H) \le 4\|\partial_{x_i}U\|\|\partial_{x_j}U\|. \tag{A10}$$

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,

$$\|\partial_{x_p}U\| = \|e^{iH_-t}\partial_{x_p}e^{-iHt}\| = \|\partial_{x_p}e^{-iHt}\|. \tag{A11}$$

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

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

This leads us to the conclusion that

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

The Cramér–Rao bound then states that, for any unbiased estimator $\hat{\boldsymbol{a}}$ of the Hamiltonian parameters,

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

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

We also consider the Bayesian Cramér-Rao bound [39–41], which bounds the performance of biased estimators by taking the expectation of the Cramér-Rao bound over a prior π,

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

Here, we note that the t^2 scaling obtained in (A13) is independent of $\boldsymbol{x}$, 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 inversely linear in the desired standard deviation.

This implies that as $t \to 0$ the lower bound on the variance of the optimal estimator for $\boldsymbol{x}$ 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

Our goal in this section is to provide rigorous estimates for the truncation error in cases where the Hamiltonian is non–commuting. 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

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

with $A^{(0)} = A$. Now, let $H_- = H_{\text{in}} - \Lambda$, were 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

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

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

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

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

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

Eq. (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 (12) 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. Then by using the triangle inequality, we arrive at (12).

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

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

Using the result of Huyghebaert and De Raedt [42], 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 $e^A e^B e^{-(A+B)} = \mathbb{1} + \frac{C}{2}\|[A, B]\|$.

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

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

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

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

Eq. (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 $H_- = \mathcal{H}_{\text{in}} - \Lambda$.

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

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

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

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

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 $H_{\text{int} \cap A}$ and $H_{\text{int} \setminus A}$, such that

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

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

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

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

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

Using Hadamard's lemma, this can be written as

$$\|\delta^{(n+1)}\| \leq \|\delta^{(n)}\| + (\|[H_{\text{in}}, \Lambda]\| + \|[H_{\text{int}}, H_{\text{in}}]\|)\|A\|\frac{t^2}{r^2} + 2\|H_{\text{int}\cap A}\|\|A\|t/r \\ + \|i[H_{\text{out}} + H_{\text{int}\setminus A}, \tilde{A}^{(n+1)}]t/r - \frac{1}{2!}[H_{\text{out}} + H_{\text{int}\setminus A}, [H_{\text{out}} + H_{\text{int}\setminus A}, \tilde{A}^{(n+1)}]]t^2/r^2 + \cdots\|. \tag{B14}$$

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

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

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 [14] such that

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

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

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

Applying (B17) recursively, it is then clear that

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

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 particle guess heuristic 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 particle guess heuristic 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 [16]. 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: Scaling with n

All of the examples considered so far examine compressed QHL 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 compressed QHL 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.

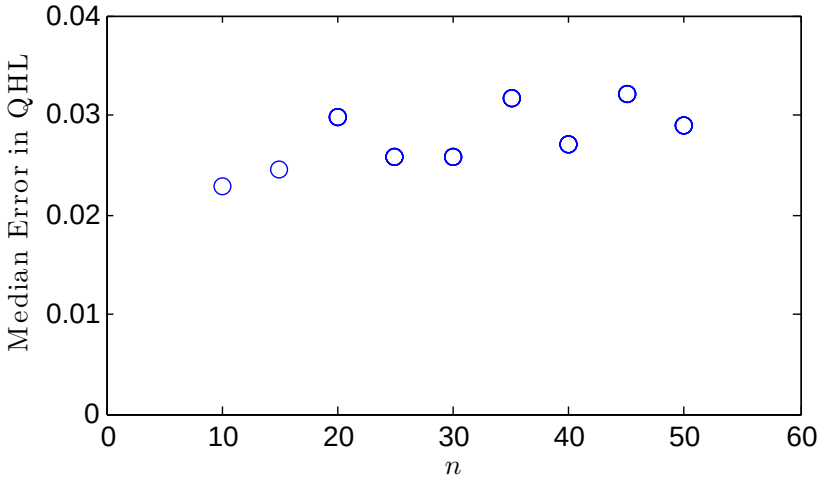


Figure 7: Median error in compressed QHL as a function of the number of qubits in the model.

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 compressed QHL 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 compressed QHL.

Appendix D: Justification of scalings for Ising model

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

$$\sum_{j=\lfloor(w-a)/2\rfloor+1}^{\ell-a} af(j) \leq \sum_{j=\lfloor(w-a)/2\rfloor+1}^{\infty} abe^{-(j-1)\alpha} = \frac{abe^{-\lfloor\frac{w-a}{2}\rfloor\alpha}}{1-e^{-\alpha}}. \tag{D1}$$

This justifies the claim made in the main body.

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

$$\begin{aligned}\sum_{j=\lfloor\frac{w-a}{2}\rfloor+1}^{\ell-a} \frac{ab}{j^\alpha} &= \frac{ab}{(\lfloor\frac{w-a}{2}\rfloor+1)^\alpha}\sum_{k=0}^{\ell-a-\lfloor\frac{w-a}{2}\rfloor-1}\frac{1}{(1+\frac{k}{(w-j)/2})^\alpha}. \\ &\leq \frac{ab}{(\lfloor\frac{w-a}{2}\rfloor+1)^\alpha}\left(1+\int_0^{\ell-a-\lfloor\frac{w-a}{2}\rfloor-1}\frac{1}{(1+\frac{k}{(w-j)/2})^\alpha}\mathrm{d}k\right).\end{aligned} \tag{D2}$$

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 $\ell \to \infty$. Evaluating the integral and some elementary simplifications leads t

$$\sum_{j=\lfloor \frac{w-a}{2} \rfloor + 1}^{\ell - a} \frac{ab}{j^{\alpha}} \leq \frac{ab\alpha}{(\lfloor \frac{w-a}{2} \rfloor + 1)^{\alpha}(\alpha - 1)}. \tag{D3}$$

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 E: Derivation of (19)

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

$$\|((G_1 + \mathcal{E}_1)G_1^{+} - \mathbb{1})\mathcal{H}_k\| \leq (\|G_1 G_1^{+} - \mathbb{1}\| + \|\mathcal{E}_1\|\|G^{+}\|)\|\mathcal{H}_k\|. \tag{E1}$$

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

$$\|(((G_L + \mathcal{E}_L)G_L^{+})(G_{L-1} + \mathcal{E}_{L-1})G_{L-1}^{+} \cdots (G_1 + \mathcal{E}_1)G_1^{+} - \mathbb{1})\mathcal{H}_k\|. \tag{E2}$$

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 (E2) we obtain from the triangle inequality that the error is at most

$$\sum_{j=1}^{L} (\|G_j G_j^{+} - \mathbb{1}\| + \|\mathcal{E}_j\|\|G_j^{+}\|) \prod_{k=1}^{j-1} \left(\|G_k\|\|G_k^{+}\| + \|\mathcal{E}_k\|\|G_k^{+}\|\right). \tag{E3}$$

Noting that the condition number, κ_k, for G_k is $\|G_k\|\|G_k^{+}\|$, (E3) can be upper bounded by the maximum values of of each of the terms involved. If we specifically define $\Gamma_{\max}$ to be the maximum value of $\|G_j G_j^{+} - \mathbb{1}\| + \|\mathcal{E}_j\|\|G_j^{+}\|$ and $\kappa_{\max}$ to be the maximum condition number then

$$\sum_{j=1}^{L} (\|G_j G_j^{+} - \mathbb{1}\| + \|\mathcal{E}_j\|\|G_j^{+}\|) \prod_{k=1}^{j-1} \left(\|G_k\|\|G_k^{+}\| + \|\mathcal{E}_k\|\|G_k^{+}\|\right) \leq L\Gamma_{\max}(1 + (\kappa_{\max} - 1) + \|\mathcal{E}_{\max}\|\|G_{\max}^{+}\|)^{L-1}. \tag{E4}$$

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 $G_j G^{+} - \mathbb{1}$ 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.