Quantum Computation of Molecular Structure Using Data from Challenging-To-Classically-Simulate Nuclear Magnetic Resonance Experiments

Thomas E. O’Brien,1,* Lev B. Ioffe,1,† Yuan Su,1 David Fushman,2 Hartmut Neven,1 Ryan Babbush,1,‡ and Vadim Smelyanskiy1,§

1 Google Quantum AI, Venice, California 90291, USA
2 Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, USA

*teobrien@google.com
†ioffel@google.com
‡babbush@google.com
§smelyan@google.com

We propose a quantum algorithm for inferring the molecular nuclear spin Hamiltonian from time-resolved measurements of spin-spin correlators, which can be obtained via nuclear magnetic resonance (NMR). We focus on learning the anisotropic dipolar term of the Hamiltonian, which generates dynamics that are challenging to classically simulate in some contexts. We demonstrate the ability to directly estimate the Jacobian and Hessian of the corresponding learning problem on a quantum computer, allowing us to learn the Hamiltonian parameters. We develop algorithms for performing this computation on both noisy near-term and future fault-tolerant quantum computers. We argue that the former is promising as an early beyond-classical quantum application since it only requires evolution of a local spin Hamiltonian. We investigate the example of a protein (ubiquitin) confined on a membrane as a benchmark of our method. We isolate small spin clusters, demonstrate the convergence of our learning algorithm on one such example, and then investigate the learnability of these clusters as we cross the ergodic to nonergodic phase transition by suppressing the dipolar interaction. We see a clear correspondence between a drop in the multifractal dimension measured across many-body eigenstates of these clusters, and a transition in the structure of the Hessian of the learning cost function (from degenerate to learnable). Our hope is that such quantum computations might enable the interpretation and development of new NMR techniques for analyzing molecular structure.

DOI: 10.1103/PRXQuantum.3.030345

1. INTRODUCTION

Quantum computing researchers are struggling to find near-term, “beyond-classical” applications of quantum computers: problems whose solution has scientific or commercial value but that cannot be solved on classical devices alone. Fault-tolerant (FT) quantum computers able to solve valuable beyond-classical problems in chemistry [1–4] and materials science [5–7] are predicted to be years away, leaving us in the noisy intermediate-scale quantum (NISQ) era [8]. And, while initial quantum experiments beyond classical [9,10] or on the beyond-classical boundary [11,12] have proven to be of great interest, it has been difficult to extend these to solve practical problems in other fields. This is due to the large error rates on quantum computers, and to the overhead that comes from mapping non-native problems onto a quantum computer. Here, non-native means problems beyond studying statics and dynamics of local spin systems, such as fermionic quantum simulation, linear algebra, or optimization. Algorithms to solve these problems incur costs in circuit compilation [13–16] and measurement [17–20], creating a gap in hardware requirements before these can achieve beyond-classical results. Before this gap is crossed, it makes sense to focus on those practical applications that do not require this overhead.

Nuclear magnetic resonance (NMR) spectroscopy has been a lauded cornerstone of analytical and organic chemistry since its discovery 80 years ago [21,22], being applicable to any molecule or material containing atoms with nonzero nuclear spin. In a NMR experiment, these spinful nuclei are excited by a radio-frequency pulse and allowed to evolve for some period of time under a typically strong (multiple tesla) magnetic field, yielding a small time-dependent response in said magnetic field. This response, or free induction decay, contains information...
about its generating nuclear spin Hamiltonian, which itself contains information about the chemical structure of the molecule or material under probe. The nuclear spin Hamiltonian is typically a strongly interacting quantum Hamiltonian due to its strong dipolar coupling (making it one of the earliest suggested candidates for a quantum computer [23–27]). From accurate knowledge of the dipolar couplings between different spins, one can infer the real-space molecular structure. However, when molecules are free to quickly rotate (e.g., when in solution), this interaction averages out, leaving only chemical shifts (local fields), a weak electron-mediated Heisenberg coupling term [27], and residual incoherent (classical) processes. Such Hamiltonians can be easily classically analyzed (sometimes even by intuition), one of the reasons why NMR has shown such success to this day. However, NMR spectra are considerably more complex for systems that are not free to tumble in all directions (as they are in a solution [28]): e.g., solid-state materials [29,30], molecules in gels [31], stuck on surfaces [32], or in membranes [33,34]. Spectrum prediction can also pose a challenge for experiments operating in low magnetic field such as “zero-field NMR” experiments [35–39], which promise more affordable and practical NMR technologies not requiring huge magnetic fields. These systems present a region of parameter space where data cannot be analyzed by classical computers, yielding a potential area for beyond-classical quantum computation. This ability to analyze classically intractable data sets could enable new types of NMR techniques to characterize previously difficult-to-characterize systems.

Quantum dynamics are generated by a system’s Hamiltonian; from a sufficient set of experimental data, it should be possible to learn the Hamiltonian that generated it. Hamiltonian learning is the inverse problem to predicting experimental outcomes given a system’s Hamiltonian [40], and is well established in the field of quantum information as an approach to device characterization. Sufficiently small devices may be characterized via classical post-processing of experimental data via Bayesian [41,42] or machine-learning methods [43–45]. Exact classical methods are intractable in large systems whenever the forward problem becomes beyond classical, but this can be avoided when possible by careful experiment design. For example, given sufficient control, one can dynamically decouple a small subsystem from its environment, even in the presence of large background magnetic fields, and then learn the global structure piece by piece [46–48]. It is also possible to learn Hamiltonians from expectation values of thermal or long-time average states, which are more easy to classically approximate [49,50]. These methods also allow one to learn a Hamiltonian from highly accurate measurements at ultrashort periods of time, where $e^{iHt} \simeq 1 + iHt$ is a nearly exact approximation. However, when none of the above methods are viable, the Hamiltonian learning problem becomes classically challenging, giving a potential beyond-classical quantum computing application. This still requires that the experiment yield sufficient data that the Hamiltonian can be inferred.

If a quantum system is chaotic, or ergodic, following some perturbation its state will explore its entire Hilbert space, showing little dependence on the precise Hamiltonian parameters and washing out long-time correlation functions [51]. This suggests that the criteria for a system to be learnable is that it correspond to the breakdown of ergodicity. This breakdown has been well studied in many-body physics for many years, most famously in the case of many-body localization [52,53], and corresponds to many interesting phenomena such as area law entanglement scaling [54,55] and the emergence of fractal many-body wave functions [56–60]. To the best of our knowledge, little has been done to tie these fundamental physics concepts to the notion of learnability of a quantum system.

When considering quantum Hamiltonian learning as a beyond-classical experiment, there are actually two classically intractable quantum computations being performed. The first is the Hamiltonian evolution itself (e.g., NMR experiment), which is an analog experiment being performed by the spectrometer and the sample. This produces a set of data that is beyond classical whenever the experiment is beyond classical. (To be beyond classical does not require an experiment to be hard for the complexity class BQP [Bounded-error Quantum Polynomial-time]; many of the experiments considered in this work would lie within the complexity class DQC-1 [Deterministic Quantum Computing with 1 clean qubit] to simulate [46,61].) A key part of this work lies in identifying those NMR experiments that are classically challenging to simulate, which is one quality distinguishing our work from previous suggestions to study classically tractable NMR signals on a quantum device [62]. The second classically intractable quantum computation is to learn the quantum Hamiltonian from the experimental data, which can be executed on a digital quantum computer. We assume that, for the foreseeable future, the connection between the spectrometer and the computer are classical. This prevents the quantum Hamiltonian learning technique of Refs. [63–65] or algorithms that require access to the quantum state [66,67] being implemented. (Access to a quantum connection between spectrometer and computer would be of immense interest if it could be achieved, as it could provide additional exponential speedups [67].) Alternatively, one can approach this problem by generating classically hard spectra (from a quantum simulation or NMR experiment) and using this to train a classical machine agent to infer Hamiltonians from spectra in the same phase (in a manner similar to Ref. [68]). However, this requires access to this additional training data.

Given a sample and a NMR spectrometer, chemists have many techniques at their disposal to infer molecular
In this paper we propose a protocol to learn the nuclear spin Hamiltonian of a molecular or material system using a digital quantum computer and time-resolved measurements from a NMR spectroscopy experiment. We expect this protocol to present a beyond-classical quantum computation (in lieu of classical access to additional data such as a training set for a machine-learning algorithm) when the dataset from the NMR experiment is hard to classically simulate. In Sec. II, we design quantum algorithms to estimate the cost function, Jacobian, and Hessian of the learning problem. We attempt to identify those systems and situations where a beyond-classical application can be found, following some general discussions of learnability in Sec. III. In Sec. IV, we describe and cost circuits to implement our quantum algorithms in both fault-tolerant and NISQ cost models. We find that in both cost models the effect of integration error can be logarithmically suppressed or better, and that the ability to run deep coherent circuits in FT yields polynomial speedups in terms of various problem parameters: the duration and error in the experiment to be simulated, and the desired error in the final gradient itself. In Sec. V, we identify NMR spectroscopy of proteins in or on membranes as one potential application, as the physical pinning of these systems to the membrane prevents the tumbling that would wash away strong correlations in solution. In Sec. VA, we study an example protein, ubiquitin, as a benchmark with known molecular structure. We identify sets of clusters of $^1$H spins within this molecule with strong intracluster coupling and weak coupling to the environment that should produce a strongly coupled signature able to be studied by a quantum device. We calculate the multifractal dimension of small clusters to study their ergodic to nonergodic phase transition as the dipolar term is suppressed (e.g., by magic-angle spinning or decoupling pulse schemes), and find the “quantum-feasible” region to require a suppression factor between around $\alpha = 5$ and $\alpha = 100$ (assuming a background magnetic field of 23.5 T, corresponding to a proton frequency of 1 GHz). This gives a large window within which quantum computers could be expected to assist in NMR interpretation. We demonstrate the application of our learning algorithm to small clusters within this region, demonstrating its convergence on a small spin cluster in the presence of sampling noise. Finally, we show a direct correspondence between the loss of ergodicity (as measured in the multifractal dimension of ubiquitin spin clusters as their dipolar coupling is suppressed) and the onset of learnability (as measured by the analytical Hessian of our learning problem at the global minimum). To the best of our knowledge, this is the first demonstration of a clear connection between the notion of fractal eigenstates in a quantum system and its learnability by quantum or classical means.

II. QUANTUM-ASSISTED HAMILTONIAN LEARNING

We now consider the problem of learning a Hamiltonian $H$ of some system from a set of time-resolved experimental data $S_x(t)$, where $x$ indexes different sets of experiments. As the Hamiltonian of a system dictates the time dynamics, this is a natural thing to learn from time series data; we discuss later how one can infer molecular structure from a Hamiltonian of nuclear spins. Each experiment consists of an initial state preparation $\rho_s$, time evolution by $H$ plus an external time-dependent driving field $H_d(t)$, and final measurement of some observable $O_x$. The signal $S_x(t)$ is then given by

$$S_x(t) = \text{Trace}[U_s(t,0)\rho_sU^\dagger_s(t,0)O_x],$$

where $U_s(t_2,t_1)$ is the time evolution operator generated by the Hamiltonian $H + H_d(t)$ from $t = t_1$ to $t_2$, i.e.,

$$U_s(t_2,t_1) = \mathcal{T} \exp \left\{ i \int_{t_1}^{t_2} [H + H_d(t)] dt \right\},$$

where $\mathcal{T}$ is the time-ordering operator. Note that one may consider $\rho_s = \rho$ and $O_x = O$ independent of the experiment $x$ by encoding preparation and measurement terms onto the driving Hamiltonian $H_d(t)$. (As we discuss later, this is often an accurate description of a real-world NMR experiment.) Alternatively, if the driving Hamiltonian $H_d(t)$ is only used for preparation and measurement, this may be encoded entirely in $\rho_s$ and $O_x$, setting $H_d(t) = 0$ and $\rho_s(t_2,t_1) = U(t_2,t_1) = e^{iH(t_2-t_1)}$.

To define our learning problem, we write our system Hamiltonian in the form

$$H = \sum_n h_n V_n,$$

where the $h_n$ are a set of parameters and $V_n$ a set of Hermitian operators. We then consider the case where some or all of the $h_n$ are unknown, and we wish to estimate these by a set $\{\tilde{h}_n\}$; this defines our learning problem. Given some prior $|\tilde{h}_n^{(0)}\rangle$ with standard deviation $w_n$, and assuming that each datapoint $S_x(t)$ is drawn from a normally distributed...
experimental population with standard deviation $\sigma^2_{\epsilon}$, the maximum-likelihood estimation of the true parameters can be found by minimizing the cost function

$$C[\hat{H}] = \sum_n \frac{(\hat{h}_n - h_n^{(0)})^2}{2\omega_n^2} + \sum_{x,j} \frac{(\hat{S}_x(t) - S_x(t))^2}{2\sigma_{x,t}^2}, \quad (4)$$

where $\hat{S}_x(t) = \text{Trace}[\hat{U}_x(t,0)\rho_0 \hat{U}^\dagger_x(t,0)O_x]$ is the estimated signal with our estimates of the parameters $\hat{h}_n$. (Throughout this work, we use bars to denote quantities derived from estimated parameters rather than hidden ones.) Though $\hat{S}_x(t)$ cannot be estimated on a classical device, implementing it on a quantum computer simply requires a circuit to simulate the time evolution $\hat{U}_n(t,0)$. However, performing such an optimization gradient-free on a higher-dimensional surface is a costly endeavor. The first key result of this work is to give practical forms for the gradient and Hessian of Eq. (4):

$$\frac{dC[H]}{dh_n} = \frac{1}{w_n^2} \hat{h}_n - h_n^{(0)} + \sum_{x,j} \frac{i}{\sigma_{x,t}^2} [\hat{S}_x(t) - S_x(t)] J^n_x(t), \quad (5)$$

$$\frac{d^2C[H]}{dh_n dh_m} = \frac{\delta_{nm}}{w_n^2} - \sum_{x,j} \frac{1}{\sigma_{x,t}^2} [\hat{J}^n_x(t)] J^m_x(t) + 2[\hat{S}_x(t) - S_x(t)] \hat{K}^{mn}_x(t). \quad (6)$$

Here we define

$$\hat{J}^n_x(t) = \int_0^t ds \hat{J}^n_x(t,s), \quad (7)$$

$$\hat{J}^n_x(t,s) = \text{Trace}[\hat{O}_x [\hat{V}_n(t,s), \rho_x(t)]], \quad (8)$$

$$\hat{K}^{nm}_x(t,s,r) = \int_0^t ds \int_0^s dr \hat{K}^{nm}_{x}(t,s,r), \quad (9)$$

$$\hat{K}^{nm}_x(t,s,r) = \text{Trace}[\hat{O}_x [\hat{K}_{x}^{nm}(t,s,r), \hat{V}_m(t,s), \rho_x(t)]], \quad (10)$$

where $\hat{V}_n(t,s) = \hat{U}_n(t,s) V_n \hat{U}_n^\dagger(t,s)$ is the (estimated) operator $V_n$ evolved forwards in time from $s$ to $t$, and $\rho_x(t) = \hat{U}_x(t,0)\rho_0 \hat{U}^\dagger_x(t,0)$ is the (estimated) state $\rho$ at time $t$. As the term dependent on $\hat{K}^{nm}_{x}$ in Eq. (6) disappears in the limit $\hat{S}_x(t) \rightarrow S_x(t)$, it may be practical when near the global minimum of $C[H]$ to approximate

$$\frac{d^2C[H]}{dh_n dh_m} = \frac{\delta_{nm}}{w_n^2} - \sum_{x,j} \frac{\hat{J}^n_x(t) \hat{J}^m_x(t)}{\sigma_{x,t}^2}, \quad (11)$$

which may be obtained at no extra cost to the gradient estimation [assuming that $\hat{S}_x(t)$ and $\hat{J}^n_x(t)$ are measured to the same relative precision]. This is important, as we can approximate the covariance matrix $\Sigma$ of our final estimation of the $\{\hat{h}_n\}$ as

$$\Sigma = [\nabla^2_h C]^{-1}. \quad (12)$$

These equations may be alternatively derived via optimal control theory, which yields a conjugate field to the state $\rho_x(t)$ that is generated by deviations $\hat{S}_x(t) - S_x(t) \neq 0$ and propagates backwards in time via the Schrödinger equation (see Appendix C for details).

### III. QUANTUM AND CLASSICAL LEARNABILITY

In the following sections, we describe how the procedure outlined in Sec. II may be developed into a complete quantum-assisted algorithm to learn Hamiltonian parameters via gradient optimization. However, there are two issues that limit the usefulness of our proposal to implement these algorithms on a quantum computer. We summarize these issues and give a sketch of the region in parameter space where the techniques we develop in this work are relevant in Fig. 1.

The first issue is that the experimental data taken may not contain enough information to learn the desired couplings at all. A large amount of nature is chaotic, or ergodic—where the system tends to approximately explore its entire phase space. In such ergodic systems, properties such as local correlation functions die off quickly without any dependence on the internal structure [11,51]. When this is the case, the set of, e.g., local spin-spin correlations generated by two different Hamiltonians may be indistinguishable up to corrections smaller than any experimental noise. This makes learning impossible. Given sufficient local disorder (relative to the strength of spin-spin interactions), systems tend to localize, yielding a nonergodic many-body localized regime characterized by an absence of transport [52]. In a nuclear spin system, a lack of transport corresponds to spin precession without diffusion or decoherence. This is to say that most NMR experiments are performed in the localized phase (yielding sharp spectral lines), though the many-body localized ergodic phase transition has been recently observed in a NMR experiment [79]. (By contrast, the delocalized phase is associated with either peak broadening or sharp peaks corresponding to delocalized effective spins; see, e.g., Ref. [37].) When spin precession can be observed for long times, or other local correlators do not decay, time-resolved measurements may be used to distinguish between, and thus learn, different Hamiltonians. In certain systems, between the fully localized and ergodic regimes there has been proposed an intermediate nonergodic regime, where the system explores a large fraction of its Hilbert space but does not completely thermalize [56–59]. In other systems, prethermal regions have been proposed, where thermalization occurs at exponentially large timescales, making
FIG. 1. Sketch of the “phases of learnability” of a quantum Hamiltonian. In the red region, the set of experimental data is insufficient to distinguish candidate Hamiltonians. This becomes increasingly likely for systems beyond a localization transition (red dashed line), as after this point experiments such as local correlators tend to provide little information about the system structure. In the blue region, the experimental data is sufficient to learn the system’s structure, but the data processing may be achieved classically, rendering a quantum computer unnecessary. This classical processing may either be achieved by the system being well approximated by a classically computable model, or by the experiment having sufficient control (gray dashed line) to isolate smaller subsystems. At the limit of good control, an experiment has the ability to dynamically decouple individual terms (blue dashed line), after which techniques such as those in Refs. [46,48] can be used to estimate terms individually. A simpler limit is to simply have the ability to address individual (or small frequency regions of) spins in the system (green dashed line), which renders the system more learnable.

these systems local for all practical purposes [80], a phenomenon that has also been observed in NMR [81]. In these situations, some learning should also be possible. The transition to ergodicity and the loss of learnability may be mitigated somewhat given the ability to perform more complicated experiments (e.g., magic-angle spinning [69,70], decoupling pulses [76], or other composite pulse sequences [82]). We summarize these notions going from left to right in Fig. 1. Here, the x axis denotes a rough measure of ergodicity of an arbitrary system (e.g., the localization length left of the localization transition, and the rate of entanglement growth on the right). This can in principle be changed by experimental control (i.e., by suppressing the dipolar interaction strength), which suggests that the localization transition should not be a vertical line.

The second issue we face is that a quantum computer may not be required to interpret the spectrum of a given Hamiltonian: it may be entirely possible to solve the learning problem with a classical device [83–87]. While this reduction in complexity is a good thing for the experiment in question; nevertheless, it limits the utility of using quantum computers in that context. The Hamiltonian learning problem may be solved classically for two reasons. Firstly, the Hamiltonian itself may be classically simulatable, or approximately simulatable. For example, in a many-body localized system, the forward-scattering approximation or other perturbative expansions may be sufficiently accurate for learning. This suggests that a quantum computer will find the most relevance studying either intermediate non-ergodic phases that are not completely localized, or the region in the proximity of a direct many-body localization transition where the localization length is too large for classical simulation. The second reason why the Hamiltonian learning problem may be solved classically is if the experiment is controllable enough to isolate smaller subsystems or otherwise simplify the system. This can, for instance, be achieved if one has the ability to spatially resolve individual spins with a magnetic field. In this case it is possible to apply dynamic decoupling pulse sequences that isolate local Hamiltonian terms while canceling out the remainder in an experiment [46–48], making local characterization possible. We summarize these notions along the y axis of Fig. 1; as one gains more control over an experiment, it becomes possible to learn Hamiltonians in more systems classically, until the barrier of dynamic decoupling is reached and all Hamiltonians are classically learnable.

These two concerns leave our quantum Hamiltonian learning algorithm with only a “goldilocks” zone of applicability. We are interested in those experiments where we have some control over our input state and Hamiltonian, but not those where enough control is available to isolate individual terms. We are also interested in those experiments where our system is somewhat delocalized, but not completely. Experimentally, this can be summarized by saying that we can study those systems where some signal can be extracted, but where that signal is complicated by features (e.g., spectral line shifts) that cannot be easily understood perturbatively.

A. Robustness of learning

Assuming access to the derivatives in Sec. II, one may in principle solve the Hamiltonian learning problem using many well-known gradient-based or Hessian-based optimization techniques. [Using the approximate Hessian in Eq. (11) for minimization results in the well-known Levenberg-Marquardt algorithm [88,89].] One may ask whether this can be made robust, to avoid being stuck in local minima. Similar questions have been asked and answered previously for single parameter estimation [46], quantum phase estimation [90], and device calibration [91,92]. We give a sketch of an argument here for the
robustness of our algorithm given a time-independent Hamiltonian \([H_n(t) = 0]\) that works under the assumption that we start with an initial guess \(h_n^{(0)}\) sufficiently close to our true parameters \(h_n\). We stress that this is not a proof, and examining the landscape around the global minimum of our learning problem is a clear target for future work.

In this case, we may work in the eigenbasis \(|\xi_a\rangle\) of the system Hamiltonian \(H|\xi_a\rangle = E_a|\xi_a\rangle\). Inserting two resolutions of the identity, our signal then takes the form 

\[
S_n(t) = \sum_{a,b} s_n^{a,b}(t),
\]

where

\[
s_n^{a,b}(t) = \langle \xi_a|\rho_t|\xi_b\rangle \langle \xi_b|O_x|\xi_a\rangle e^{i(E_{a} - E_{b})t}.
\]

If our estimate deviates by some parameter \(h_n \rightarrow \tilde{h}_n = h_n + \delta\) then to lowest order in perturbation theory our estimated signal takes the form

\[
\tilde{S}_n(t) = \sum_{a,b} \tilde{s}_n^{a,b}(t) + \delta X,
\]

\[
\tilde{s}_n^{a,b}(t) = s_n^{a,b}(t)e^{i\delta \langle \xi_a|V_n|\xi_b\rangle - \langle \xi_b|V_n|\xi_a\rangle}.
\]

\[
X = \sum_{a,b,c\neq c} e^{-iE_{a}t} \left( e^{iE_{a}t} - e^{iE_{b}t} \right) E_a - E_c
\]

\[
\times \langle \xi_c|\rho_t|\xi_b\rangle \langle \xi_b|O_x|\xi_c\rangle \langle \xi_a|V_n|\xi_c\rangle
\]

\[
+ \sum_{a,b,c\neq c} e^{iE_{a}t} \left( e^{-iE_{b}t} - e^{-iE_{c}t} \right) \frac{E_b - E_c}{E_b - E_c}
\]

\[
\times \langle \xi_c|\rho_t|\xi_b\rangle \langle \xi_b|V_n|\xi_c\rangle \langle \xi_b|O_x|\xi_a\rangle.
\]

Note here that \(X\) is independent of \(\delta\). The second term in our cost function [Eq. (4)] takes the form

\[
\sum_{x,t} \frac{(\tilde{S}_n(t) - S_n(t))^2}{2\sigma_{x,t}^2} = \sum_{x,t} \frac{1}{\sigma_{x,t}^2} \left[ \delta X + \sum_{a,b} \tilde{s}_n^{a,b}(t) \right]^2
\]

\[
\times \left[ 1 - e^{i\delta \langle \xi_a|V_n|\xi_b\rangle - \langle \xi_b|V_n|\xi_a\rangle} \right]^2.
\]

This oscillates as a function of \(\delta\) with a frequency bounded by \(4t \max\langle \xi_a|V_n|\xi_b\rangle\), which is independent of the system size. This implies that we know that local minima in our parameter space must be separated by at least \(4t \max\langle \xi_a|V_n|\xi_b\rangle^{-1} \leq [4t \max\langle |V_n|\rangle^{-1}].\) Flipping this around, let us suppose that we know that our initial guess \(h_n^{(0)}\) of our parameters \(h_n\) lies within some \(\delta \leq \sum_n |h_n - h_n^{(0)}|\); we can perform robust Hamiltonian learning by first learning \(H\) from only experiments at times

\[
t < t_{\text{max}} = \frac{\pi}{4g \max_n \|V_n\|},
\]

After converging on this data we may estimate the variance of our parameter guess [using Eq. (6)], refine our estimate of \(\delta\), and increase the range of allowed \(t\). Assuming that estimation at each \(t_{\text{max}}\) yields an error \(\delta_n \leq c\pi/(4\max_n \|V_n\|t_{\text{max}})\) for some \(c < 1\), repeating this procedure over multiple orders will converge to some final error \(\epsilon \in O(\log(1/e))\) iterations of this procedure. If our initial uncertainty \(\sum_n |h_n - h_n^{(0)}|\) grows with the system size (which can simply be due to the increase in the number of parameters), this will necessarily shrink the initial choice of \(t_{\text{max}}\) by the same amount. However, for a local Hamiltonian, we expect the number of parameters and the magnitude of terms to only grow linearly in \(N\).

The ability to estimate out to large times allows us to avoid learnability issues in localized systems where small long-range couplings contribute mostly to the off-diagonal part of the Hamiltonian (and thus do not affect the eigenstructure significantly). However, it is of no help in ergodic systems where signals \(S_n(t)\) disappear quickly with \(t\). As we will see, in those (ergodic) systems learning anything would be a significant challenge. It is further limited in a real NMR experiment by decoherence due to interaction with the environment. This gives a dephasing time \(T_2^*\) beyond which all signal quickly decays back to its thermodynamic average. This in turn is a natural cutoff as the longest-possible experiment (and simulation) time worth learning from. If this decoherence time is too short, it will in turn place restrictions on our ability to learn weak couplings to high accuracy.

IV. QUANTUM ALGORITHMS

In this section, we outline algorithms to efficiently calculate our cost function \(C[H]\) and its first and second derivatives using a quantum computer. Quantum algorithm optimization is significantly different when targeting noisy near-term versus fault-tolerant long-term devices; we present algorithms for both situations. In either case, we discretize the integral in Eq. (7) as

\[
\int_0^t dsf(s) \sim \sum_{i=0}^{t-1} z_i f(s_i),
\]

where the weights \(z_i > 0\) are chosen such that \(\sum_i z_i = t\). [A two-dimensional discretization is similarly possible for the integral \(K_{n,m}(t)\).] There are many possible methods for choosing both the weights and the points \(s_i\); one may use a simple trapezoidal or midpoint rule, or more complicated Gaussian quadrature methods, or one may take a Monte Carlo approach and choose the points at random. Each method incurs a discretization error that goes to 0 as \(I \rightarrow \infty\). For both the FT and NISQ quantum methods that we propose, cost depends primarily on \([\sum_i z_i] = t\), and has only at most a logarithmic dependence on \(I\). This allows us to choose our method of integration for simplicity or ease in circuit design.
A. Algorithms for near-term quantum computers

In the current NISQ era we want to run the shortest quantum circuits possible for any application. To this end, we propose estimating the signal \( \bar{S}_x(t) \), and the integrands \( \bar{J}_x(t) \) and \( \bar{K}_{x,m}(t) \) on a quantum computer, and performing the integration and summation to yield \( dC[H]/dh_n \) and \( \bar{d}C[H]/dh_n dh_m \) classically. The signal \( \bar{S}_x(t) \) is already in the form where it can be read as the expectation value of a quantum state following the application of a unitary circuit (Fig. 2, top). This requires that it take the form [Eq. (1)]

\[
\text{Trace}[U \otimes \mathcal{M}] \text{ for a quantum state } \mathcal{E}, \text{ unitary } U, \text{ and hermitian operator } \mathcal{M}. \text{ [This can be compared directly to Eq. (1).]} \]

The integrands \( \bar{J}_x(t) \) [Eq. (8)] and \( \bar{K}_{x,m}(t) \) [Eq. (10)] are not quite of this form. However, the integrands may be put in the correct form by adding control terms to calculate the first and second derivatives of our cost function \( C \).

The desired integrand can be found to be the expectation value of the product of the indicated operators, which (in NISQ) must be read out by repeated preparation and measurement.

\[
\text{Trace}[U \rho U^\dagger O_x] = \sum_{m,n} \langle n | \rho_x | m \rangle \langle m | U^\dagger O_x U | n \rangle
\]

As we know the initial distributions \( \langle n | \rho_x | n \rangle \), the estimations of \( \langle n | U^\dagger O_x U | n \rangle \) may be used to compute the target trace. In practice, we do not need to prepare all states; it suffices to sample from a distribution proportional to \( \langle n | \rho_x | n \rangle \). In state-of-the-art quantum experiments this presents a small difficulty, as uploading a new pulse sequence to reprepare each state may be impractical. One solution may be to initially prepare qubits in the \( |+\rangle \) basis and measure them prior to performing a simulation, which results in a new preparation each time.

Repeating the above procedure at multiple points \( s_i \) and \( r_i \) allows for parallel estimation of \( J_x^n(t) \) or \( K_{x,m}^{n,m}(t) \) via Eq. (19) (for fixed \( n \) and \( m \)). In principle, in NISQ we are free to choose these times at random from the ranges \([0, t]\) and \([0, s]\) and choose the initial state. (In practice, changing evolution times is more difficult than choosing initial states; making this scheme more practical is an important direction for future work.) In this case, each choice of starting state and time produces an independent random variable and Hoeffding’s inequality may be applied. For the estimation of \( J_x^n(t) \), \( M \) repetitions of our experiment to transform a commutator into the above form (with \( Y \otimes O = \mathcal{M}, c-U = \mathcal{U}, \text{ and } |+\rangle\langle+| \otimes \rho = \mathcal{E} \). Here the symbol \( c-U := I \otimes U \) denotes the unitary \( U \) controlled by the control qubit. These circuits require only local control of the \( V_n \) unitary (see below) and uncontrolled time evolution, making them rather NISQ friendly (using, e.g., the randomized Trotterization methods of Ref. [93] to simulate the time evolution). We assume in Fig. 2 that the terms \( V_n \) are unitary (i.e., tensor products of Pauli operators for pairs of spins), but if this is not the case, one may write \( V_n \) as a linear combination of unitary operators, execute the circuits for each unitary component separately, and sum the resulting expectation values to yield the desired result [94].

The circuits in Fig. 2 assume the ability to prepare the initial states \( \rho_x \). In the applications we consider in this paper, these will be mixed diagonal states in the computational basis. The measurement operators \( O_x \) will similarly be diagonal in the computational basis. Preparing mixed states requires that we average over many pure state preparations. Consider the case where we prepare a computational basis state \( |n\rangle \), then perform a circuit \( U \) and measure the expectation values of a set of \( O_x \) in parallel, yielding a set of estimates of \( \langle n | U^\dagger O_x U | n \rangle \). If we have repeated this independently for all computational basis states, and each \( \rho_x \) is diagonal in the computational basis, we can write

\[
\text{Trace}[U \rho_x U^\dagger O_x] = \sum_{m,n} \langle n | \rho_x | m \rangle \langle m | U^\dagger O_x U | n \rangle
\]
yields an estimator \( \overline{J_x}(t) \) that satisfies

\[
P\left( \left| \overline{J_x}(t) - \overline{\tilde{J}_x}(t) \right| > \epsilon \right) \leq 2e^{-M\epsilon^2/2|\Omega_x|^2}\epsilon^2, \tag{22}
\]

and the number of samples required to estimate this at a constant failure rate scales as \( M = \tilde{O}(\epsilon^{-2}|\Omega_x|^2t^2) \) (where we use \( \tilde{O} \) to denote asymptotic complexity suppressing polylogarithmic factors). Similarly, for the estimation of \( \overline{K_x^{m,n}}(t) \), \( M \) repetitions of our experiment yields an estimator \( \overline{K_x^{m,n}}(t) \) that satisfies

\[
P\left( \left| \overline{K_x^{m,n}}(t) - \overline{\tilde{K}_x^{m,n}}(t) \right| > \epsilon \right) \leq 2e^{-M\epsilon^2/2|\Omega_x|^2t^2}, \tag{23}
\]

and the number of samples required to estimate this at a constant failure rate scales as \( M = \tilde{O}(\epsilon^{-2}|\Omega_x|^2t^4) \). This scaling in \( t \) is to be expected, as the integrals tend to scale as \( J_x(t) \sim t \), \( K_x^{m,n}(t) \sim t^2 \). [To see this, note that the diagonal terms of \( \int_0^t ds V_{m,n}(t,s) \) in the Hamiltonian basis grow linearly in \( t \).] This implies that we may estimate \( J_x(t) \) and \( K_x^{m,n}(t) \) to constant relative error with a number of samples independent of \( t \).

We now give an analysis of the complexity of this sampling approach where we assume quantum simulation is performed under a Hamiltonian quantum model. The no-fast-forward theorem \( [97] \) requires a number of queries of \( \Omega(t) \) to execute each of the circuits in Fig. 2, so the total gate counts to estimate \( J_x(t) \) and \( K_x^{m,n}(t) \) to error \( \epsilon \) using these methods scale at best as \( \tilde{O}(\epsilon^{-2}|\Omega_x|^2t^2) \) and \( \tilde{O}(\epsilon^{-2}|\Omega_x|^2t^4) \), respectively. By comparison, the total query count to estimate \( S_x(t) \) to error \( \epsilon \) using the circuit in Fig. 2 is bounded asymptotically as \( \tilde{O}(\epsilon^{-2}|\Omega_x|^2t^2) \). These estimates need to be combined to estimate the derivative in Eq. (5). To compute the total cost to estimate this to constant error, let us assume that \( J_x(t) \sim t \), \( S_x(t) \sim t^2 \), that \( \sigma_x, t \sim \sigma \) and \( |\Omega_x| \sim 1 \) are independent of \( x \) and \( t \), and that we can optimize the number of repetitions of each experiment to minimize the total query count. Let us also assume that each experiment involves preparation and measurement in a commuting basis, and let us assume no covariance between parallel measurements. Then, we find that the total number of queries required to estimate single derivative terms to error \( \epsilon \) is bounded asymptotically by (see Appendix A for details)

\[
\tilde{O}\left( \sigma^{-4}\epsilon^{-2}N_s \left[ \sum_{\text{sampled } t} f^{1/2} \right]^2 \right), \tag{24}
\]

where \( N_s \) is the number of distinct experiments performed. The evaluation of the sum over \( t \) depends on whether \( S_x(t) \) are sampled logarithmically sparsely [in which case \( \sum_t f^{1/2} \sim T^{3/2} \log(T) \), or densely (in which case \( \sum_t f^{1/2} \sim T^{3/2} \), where \( T = \max(t) \). In the former case, the total number of queries is bounded asymptotically by \( \tilde{O}(\sigma^{-4}\epsilon^{-2}N_s T^3) \), while in the latter it is bounded asymptotically by \( \tilde{O}(\sigma^{-4}\epsilon^{-2}N_s T^3) \).

**B. Improved estimation of the gradient term on a fault-tolerant quantum computer**

In a fault-tolerant cost model, it is preferable to perform the integration, multiplication, and summation over \( t \) and \( x \) in the second term of Eq. (5) entirely coherently. This because on an error-corrected quantum computer the key resource to minimize is the total number of gates (i.e., the sum of the gate count of each circuit applied) rather than just the depth of the longest circuit. We now outline how this may be achieved for the gradient term.

Assuming that \( \Delta_x = \frac{W_x}{\sqrt{2}} \langle 0 | 0 \rangle W_x^\dagger \), and using the fact that \( \text{Trace}(A) \cdot \text{Trace}(B) = \text{Trace}(A \otimes B) \), we have

\[
\sum_{x,t} \frac{i}{\sigma_x} \overline{S_x(t)} \overline{J_x^n(t)} \sim \left\{ 0 \otimes 0 \otimes + \left| \sum_{x,t,l} \frac{z_{l,t}}{\sigma_x} \mathcal{U}^0(x, t, s, l) \left| 0 \otimes 0 \otimes + \right. \right. \right\}, \tag{25}
\]

\[
\mathcal{U}^0(x, t, s) = W_x^\dagger \mathcal{U}_x^0(t, 0) O_x U_x(t, 0) W_x \otimes W_x^\dagger \mathcal{U}_x^0(s, 0) |0\rangle \langle 0| - |1\rangle \langle 1| \otimes V_n \times \mathcal{U}_x^0(t, s) O_x U_x(t, s) \langle 0 | 0 \rangle \otimes V_n + |1\rangle \langle 1| U(s, 0) W_x, \tag{26}
\]

where the approximation is the approximation from our numerical integration. One can confirm that \( \mathcal{U}^0(x, t, s) \) is unitary as long as \( V_n \) and \( O_x \) are unitaries (and if this is not the case, they may be decomposed as a linear combination of unitaries themselves). The second part of the second term in Eq. (5) requires multiplying by the experimental signal \( S_x(t) \). This signal then needs to be loaded onto the device; if done naively, this could easily become the dominant cost in our circuit. To lower this cost, we make the reasonable assumption that \( S_x(t) \) consists of a small number \( N_{a_x} \ll T \) of Fourier components

\[
S_x(t) = \sum_{k=1}^{N_{a_x}} a_x(k) \cos(t \omega_x + \phi_x), \tag{27}
\]

where \( \omega_x = 0 \) or \( \omega_x = \pi/2 \) and \( a_x(k) > 0 \) is expected from the \( t = 0 \) behavior of our signal. Then, writing \( \cos(a_x + \phi_x) = \frac{1}{2}(e^{i(a_x+\phi_x)} + e^{-i(a_x+\phi_x)}) \), we have

\[
\sum_{x,t} \frac{i}{\sigma_x} S_x(t) \overline{J_x^n(t)} \sim \left\{ 0 \otimes 0 \otimes + \left| \sum_{x,t,l,k} \frac{z_{l,t}}{\sigma_x} \mathcal{U}^1(x, t, s, l, k) \left| 0 \otimes 0 \otimes + \right. \right. \right\}, \tag{28}
\]
\[ U^1(x,t,s,k) = e^{-i\delta t (000, k^+\phi_1)} W_x^t U_x^t (s,0) \]
\[ \times [ |000\rangle \langle 000| - |100\rangle \langle 100| V_n ] \]
\[ \times U_x^t(t,s) O_s U_x^t(t,s) \]
\[ \times [ |000\rangle \langle 000| \otimes V_n + |100\rangle \langle 100| ] U_x(s,0) W_x, \]

(29)

where we have labeled the operations acting on the different control qubits 0 and 1. Under the above assumptions, \( U^1(x, t, s) \) is also a unitary operator. In the above, the \( s_{ti} \) and \( z_{ti} \) points are our integration points and weights, respectively (following Sec. IV A), but allowing for the fact that the limits of integration (and thus both the points we should sample over and the total width we need to multiply by) are dependent on \( t \). Both summations may then be block encoded as linear combinations of unitaries \([98]\). These require control registers \(|x\rangle\), \(|t\rangle\), \(|l\rangle\), and \(|k\rangle\) to encode the summation variables (as well as some additional registers we introduce later), and SELECT and PREPARE unitaries. (We assume here that our times \(|t\rangle\) have some finite binary representation.) These registers contain in turn \( n_z \sim \log(N_z^l) \), \( n_t \sim \log(KT) \), \( n_l \sim \log(L) \), and \( n_k \sim \log(N_k^l) \) qubits, where \( 1/K \) is the precision to which we store our times \( t \); see Appendix B 2 for a detailed analysis of the truncation and discretization error.

The \( \text{SEL}_0 \) unitary selects the correct \( \tilde{U}^0(x, t, s, t, k) \) unitary to implement based on the control register; in other words,

\[ \text{SEL}_0 = \sum_{s, t, l, k} |x\rangle \langle t| \langle l| \langle |k\rangle \langle l| |x\rangle \langle U^0(x, t, s, t, k). \]

Similarly,

\[ \text{SEL}_1 = \sum_{s, t, l, k} |x\rangle \langle t| \langle l| \langle k| \langle |l\rangle \langle x| \langle U^1(x, t, s, t, k). \]

(30)

(31)

(32)

(33)

In Fig. 3, we show how this can be implemented using oracular access to \( U_x(t, s) \), \( O_s \), and \( W_x \) (which we shortly give implementations for). The \( \text{PREP}_n \) unitaries prepare the corresponding control states

\[ |\Psi^0_x = \frac{1}{\sqrt{\lambda_0}} \sum_{s, t, l} \sqrt{\frac{z_{ti}}{\sigma_{s, t}}} |x\rangle \langle t| \langle l|, \]

\[ \lambda_0 = \sum_{s, t, l} \frac{z_{ti}}{\sigma_{s, t}}, \]

\[ |\Psi^1_x = \frac{1}{\sqrt{\lambda_1}} \sum_{s, t, l, k} \sqrt{\frac{z_{tk}}{\sigma_{s, t}}} |x\rangle \langle t| \langle l| \langle k|, \]

\[ \lambda_1 = \sum_{s, t, l, k} \frac{z_{tk}}{\sigma_{s, t}} ^2, \]

from an initial state \(|0\rangle\) on the control register. [Note that the absolute value in Eqs. (31) and (33) are technically unnecessary as all summands are positive.] We omit garbage registers in these steps for simplicity. Given these, one can check that

\[ \langle 0| \text{PREP}^+_0 \text{SEL}_0 \text{PREP}_0 |0\rangle \sim \frac{1}{\lambda_0} \sum_{s, t} i \bar{S}_x(t) \bar{S}_y(t), \]

(34)

\[ \langle 0| \text{PREP}^+_1 \text{SEL}_1 \text{PREP}_1 |0\rangle \sim \frac{1}{\lambda_1} \sum_{s, t} i \bar{S}_x(t) \bar{S}_y(t), \]

(35)

where the circuits act on the combined system and control register set. For \( \alpha = 0, 1 \), we may then use the overlap estimation algorithm of Ref. [99] to estimate these values to error \( \epsilon_\alpha \) with confidence \( 1 - \delta \) using \( \mathcal{O}(\log(\delta^{-1})\epsilon_\alpha^{-1}) \) queries to \( \text{PREP}_n \), \( \text{SEL}_\alpha \); see Appendix B 1 for a review.
of this algorithm. The number of additional gates used in Ref. [99] is negligible compared to the cost of block encoding. Setting \( e_\alpha = O(\epsilon/\lambda_a) \) allows us to obtain an estimate of \( dC[H]/dH_a \) that is within \( \epsilon \) with confidence \( 1 - \delta \). When we have sampled at time \( t \), \( \sum z_{ij} = t \), so if we assume that \( |S_i(t)| \leq 1 \) and \( \sigma_{ij} = \sigma \), we have \( \lambda_1 = \lambda_2 = (N_\chi/\sigma^2) \sum_{\text{sampled } t} t_1 \). The number of oracle calls to PREPARE and SELECT then scales as [in comparison to Eq. (24)]

\[
\tilde{O} \left( \sigma^{-2} e^{-1} N_\chi \sum_{\text{sampled } t} t \right). \tag{36}
\]

To make a heuristic comparison to the NISQ results, we again consider an oracular model. Our SELECT oracles require time evolution by up to \( T = \max(t) \), so a quantum simulation algorithm with linear scaling in the evolution time would make \( O(T) \) queries to the Hamiltonian oracle. Thus, in the sparse or dense case the total number of oracle calls to \( 1 \) is bounded by \( O(\sigma^{-2} e^{-1} N_\chi T^2) \) (a saving of \( \sigma^{-2} e^{-1} T \)), while in the dense case the total number of oracle calls is \( O(\sigma^{-2} e^{-1} N_\chi T^2) \) (a saving of \( \sigma^{-2} e^{-1} T^2 \)). We expect much larger savings for simulating concrete Hamiltonians using fault-tolerant quantum algorithms, as NISQ approaches [93,94] typically cannot achieve linear scaling in the simulation time and also have worse scaling in the target precision.

The complexity of the SELECT unitaries is dictated by the need to implement the controlled time evolutions \( \sum x_{L} |x \rangle \langle t|\langle l| \langle l|t\rangle \langle x|U_x(s_{L},0) \rangle \) and \( \sum x_{L} |x \rangle \langle t|\langle l| \langle l|t\rangle \langle x|U_x(t,s_{L}) \rangle \). We discuss the initial preparation \( \sum_{x} |x \rangle \langle x|W_x \) later. There are a wide range of fast quantum algorithms for simulating time evolution [97,100–103]; which choice is optimal depends on the details of the Hamiltonian being studied. Here, we give an example implementation of the SELECT unitaries using higher-order product formulas, following the analyses developed in Ref. [104]. We expect higher-order formulas to provide the fastest approach for simulating many spin Hamiltonians [104,105], at least asymptotically. We assume for practical purposes here that our time evolution is experiment independent—\( U_x(t,s) = U(t,s) = e^{iH(t-s)} \). This removes the need to consider the \( |x \rangle \) register in our implementation of SELECT. Our implementation requires that we fix the discretization of the integral in Eq. (7). We choose \( L \) points \( s_{L} = tL \), with even weights \( w_i = t/L \). The error in this approximation can be shown to be bounded by \( O(\sigma^2 ||O_x||H,V_n)||/L) \), which is negligible if \( L \gg \sigma^2 ||O_x||H,V_n||/L \); see Appendix B2 for details. The controlled time evolution parts of the SELECT unitary then take the form

\[
c-U(s,0) = \sum_{l=0}^{L} \sum_{t=0}^{T} \langle l|\langle t|\langle l|e^{-i\frac{tH}{L}}. \tag{37}
\]

We implement this by dividing the total time interval \([0,t]\) into \( R \) Trotter steps and implementing a higher-order product formula in each step. To ensure that the simulation has error at most \( \eta \), we take [104]

\[
R = X_{1}/T(X_{2}/T)\eta^{o(1)}, \tag{39}
\]

where the lowercase \( o(1) \) here represents a constant that can be taken to be arbitrarily small, and the \( X_{1} \) and \( X_{2} \) coefficients depend on the system size and graph connectivity. For a linear chain of \( N \) qubits, we have \( X_{1} \sim 1 \) and \( X_{2} \sim N \) [106]. By contrast, assuming a model of clustered Hamiltonians [107]

\[
H = \sum_{\mathcal{L}} \sum_{k,k' \in \mathcal{L}} H_{k,k'} + \sum_{\mathcal{K} \neq \mathcal{L}} \sum_{k,k' \in \mathcal{K}} H_{k,k'}, \tag{40}
\]

where \( H_{k,k'} \ll H_{k,k'} \) whenever \( k,k' \in \mathcal{K} \neq \mathcal{L} \), \( \neq l \), we have \( X_{1} \sim \Lambda_{\text{ind}}, X_{2} \sim \Lambda \), where

\[
\Lambda_{\text{ind}} = \max_{\mathcal{L}} \max_{k \in \mathcal{K}} \sum_{k \in \mathcal{L}} ||H_{k,k}||, \tag{41}
\]

\[
\Lambda = \sum_{\mathcal{K},\mathcal{L}} \sum_{k \in \mathcal{K},k \in \mathcal{L}} ||H_{k,k}||. \tag{42}
\]

Alternatively, we can apply a partial Trotter decomposition without splitting the terms within each cluster [107]. This reduces the Trotter error to instead scale with \( X_{2} \sim \Lambda_{\text{int}} \), where

\[
\Lambda_{\text{int}} = \sum_{\mathcal{K} \neq \mathcal{L}} \sum_{k \in \mathcal{K},k \in \mathcal{L}} ||H_{k,k}|| \ll \Lambda. \tag{43}
\]

Each cluster can then be simulated using either product formulas or more advanced quantum algorithms. We expect that such a hybrid approach can improve the runtime of our approach, but a detailed study of such an improvement is out of the scope of the present paper and will be left as a subject for future work.

To analyze how the error of quantum simulation affects the accuracy of the overlap estimation, we use the block-diagonal structure of Eqs. (37) and (38). We see that the controlled time evolution has error at most \( \eta \) provided that quantum simulation is performed with accuracy \( \eta \). To achieve an accuracy of \( \epsilon \) in the estimate of overlap, we set \( \eta_{\alpha} = O(\epsilon/\lambda_{\alpha}) \) for \( \alpha = 0,1 \). This sets the minimum number of Trotter steps \( R \) in Eq. (39).

We now explain how to add the double control by the \( |l \rangle \) and \( |t \rangle \) registers to a general product formula \( S_p \). The single control by the \( |t \rangle \) register also required for the \( \text{SEL}_{\alpha} \) oracle can be implemented by the following techniques as
The near-linear dependence of the evolution time in $R$ and requirement that $L \gg \mathcal{O}(t^2)$ imply that we need $L \gg R$. This in turn implies that $l/t/L$ is not necessarily an integer multiple of $t/R$. For simplicity, we assume that $L$, $R$ are powers of 2, and write $l/t/L = rt/R + q$ for $0 \leq q < t/R$. We write $q' = qR/t < 1$, and then $lR/L = r + q'$ gives the number of integer $(r)$ Trotter steps and the fractional remainder $q'$. Because $L$ and $R$ are powers of 2, the integers $r$ and $q'$ are already stored in the first $\log(R)$ and the last $\log(L) - \log(R)$ bits of the $l$ register, and may be identified by renaming $l$ as $(r, q')$. The integer part $(r)$ determines the number of times for which $S_p(t/R)$ needs to be applied: controlling $S_p^{2^{by}}(t/R)$ by the $b_i$th bit of the $|r\rangle$ control register and the $|t\rangle$ register implements the unitary

$$
\sum_{r,t} |r\rangle\langle r| |t\rangle |t| S_p(t/R), \quad (44)
$$

where here the $|t\rangle$ register dictates the angle of rotation of each component of the product formula. For example, we would directly implement $\sum_t |t\rangle (t) e^{i\theta_{Z_i}}$, bitwise, using the $b_i$th bit of the $t$ register to control a rotation by $e^{i2^{by}/2^n\theta Z_i}$. This has a gate complexity polylogarithmic in the input parameters, and so we neglect it. We can similarly implement the final fractional Trotter step controlled by the $|q'\rangle$ register; i.e., we implement the unitary

$$
\sum_{q',t} |q'\rangle |t\rangle \langle t| q' S_p(tq'/R). \quad (45)
$$

This also has a similar cost that is polylogarithmic in the input parameters. The final scaling of our doubly controlled time evolution is then identical up to logarithmic factors to the cost of implementing the Trotter evolution without control.

Our above implementation of controlled quantum simulation is developed and optimized specifically for product formulas. Another possible circuit implementation that works for not only product formulas but also more advanced quantum simulation algorithms is to use a binary representation of the evolution time and simulate for time $2^k$ with integer $k$; see Ref. [108] for details. In any case, the complexity only scales logarithmically with the input parameters and the overhead is negligible. This justifies the comparison between the oracular models in Eqs. (36) and (24), as long as the PREPARE and controlled-$W_s$ circuits have lower costs than SELECT.

A naive implementation of the Trotter steps requires that we exponentiate all the terms in the Hamiltonian. For the clustered model in Eq. (40), this implies a gate complexity of $O(N^2)$ to implement each Trotter step. However, this may be improved by truncating Hamiltonian terms with very small magnitudes or by switching to an advanced quantum simulation algorithm. We also need to synthesize the rotation gates with respect to a fault-tolerant gate set, but the overhead in the circuit synthesis is asymptotically negligible.

The PREP0 and PREP1 oracles require loading the coefficients in Eqs. (30) and (32), respectively, onto a quantum register. As we have chosen a uniform integration measure, the integration weights for both PREP0 and PREP1 are independent of the value of the $l$ register, which may be prepared by simply applying a Hadamard gate to all qubits. The remainder of our PREPARE oracles relies heavily on the quantum read-only memory (QROM) and the coherent alias sampling (CAS) technique of Ref. [15], which can be used to perform the mappings $|j\rangle |0\rangle \rightarrow |j\rangle |a_j\rangle$ and $(1/\sqrt{N_d}) \sum_j |j\rangle \rightarrow (1/\sqrt{\sum_j a_j}) \sum_j \sqrt{a_j} |j\rangle$ with $O(N_d)$ Toffoli gates, where $N_d$ is the number of unique data points or indices $j$. This is important as we do not assume that our times $t$ are chosen uniformly, so preparing the $|t\rangle$ register is nontrivial. If we index our times by some uniform index $j$, i.e., writing $t = t_j$, we can use QROM to map $|j\rangle |0\rangle \rightarrow |j\rangle |t_j\rangle$. This requires the $|j\rangle$ register be of size $N_d = \log(N_d)$. Coherent alias sampling allows us to prepare the state $(1/\sqrt{k_1}) \sum_{j,k} (\sqrt{k_1}/\sqrt{a_j}) |j\rangle |t_j\rangle |x\rangle$, with a cost equal to the number of unique datapoints. Combining this with the prepared $|t\rangle$ register above yields the PREP0 oracle. If $a_{j,k} = \sigma_j$ (i.e., all separate experiments are performed with the same error, which is a reasonable assumption), this has an identical cost of $N_d$. We assume $N_d$ scales at worst linearly in $T$ (i.e., for dense sampling), and so the cost of implementing the PREP0 oracle is bounded by $O(T)$ and dominated in the block encoding by the additive cost of the SEL0 oracle. The PREP1 oracle differs from the PREP0 oracle only by the additional amplitudes $a_{j,k}$. These may be mapped onto the device using coherent alias sampling at a cost scaling as $N_t N_w$. This cost is additive to the $N_d$ cost above, and as we expect $N_t N_w \ll N_d$, we expect this oracle to also be dominated by the cost of SEL1.

As an additional part of the SEL1 subroutine, we need to implement the controlled $Z$ rotation $e^{-i\theta_2(\omega_{x,k} + \phi_1)}$. The classical values $\omega_{x,k}$ and $\phi_1$ here need to be loaded onto the quantum device. In order to do this, we rely on the QROM technique of Ref. [15]. Given the set of $N_w$ classical data points $f_{x,k} = 2\pi \omega_{x,k}$ to some fixed precision, QROM can be used to perform the mapping $|k\rangle |x\rangle |0\rangle \rightarrow |k\rangle |x\rangle |f_{x,k}\rangle$, using $O(N_w N_t)$ Toffoli gates. We can similarly map $\phi_1$ onto a single qubit, $|x\rangle |0\rangle \rightarrow |x\rangle |b_x\rangle$, where $b_x = 0$ if $\phi_1 = 0$ and $b_x = 1$ if $\phi_1 = \pi/2$, at a cost of $O(N_t)$ Toffoli gates. These mappings are more appropriate to implement during the PREP1 step, so we shall insert them there. Then, in the SEL1 subroutine, we may assume access to these registers, in which case the controlled-$Z$ rotation may be implemented by arithmetic of the same form as in
product formulas at a cost polylogarithmic in the size of the \(|f_{j,k}\) and |f\) registers. We may alternatively use the phase gradient method described in Ref. [109].

It remains to specify a preparation scheme for \(\rho_x = W_x |0\rangle \langle 0| W_x^\dagger\). This is especially important to consider as \(\rho_x\) is not a pure state, so it is impossible to prepare it from an initial register with a unitary operation on an \(N\)-qubit quantum register. We require \(W_x\) to be a unitary operation for the expectation value estimation algorithm, as it requires repeated access to \(W_x\) and \(W_x^\dagger\) (or equivalently the ability to reflect around \(\rho_x\)). To solve this problem, we expand the size of our system register, and prepare a purified state \(|\psi_x\rangle = W_x |0\rangle\) such that, for all observables \(O\) within our original system, \(\text{Trace}[O\rho_x] = \text{Trace}[O \otimes I |\psi_x\rangle \langle \psi_x|]\). This requires that we at most double the number of qubits of our system \(N\), and for all operations other than \(W_x\) and \(W_x^\dagger\) we can ignore the additional qubits. As we will see below, for applications in NMR, we are mostly interested in preparing states such as

\[
\rho_x = \frac{1}{2} (I + Z_{j_n}),
\]

which is the maximally mixed state on all qubits except qubit \(j_n\). To achieve this with an additional \(N\) qubits, we begin with \(N\) copies of the Bell state \((1/\sqrt{2}) (|00\rangle + |11\rangle) \otimes \overline{N}\), which can be prepared using only Clifford gates. We then perform a Toffoli gate with \(x\) and the \(j, t\)th qubit as controls and \(j_n + N\) as the target, followed by a Hadamard gate on the \(j, t\)th qubit controlled by \(x\). Each controlled Hadamard can be implemented using a single Toffoli gate [4, Fig. 17]. This prepares the state

\[
|x\rangle |\psi_x\rangle = \frac{1}{\sqrt{2^{N-1}}} \sum_{j_n} |0_j\rangle |0_{j+N}\rangle \prod_{j \neq j_n} (|0_j\rangle + |1_j\rangle) |1_{j+N}\rangle,
\]

which has our desired properties. More generally, any thermal state of the classical one-dimensional (1D) Ising model can be prepared as a \(2N\)-qubit thermofield double state with perfect fidelity using a depth-\(N/2\) circuit [110].

V. APPLICATION TO NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

We now focus on the application of our quantum learning algorithm to NMR spectroscopy. In a NMR experiment, a sample of a molecule, crystal, or other material is placed in a strong magnetic field, which interacts with the magnetic moment of any spinful nucleus. These also couple to each other through dipole-dipole and electron-mediated interactions, making the full spin Hamiltonian [28]

\[
H = \sum_i \chi_i S_i \cdot B + \sum_{i \neq j} J_{ij} S_i \cdot S_j
\]

\[
+ \sum_{i \neq j} [S_i \cdot S_j - \frac{3}{2} |\mathbf{r}_{ij}|^2 (S_i \cdot \mathbf{r}_{ij}) (S_j \cdot \mathbf{r}_{ij})],
\]

Here, \(\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j\) is the vector between the \(i\)th and \(j\)th nuclei, \(S_i\) is the spin vector, \(B\) is the external magnetic field, \(\chi_i = \gamma_i \delta_i\) is the shielded magnetogyric ratio of the \(i\)th nucleus (with \(\delta_i\) the chemical shift and \(\gamma_i, the unshielded magnetogyric ratio), \(\Gamma_{ij} = \mu_0 \gamma_i \gamma_j h/8\pi |\mathbf{r}_{ij}|^3\) is the dipole-dipole interaction strength (with \(\mu_0\) the vacuum permeability and \(h\) Planck’s constant), and \(J_{ij}\) is the electron-mediated coupling strength. Assuming that our nuclei are spin-1/2, \(S_i = \frac{1}{2} (X_i, Y_i, Z_i)\). (Simulation of nuclei with spin > 1/2 is also possible on a quantum computer with only small overhead [111].) As the coupling constants \(\Gamma_{ij}\) depend directly on the physical distance \(|\mathbf{r}_{ij}|\) between the molecular spins, knowledge of these distances is sufficient to infer the molecular geometry (modulo global translations, rotations, and reflections) [112]. A NMR spectroscopy experiment follows the protocol outlined in Sec. II. The system begins at thermal equilibrium, is perturbed by one or more magnetic field pulses, and has its free-induction decay read out (which is equivalent to measuring \(\sum_i X_i\)). In this formalism the starting state \(\rho_x\) and measurement operator \(O\) are the same for all experiments [which differ in their choice of \(H_x(t)\)]. However, the external perturbation is often chosen to polarize the initial state and flip spins in the end, making the signal \(S_x(t)\) a local spin-spin correlation measurement. (We discuss how to implement this in a strongly correlated system shortly.)

Whether a NMR-relevant spin Hamiltonian lies within the ergodic, classically feasible, or quantum-feasible regimes in Fig. 1 depends on the relative energy scales of the terms in Eq. (48). These are typically

\[
\beta^{-1} \gg B X \gg B |X_i - X_j| \sim \Gamma_{ij} \gg J_{ij},
\]

where \(\beta\) is the inverse temperature of the system, and \(X = \langle 1/N \rangle \sum_i X_i\). Only terms coupled to the dipolar term (\(\Gamma_{ij}\)) and the electron-mediated interaction (\(J_{ij}\)) generate classically challenging dynamics, so one of them must be large for us to lie outside the classically feasible region of Fig. 1. However, in solution a molecule tumbles rapidly, averaging out the dipolar term to 0. In a strong magnetic field, the remaining Hamiltonian can be treated perturbatively and solved classically, rendering our quantum learning algorithm unnecessary. Previous proposals [62] that suggested using quantum computers to learn the structure of molecules in solution suffer from this inapplicability issue. [Note that in zero- or ultralow-field NMR [35,36,38,39] Eq. (49) does not hold; in these experiments \(B X \lesssim J_{ij}\) and
quantum computers may have a role to play in learning structure.

We propose to instead study systems where molecules are out of solution and not free to move. Indeed, NMR is used in a wide range of systems out of solution: gels [31], surfaces [32], proteins in or on membranes [33,34, 113,114], and in the solid state [28]. In these systems the uniform magnetic field interaction term $B \chi$ is still the dominant energy scale, and terms that do not commute with this cancel out, leaving

$$H = B \sum_i \chi_i Z_i + \sum_{i \neq j} J_{ij} S_i \cdot S_j$$

$$+ \sum_{i \neq j} \Gamma_{ij} (3 \cos^2(\phi_{ij}) - 1)[S_i \cdot S_j - 3Z_i Z_j],$$

where $\phi_{ij}$ is the angle between $\mathbf{B}$ and $\mathbf{r}_{ij}$. (This is commonly known as the secular approximation.) This approximation requires $\phi_{ij}$ to be well defined (e.g., by stacking membranes so that all proteins are similarly aligned with the magnetic field). Simulating randomly scattered molecules would require classical averaging over many such alignments, presenting an additional simulation challenge (and in practice broadening out spectral lines). In solid-state NMR experiments one typically removes the dipolar coupling by magic-angle spinning; spinning at a high frequency around an angle $\theta = 54.74^\circ$ to the magnetic field. This spinning suppresses the dipolar term by a factor $(3 \cos^2(\theta) - 1) \sim 0$ as long as the spinning frequency is much higher than the dipolar coupling strength. Combining this with frequency-selective dipolar recoupling [115] has achieved remarkable success in biochemistry (see, e.g., Ref. [116]). However, this is not typically achievable to high precision for proton NMR, where dipolar couplings are typically of the order of 30–40 KHz [117] (by comparison, the highest frequency centrifuges are around 100 KHz [33]). Even when magic-angle spinning is combined with decoupling pulse schemes, significant residual coupling in proton-NMR spectra can be observed [33,118,119]. (Moreover, suppressing the dipolar coupling term removes valuable structural information about a system.) Proton NMR in solutions has achieved great success in biochemistry, but the folding of proteins can be quite different in vitro versus in vivo. For example, a large number of proteins in cell walls and membranes are folded precisely according to this external environment, and lose their shape in solution [33,34]. This suggests that learning the structure of proteins confined in or on membranes via proton NMR is a potentially valuable beyond-classical quantum computing application.

Proposing to study systems with strong dipolar coupling presents a challenge in designing realistic state preparation and measurement schemes. As the temperature is larger than all energy scales in our experiment, our initial state is a thermal state

$$\rho_0 = e^{-\beta H} \approx 1 - \sum_i \beta B \chi_i Z_i,$$

and we must perturb this in order to generate any signal at all. (The approximation here is quite good, as even for a 23.5-T background magnetic field $\beta B \chi \sim 10^{-4}$.) We have two external handles to perturb our system: the ability to apply time-dependent rf pulses to modulate the background magnetic field $B$, and magic-angle spinning [69,70]. The latter is a double-edged sword; we cannot alter the direction of the magnetic field nor the spinning sample on the timescale of our system, so it is only practical to use this to suppress the dipolar coupling by a fixed amount for the entire experiment. However, this may be adjusted during the experiment via dipolar decoupling and recoupling pulse schemes [71,72,75,76,115]. As a simple example, consider the four-step Waugh-Huber-Haeberlen scheme [120,121], which consists of (1) a rest of time $dt$ and a $\pi/2$ pulse around the $x$ axis, (2) a rest of time $dt$ and a $-\pi/2$ pulse around the $y$ axis, (3) a rest of time $2dt$ and a $\pi/2$ pulse around the $y$ axis, and (4) a rest of time $dt$, a $-\pi/2$ pulse around the $x$ axis and a final rest of time $dt$. By performing a Magnus expansion to first order in $dt$ on the above combined unitary scheme, the dipolar term averages to zero. (This is true of any scheme that rotates the $X, Y, Z$ Pauli spins to the $z$ axis of the magnetic field for equal periods of time.) Once the dipolar field is sufficiently suppressed, individual spins may be flipped by, e.g., applying a low-amplitude magnetic field oscillating at a frequency $\omega$ that addresses spins for which $B \chi_i = \omega$. This simple scheme is likely difficult to achieve sensitivity below about 1 KHz, but this may be improved by frequency-selective pulsing schemes [71,122] such as the DANTE (delays alternating with nutations for tailored excitation) scheme [123]. After applying such a scheme, the perturbation to the system is roughly

$$\rho - \rho_0 = \beta B \sum_{|\chi_i B - \omega| < \delta \omega} \chi_i Z_i,$$

where $\delta \omega$ gives the accuracy of the technique. In a NMR experiment the signal is rescaled by the factor $\beta$ in Eq. (52). However, in a quantum computation we may divide by $\beta$ before performing our estimation, which makes the error requirements in our final signal independent of $\beta$. Approximations in the above can be accounted for in our learning scheme by adjusting the starting state, or by incorporating the pulse sequence into the unitary $U_x$, giving an additional advantage over classical Hamiltonian inference.
A. Determining the structure of ubiquitin as an example application

A significant body of literature on protein structures already exists, which we can use as a benchmark for our quantum learning algorithm. The protein ubiquitin, named for its abundance throughout eukaryotic organisms, was discovered in 1975 [124,125]. The structure of ubiquitin is well known, making it a good benchmark. The protein contains over 600 protons, but these tend to cluster (in terms of their relative coupling strengths, which correspond to their location in space); we propose to divide the full molecule into smaller clusters that may be studied individually. This corresponds to assuming our experimental signature

$$S_x(t) = \sum_{\text{cluster } c} S_x^c(t),$$  \hspace{1cm} (53)

and we can either subtract the contribution of individual pieces from the total signal, or learn this linear combination in a single step. The latter has a linear overhead in the number of pieces, and does not require larger quantum computers or longer circuits. In practice, we do not need to determine the clustering ahead of time; clusters will present themselves as minibands in the spectrum that cannot be separated as in Eq. (53), and we may account for these in our learning algorithm by allowing our optimization to adjust the number of spins in any given band.

To investigate this clustering, we write the ubiquitin molecule as a graph (with edges weighted by the dipolar interaction strength). We take the atomic coordinates of ubiquitin in solution from previous NMR data (protein data bank ID 1D3Z), and proton chemical shifts from the Biological Magnetic Resonance Data Bank (BMRB), entry 17769 (both in turn taken from Ref. [126]). For simplicity, we assume that all magnetically active nuclei in our protein are $^1$H, and that all other magnetically active nuclear species are shifted far enough in frequency from the $^1$H nuclei that their spin dynamics can be neglected. (This is a simplification of the system in question, and would likely need to be corrected for accurate simulation.) We then define a cluster as any connected subgraph where all edges are higher weight than any edges (in the larger graph) that point from the subgraph out. (The above definition in principle removes some edges from the subgraph that correspond to couplings in the cluster. However, upon identification of the cluster we consider all couplings between the spins regardless of their size.) Such subgraphs may be found by thresholding; setting a truncation threshold $V_{\text{min}}$ and eliminating all couplings lower than this threshold in the nuclear spin Hamiltonian cuts the graph of spins into a set of disconnected subgraphs, each of which is a cluster. In the ubiquitin molecule (Fig. 4), by adjusting this $V_{\text{min}}$ we can first split off a 466-spin central core of the molecule (left), and then a 238-spin backbone (middle top), and finally a 60-spin cluster (middle bottom). Given its size, we expect the problem of learning the 60-spin cluster Hamiltonian to lie around the beyond-classical boundary. To investigate the cluster’s coupling to its environment, we plot the distribution of coupling strengths (Fig. 4, right) both within the full molecule, within the 60-spin cluster, and between the cluster and the rest of the molecule. We see that the mean coupling within the

![Image of ubiquitin molecular structure and coupling distribution](image-url)
ubiquitin cluster is around 2 to 3 times all couplings to the environment, and the majority of couplings to the environment are more than 10 times smaller than the couplings within the cluster. We believe that this is sufficiently weak that these couplings may be treated perturbatively, though verifying this is a clear task for future study.

We now demonstrate our Hamiltonian learning algorithm for a small cluster of six spins in the ubiquitin molecule (Fig. 5). We assume that we have access to magic-angle spinning or dipolar decoupling techniques to suppress our dipolar field by a factor \( \alpha = 10 \), and that the background field is 23.5 T. To simulate the proposal that we know the α spins, we take \( P = (N - 1)/2 \) filling.] In an ergodic system where nearly all computational basis states contribute nearly equally to each eigenstate, the participation entropy scales as \( S_1 \sim -\alpha \psi_t^2 \). For the systems considered, this is roughly \( S_1^{(\text{ergodic})} \sim -0.63N \). By comparison, a completely localized system has a constant (or logarithmically growing) participation entropy. As we increase \( \alpha \) and the system becomes nonergodic, the participation entropy follows a trend \( S_1 \sim D_1 S_1^{(\text{ergodic})} \), where the multifractal dimension \( D_1 = D_1(\alpha) \) characterizes the fraction of the Hilbert space explored by an eigenstate \([60]\).

In Fig. 6 we plot the multifractal dimension of our 127-spin clusters as we suppress the dipolar term by a factor \( \alpha \). Our quantum-feasible region corresponds to a multifractal dimension \( D_1 < 1 \), while our classically feasible region corresponds to a multifractal dimension \( D_1 \ll 1 \). We see a clear region between suppression factors of \( \alpha \sim 5 \) and \( \alpha \sim 100 \) where our system begins to localize and the multifractal dimension quickly drops, but the system is not completely localized and classically simple. In the thermodynamic limit this transition is discontinuous [60], but as we are interested in finite system sizes, we believe that the observed trend of \( D_1 \) as we cross this localization transition is relevant to our situation. On either side of the localization transition, local disorder will make individual clusters either more ergodic or more local than the mean, which implies that the boundaries between degenerate, quantum feasible, and classically feasible are not sharp as a function of the dipolar suppression. (This can be seen in the insets of Fig. 6.)
The ergodic to nonergodic phase transition observable in the Hamiltonian eigenstructure maps immediately to the learnability of the NMR spin Hamiltonian. This can be studied in the Hessian of the Hamiltonian learning problem at complete convergence, given by Eq. (11). As the Hessian corresponds to the Fisher information of the system, small eigenvalue-eigenvector pairs \((\lambda, \vec{v})\) correspond to “floppy modes” in our parameter space; linear combinations of parameters that may be adjusted in tandem without significantly altering our signal. Large eigenvalue-eigenvector pairs correspond to combinations of parameters that are well learned. In Fig. 7 we study the typical eigenstructure of the Hessians of clusters of 5–8 spins. We see a clear transition that corresponds exactly to the ergodic to nonergodic phase transition identified in Fig. 6. On the left-hand side, corresponding to the ergodic phase, the typical maximum eigenvalue (Fig. 7, top) of the ensemble shows a clear exponential decay in system size, implying that learning in a large system will be nigh impossible. Moreover, the typical participation in these systems,

\[
\exp \left( - \sum_j v_j^2 \log v_j^2 \right),
\]

(55)
grows quickly (Fig. 7, bottom), implying that these modes correspond to global data rather than specific couplings. By contrast, when the system is strongly localized, the largest eigenvalues are roughly constant in the system size, and correspond to linear combinations of only one or two parameters. As the system shifts between these two phases, we see a continuous improvement in learnability, where it appears we can learn some of but not all of the system. This can be observed in the full eigenspectrum data (Fig. 7, top inset). We note that the largest eigenvalues also correspond to smaller typical participation, which suggests that, when a system is on the ergodicity boundary, we can learn some local couplings rather than just global information. This result demonstrates the importance of having access to the Hessian when solving the learning problem, as it tells which of the converged parameters can be relied upon.

VI. CONCLUSION

In this work we introduced a new method for learning an unknown quantum Hamiltonian of a spin system
from time-resolved measurements of the system. We constructed and costed circuits within NISQ and FT frameworks to estimate gradients of the cost function of this learning problem, finding clear asymptotic speedups when one is not constrained by poor coherence in NISQ devices. We outlined an application for these algorithms in classically intractable NMR experiments, and proposed a specific region in the NMR field (when dipolar couplings are strong and cannot be simply removed) as an area where beyond-classical computations may be very useful. Taking the ubiquitin protein as an example, we identified small clusters of spins in the larger $^1$H dipolar coupling matrix, demonstrated the convergence of our learning algorithm on a toy example, and investigated the cluster-environment coupling and the learnability of the system as the dipolar coupling is suppressed.

As part of this work, we identified a direct correspondence between the ergodic to nonergodic phase transition and the learnability of a Hamiltonian from time-resolved experimental data. The latter was clearly observable in the structure of the Hessian of the learning problem. We believe this to be a more general result than in the NMR problems that we have studied in this work. As far as we know, this is the first link demonstrated between the onset of wave-function fractality and the onset of learnability of the generating Hamiltonian.

Our work opens a new field of quantum computing applications, leaving clear directions for future study. Our learnability data suggests that not all dipolar coupling parameters (that encode the 3D protein structure) may be learnable using the simple spin-spin correlations measured in this work. Developing future experiments to target the parameters relevant for structure calculations in NMR systems will be highly relevant in the future. A related question (which we have not yet determined the answer to) pertains to the conditions (if any) under which this problem is classically difficult. While we can rely on the fact that inference of NMR spectra from strongly correlated problems appears difficult (and the forward problem of generating the spectra is at least DQC1-hard [46,61]), we do not know of a complexity-theory result explaining when the inverse problem is as difficult. Also, the protocol we propose in Sec. III A can, in principle, avoid trapping in local minima; however, this is not optimized or fully costed. We suggest that, following Ref. [46], it may be possible to define a protocol using our methods that learns at the Heisenberg limit (when learning is possible). It is also unclear precisely how well our protocol behaves in a system with a large number of unknown couplings, and whether we will have problems with vanishing gradients as our system size grows. (We note however that this is avoided somewhat in NMR problems where we start with a reasonable guess of our Hamiltonian parameters.) Another clear direction for future work is to consider learning Hamiltonian parameters from the Fourier transform of the spectra $S_x(t)$ instead of the time-resolved data, as spectral information is typically more robust to noise than amplitude data. It also remains to determine what the cost is to optimize a real-world NMR experiment from a realistic initial guess. This understanding can be furthered by small- and midscale quantum experiments (such as recently performed in Ref. [128]) on simulated or experimental hardware. Finally, thanks to the generality of our Hamiltonian learning techniques, it should be possible to extend these methods to the design of new NMR experiments, or to interpret data from other types of experimental procedures used to probe condensed matter, high-energy physics, chemistry, and materials science. We look forward to further explorations of what new possibilities for quantum experiments and data analysis these techniques can bring to the scientific community.

ACKNOWLEDGMENTS

The authors wish to thank Igor Aleiner, Dmitry Abanin, Nicholas Rubin, Jarrod McClean, Jooho Lee, Ashok Ajoy, Joshua Goings, Nathan Wiebe, and William Higgins for useful advice and discussions of this work. D.F. is supported by NIH Grant No. GM065334.

APPENDIX A: OPTIMIZING QUERY COUNTS FOR NISQ ALGORITHMS

In this section we derive Eq. (24) by optimizing the allocation of oracle calls across a set of NISQ experiments. This is a slightly atypical approach; typically, in a NISQ experiment one attempts to optimize the number of experiments (or shots) to achieve some target error $\epsilon$ [17,129] rather than working with an oracular model. However, this optimization allows us to make a fair comparison between the results of Secs. IV A and IV B. Propagating variance through Eq. (5) yields

$$\text{Var}\left[\frac{dC[H]}{dh_a}\right] = \sum_{x,t} \frac{1}{\sigma_{x,t}} \text{Var}[\tilde{S}_x(t)] \tilde{J}_x^n(t)^2$$

$$+ \text{Var}[\tilde{J}_x^n(t)] (\tilde{S}_x(t) - S_x(t))^2. \quad (A1)$$

As our estimate of $\tilde{J}_x^n(t)$ is bounded, assuming that $||O_x|| = 1$ the bound in Eq. (22) yields an estimator for $J_x^n(t)$ with variance $\epsilon^2$ using $M = \bar{T}^2 \epsilon^{-2}$ repetitions of the circuit. We now assume that we may measure $S_x(t)$ and $J_x^n(t)$ for different $x$ in parallel. This is realistic for our NMR application, and will save a factor $N_x$ in the asymptotic scaling. As the number of oracle calls per circuit scales as $t$, we can achieve a variance $\text{Var}[\tilde{J}_x^n(t)] = a_{x,t} t^3 C_{x,t}$ using $C_{x,t}$ oracle calls (for some $t$ and $x$-independent constant $a_{x,t}$). Estimating $\tilde{S}_x(t)$ with a variance $\epsilon^2$ requires simply repeating the corresponding circuit (Fig. 2, top) $\epsilon^{-2}$ times, and so with $C_{x,t}$ oracle calls we achieve a variance.
where we have absorbed the constants of the \( \bar{\mathcal{S}}(t) \) and \( J_0^s(t) \) scalings into \( a_{x,t;S} \) and \( a_{x,t;J} \). To optimize this, we adopt the same Lagrangian approach introduced for measurement optimization in Ref. [129]. We write a Lagrangian

\[
\mathcal{L} = \lambda \left\{ \sum_{x,t} \frac{1}{\sigma^4} \left[ \frac{a_{x,t;S}^3}{C_{t;S}} + \frac{a_{x,t;J}^3}{C_{t;J}} \right] - \text{Var} \left[ \frac{dC[H]}{dh_n} \right] \right\} + \sum_{x,t} (C_{x,t;J} + C_{x,t;S}),
\]

and then differentiate with respect to our free parameters \( C_{x,t;J} \) and \( C_{x,t;S} \) and solve for the result being equal to 0:

\[
\frac{\partial \mathcal{L}}{\partial C_{t;S}} = 1 - \sum_x \frac{\lambda a_{x,t;S}^3}{\sigma^4 C_{t;S}^2} = 0 \quad \Rightarrow \quad C_{t;S} = \lambda^{1/2} \left[ \sum_x a_{x,t;S} \right]^{1/2} t^{3/2} \sigma^{-2},
\]

\[
\frac{\partial \mathcal{L}}{\partial C_{t;J}} = 1 - \sum_x \frac{\lambda a_{x,t;J}^3}{\sigma^4 C_{t;J}^2} = 0 \quad \Rightarrow \quad C_{t;J} = \lambda^{1/2} \left[ \sum_x a_{x,t;J} \right]^{1/2} t^{3/2} \sigma^{-2}.
\]

Substituting into the expression for a variance \( \text{Var}[dC[H]/dh_n] = \epsilon^2 \) yields

\[
\epsilon^2 = \lambda^{-1/2} \sigma^{-2} \sum_t \left\{ \left[ \sum_x a_{x,t;J} \right]^{1/2} t^{3/2} \right\}^2 + \left[ \sum_x a_{x,t;S} \right]^{1/2} t^{3/2},
\]

and rearranging for \( \lambda \) yields

\[
\lambda = \epsilon^4 \sigma^{-4} \sum_t \left\{ \left[ \sum_x a_{x,t;J} \right]^{1/2} t^{3/2} \right\}^2 + \left[ \sum_x a_{x,t;S} \right]^{1/2} t^{3/2}^2.
\]

Finally, we can write the total number of oracle calls as

\[
C = \sum_{x,t} (C_{x,t;J} + C_{x,t;S}) = \epsilon^{-2} \sigma^{-4} \left\{ \sum_t \left[ \left( \sum_x a_{x,t;J} \right]^{1/2} t^{3/2} \right\}^2 + \left[ \sum_x a_{x,t;S} \right]^{1/2} t^{3/2} \right\}^2;
\]

as \( a_{x,t;J} \) and \( a_{x,t;S} \) are constants, we asymptotically have \( \sum_x a_{x,t;J}, \sum_x a_{x,t;S} \sim O(N_x) \). Substituting this into Eq. (A8) immediately yields Eq. (24).

**APPENDIX B: FAULT-TOLENT APPROACH TO THE BACKWARDS PROBLEM**

In this appendix, we expand on the circuitry and cost analysis of the fault-tolerant estimation of the gradient of the Hamiltonian learning problem in Sec. IV B.

We consider estimating the quantity

\[
\sum_{x=1}^{N_x} \sum_{j=1}^{N_j} \frac{1}{\sigma^{2\epsilon} x_j} \left( \text{Trace}(O_x \bar{\rho}_s(t_j)) - \sum_{m=1}^{N_m} a_{x,m} \cos(t_j \omega_{x,m} + \phi_x) \right) \times \int_0^{t_f} ds \text{Trace}(O_x[V_n(t_j, s), \bar{\rho}_s(t_j)])
\]

on a fault-tolerant quantum computer for the backwards problem. Here, operators \( V \) and \( \bar{\rho}_s \) are defined as

\[
V_n(t, s) := e^{-i(t-s)H} V_n e^{i(t-s)H}, \quad \bar{\rho}_s(t) := e^{-iH} |\psi_s \rangle \langle \psi_s| e^{iH}
\]

for some Hamiltonian \( H \), an arbitrary term \( V_n \) in \( H \), and an \( x \)-dependent initial state \( |\psi_s \rangle \).

Our approach depends on a quantum overlap estimation algorithm that we review in Appendix B 1. To implement this on a fault-tolerant quantum computer, we need to discretize the time integral and re-express the target quantity as a linear combination of quantum overlaps. We also need to truncate real parameters \( \sigma_{x,t;J}, t_s, a_{x,m}, \) and \( \omega_{x,m} \) to a finite number of digits to construct the quantum circuit. We analyze the truncation and discretization error in Appendix B 2.

1. Quantum overlap estimation

To estimate Eq. (B1) on a fault-tolerant quantum computer, we first re-express it as a linear combination of
quantum overlaps:
\[
\sum_{x,\eta} \frac{1}{\sigma_{x,\eta}} \langle \psi_x | e^{i\eta} H O_x e^{-i\eta} H | \psi_x \rangle \\
\times \int_0^{t_f} ds \{ \langle \psi_x | e^{i\eta} H O_x e^{-i\eta} H V_n e^{-i\eta} H | \psi_x \rangle - \langle \psi_x | e^{i\eta} H V_n e^{-i\eta} H O_x e^{-i\eta} H | \psi_x \rangle \}
\]
\[
+ \sum_{x,\eta} \frac{1}{\sigma_{x,\eta}} \sum_m a_{x,m} \cos(t_f \omega_{x,m} + \phi_x)
\times \int_0^{t_f} ds \{ \langle \psi_x | e^{i\eta} H O_x e^{-i\eta} H V_n e^{-i\eta} H | \psi_x \rangle - \langle \psi_x | e^{i\eta} H V_n e^{-i\eta} H O_x e^{-i\eta} H | \psi_x \rangle \}.
\] (B3)

Then our goal is to find unitary operators SEL_a, quantum states |Ψ|^a = PREP_a|0⟩, and positive numbers λ_a > 0 for a = 0, 1, such that the |Ψ|^a SEL_a |Ψ|^a), compute the desired linear combinations up to scaled-down factors of λ_a, equivalently, we say that the target quantities are block encoded by |Ψ|^a and SEL_a with scaled-down factor λ_a. We describe how to construct such SEL_a and |Ψ|^a) in Sec. IV B.

We estimate the quantum overlap |Ψ⟩|SEL⟩|Ψ⟩ using the overlap estimation algorithm of Ref. [99]. Specifically, we consider the two reflections
\[
I - 2|Ψ⟩⟨Ψ|, \quad I - 2|SEL⟩|Ψ⟩|SEL⟩^\dagger.
\] (B4)

These reflections keep the two-dimensional subspace span(|Ψ⟩, SEL|Ψ⟩) invariant, on which their product
\[
(I - 2|Ψ⟩⟨Ψ|)(I - 2|SEL⟩|Ψ⟩|SEL⟩^\dagger)
\] has eigenvalues e^±i2 arcsin |⟨Ψ|SEL|Ψ⟩|. Therefore, we can perform quantum phase estimation on the above operator and take the cosine of the outcome to estimate the amplitude α ~ |⟨Ψ|SEL|Ψ⟩|. To further retrieve the phase, we introduce an ancilla qubit and estimate
\[
\beta_0 \sim |⟨+Ψ|c(SEL)+Ψ⟩| = \frac{|1 + ⟨Ψ|SEL|Ψ⟩|}{2},
\]
\[
\beta_{\pi/2} \sim |⟨+Ψ|c(\pi Z/4 ⊗ I)c(SEL)+Ψ⟩| = \frac{|1 - i⟨Ψ|SEL|Ψ⟩|}{2},
\] (B6)

where c(SEL) is the controlled operation
\[
c(SEL) := |0⟩|0⟩ ⊗ I + |1⟩|1⟩ ⊗ SEL.
\] (B7)

From these we obtain
\[
y = \frac{4\beta_0^2 - \alpha^2 - 1}{2} + i\frac{4\beta_{\pi/2}^2 - \alpha^2 - 1}{2} \sim ⟨Ψ|SEL|Ψ⟩.
\] (B8)

The standard quantum phase estimation outputs an estimate of the eigenphase with accuracy ε by making O(1/ε) queries to the reflections, succeeding with a constant probability greater than 1/2. The precision parameter ε directly translates to a maximal error of O(ε) in the estimated overlap. To succeed with a higher probability, we can repeat quantum phase estimation and take the median of the outcomes. By Hoeffding’s inequality, the success probability can be made arbitrarily close to one with only logarithmic overhead. We thus obtain the following result.

**Lemma 1 (Quantum overlap estimation):** Given quantum state |Ψ⟩ = PREP|0⟩, unitary SEL, ε > 0, and 0 < δ < 1, there exists a quantum algorithm with output y such that
\[
\mathbb{P}(|y - ⟨Ψ|SEL|Ψ⟩| ≥ ε) < δ.
\] (B9)

This algorithm makes O((log(1/δ)/ε) queries to PREP and SEL for their controlled version c-(PREP) and c-(SEL)), and uses O(N polylog(1/ε, 1/δ)) additional gates, where N is the number of qubits in the target system.

In the description of the above algorithm, we have ignored the normalization factor λ > 0 introduced by |Ψ⟩ = PREP|0⟩ and SEL. To get the target quantity, we need to multiply the outcome of quantum overlap estimation by λ. To ensure that the estimation succeeds with probability 1 − δ and accuracy ε, it then suffices to make O((log(1/δ)λ/ε) queries to PREP and SEL and use O(N polylog(λ, 1/ε, 1/δ)) additional gates.

It is instructive to compare the fault-tolerant approach with the sampling-based approach that is more suitable to implement on near-term quantum devices. That approach uses the generalized Hadamard test and produces an unbiased estimate of the real and imaginary parts of ⟨Ψ|SEL|Ψ⟩ with constant variance. By Hoeffding’s inequality, it suffices to take O((log(1/δ)/ε^2) samples to estimate with accuracy ε and probability 1 − δ. Each sample requires a constant number of queries to PREP and SEL. Therefore, we get a factor of Θ(λ/ε) saving by switching to the fault-tolerant quantum algorithm. See Secs. IV A and IV B for further discussions of these two approaches.

### 2. Truncation and discretization error

In this section, we analyze the error due to the truncation of real parameters and discretization of the time integral. We will see in Sec. IV B that this only introduces a logarithmic overhead in the overall cost.

We first consider discretizing the integral as
\[
\int_0^{t_f} ds f(s) \approx \frac{t_f}{L} \sum_{\ell=0}^{L-1} f\left(\frac{\ell}{L} t_f\right).
\] (B10)
where
\[ f(s) := \text{Trace}(O_x[e^{-i(t_j-s)H}V_n e^{i(t_j-s)H}], \]
\[ e^{-it_jH} |\psi_x\rangle\langle \psi_x| e^{it_jH}), \]
(B11)

This discretization error can be made arbitrarily small by choosing \( L \) sufficiently large. Here, we analyze how the error scales as a function of \( L \). Using the integral expansion
\[
\int_0^{t_j} ds f(s) - \frac{t_j}{L} \sum_{\ell=0}^{L-1} f\left(\frac{\ell}{L} t_j\right)
= \sum_{\ell=0}^{L-1} \int_{\ell t_j/L}^{t_j / L} ds \left(f(s) - f\left(\frac{\ell}{L} t_j\right)\right)
= \sum_{\ell=0}^{L-1} \int_{\ell t_j/L}^{t_j / L} ds \int_s^t d\tau f'(\tau),
\]
(B12)

we have
\[
\left|\int_0^{t_j} ds f(s) - \frac{t_j}{L} \sum_{\ell=0}^{L-1} f\left(\frac{\ell}{L} t_j\right)\right|
\leq \sum_{\ell=0}^{L-1} \int_{\ell t_j/L}^{t_j / L} ds \int_s^t d\tau f'(\tau) \max = \frac{t_j^2}{2L} \|f'\|_{\max},
\]
(B13)

where
\[
\|f'\|_{\max} := \max_{0 \leq s \leq t_j} |f'(s)|
\]
(B14)
is the largest derivative of \( f \) within the time interval \([0, t_j]\). The derivative \( f'(s) \) takes the form
\[
f'(s) := \text{Trace}(O_x[e^{-i(t_j-s)H}H, V_n e^{i(t_j-s)H}], \]
\[ e^{-it_jH} |\psi_x\rangle\langle \psi_x| e^{it_jH}), \]
(B15)

which gives
\[
\|f'\|_{\max} \leq 2 \|[H, V_n]\|.
\]
(B16)
The above discretization only achieves first-order accuracy, but one can improve this by switching to a higher-order scheme, which can reduce the cost of fault-tolerant implementation; see [7, Appendix H] for details.

In the following, we evaluate this bound for a model of clustered Hamiltonians acting on \( N \) sites:
\[
H := \sum_{K} H_K + \sum_{K \not\subset L} H_{K,L}
= \sum_{K} \sum_{k \not\subset K} H_{k,K} + \sum_{K \not\subset L} \sum_{k \not\subset K, l \in L} H_{k,l}.
\]
(B17)

Here, each term from the Hamiltonian acts on at most two sites and the sites are further grouped into clusters.

We use calligraphic capital letters such as \( K \) and \( L \) to denote the clusters, and use \( k \) to denote an arbitrary single site within \( K \). Assuming that Hamiltonian terms are normalized \( \|H_k\| \leq 1 \), we have
\[
\|f'\|_{\max} \leq 4 \Lambda_{ind}, \quad \Lambda_{ind} := \max_{\ell \in L} \sum_{K \not\subset L} \sum_{k \in K, l \in L} \|H_{K,l}\|.
\]
(B18)

Altogether, we have discretized Eq. (B1) with error at most
\[
\sum_{x,j} \frac{1}{\sigma_{x,j}^2} \left|\text{Trace}(O_x \tilde{\rho}_x(t_j)) - \sum_m a_{x,m} \cos(t_j \omega_{x,m} + \phi_x)\right|
\times \frac{2t_j^2 \Lambda_{ind}}{L} = O\left(\frac{\lambda \Lambda_{ind} T}{L}\right),
\]
(B19)

where
\[
\lambda := \lambda_0 + \lambda_1,
\]
\[
\lambda_0 := \sum_{x,j} \frac{t_j}{\sigma_{x,j}^2}, \quad \lambda_1 := \sum_{x,j,m} \frac{a_{x,m} t_j}{\sigma_{x,j}^2}, \quad T := \max_j t_j.
\]
(B20)

To achieve an accuracy of \( \epsilon \), it suffices to choose
\[
L = O\left(\frac{\lambda \Lambda_{ind} T}{\epsilon}\right).
\]
(B21)

We take \( T \) and \( L \) to be powers of 2 to simplify our circuit implementation.

We now consider the error due to the finite-digit truncation of the real parameters \( \sigma_{x,j} \), \( t_j \), \( a_{x,m} \), and \( \omega_{x,m} \). In general, the error in \( \sigma_{x,j} \) can be bounded under certain continuity assumptions with respect to the argument \( t_j \). Here, we take \( \sigma_{x,j} \equiv \sigma \) to be constant to simplify the analysis. In our circuit implementation, the evolution time will be loaded onto a quantum register using the QROM approach of Ref. [15] as
\[
|t_{log T-1} \cdots t_1 t_0\rangle = |t_{-1} T_2 \cdots t_{-log K}\rangle,
\]
(B22)

where we have again taken \( K \) to be a power of 2 to simplify the implementation. Here, \( T \) is the maximum possible time so log \( T \) bits suffice to represent the integer part of \( t \). The length of the decimal part should be chosen large enough to represent the time sufficiently accurate. Specifically, for
where \( |t' - t| \leq 1/K \), we have
\[
\|e^{-itH} - e^{-it'H}\| \leq |t' - t| \|H\| = O\left( \frac{N\Lambda_{\text{ind}}}{K} \right),
\] (B23)
which implies that
\[
\left| \int_0^t ds f(t',s) - \int_0^t ds f(t,s) \right| \\
\quad \leq \left| \int_0^t ds f(t',s) - \int_0^t ds f(t',s) \right| + \left| \int_0^t ds f(t',s) - \int_0^t ds f(t,s) \right| \\
\quad = O\left( \frac{N\Lambda_{\text{ind}}}{K} \right),
\] (B24)

\[
|g(t') - g(t)| \leq |t' - t| \max \left| \frac{dg}{d\tau} \right| = O\left( \frac{N\Lambda_{\text{ind}}}{K} \right).
\] (B25)

for
\[
f(t,s) := \text{Trace}(O_s [e^{-i(t-s)H} V_n e^{i(t-s)H}], e^{-itH} |\psi_s \rangle \langle \psi_s | e^{itH}]),
\] (B26)
\[
g(t) := \text{Trace}(O_s e^{-itH} |\psi_s \rangle \langle \psi_s | e^{itH}).
\] (B27)

Similarly, if \( t_j' \) and \( \omega_{x,m}' \) satisfy \( |t_j' - t_j| \leq 1/K \) and \( |\omega_{x,m}' - \omega_{x,m}| \leq 1/K \), then
\[
|\cos(t_j' \omega_{x,m}' + \phi_j) - \cos(t_j \omega_{x,m} + \phi_j)| \\
\quad \leq |t_j' \omega_{x,m}' - t_j \omega_{x,m}| = O\left( \frac{T + W}{K} \right),
\] (B28)

where \( T := \max t_j \) and \( W := \max \omega_{x,m} \). The coefficients in the Hamiltonian can be approximately prepared using the coherent alias sampling approach also described in Ref. [15]. Using that approach with \( \log K \) qubits for the inequality test, we have
\[
\left| \frac{t_j'}{\sigma^2} - \frac{t_j}{\lambda_0} \right| \leq \frac{1}{KN_d}, \\
\left| \frac{\omega_{x,m}'}{\lambda_1} - \frac{\omega_{x,m}}{\lambda_1} \right| \leq \frac{1}{KN_d N_{\omega} N_d}.
\] (B29)

The total truncation error can now be bounded by
\[
O\left( \frac{\lambda N \Lambda_{\text{ind}}}{K} + \frac{\lambda_0 N \Lambda_{\text{ind}}}{K} + \frac{\lambda_1 (T + W)}{K} + \frac{N_d \lambda_0}{KN_d} \\
+ \frac{N_{\omega} N_d \lambda_1}{KN_d N_{\omega} N_d} \right) = O\left( \frac{\lambda (N \Lambda_{\text{ind}} + T + W)}{K} \right).
\] (B30)

To ensure that this error is at most \( \epsilon \), it suffices to choose
\[
K = O(\text{poly}(\lambda, N, \Lambda_{\text{ind}}, T, W, 1/\epsilon)).
\] (B31)

**APPENDIX C: ALTERNATIVE DERIVATION OF HAMILTONIAN DERIVATIVES VIA OPTIMAL CONTROL THEORY**

To derive Eq. (5) through optimal control theory, we enforce the evolution of \( \rho_s(t) \) by \( H + H_s(t) \) variationally. We introduce an auxiliary field \( \kappa_s(t) \) as a Lagrange variable to enforce this condition, which transforms our cost function to
\[
C[\tilde{H}, \tilde{\rho}, \kappa] = \sum_n \left( \frac{\tilde{h}_n - h_n^{(0)}}{2\omega_n^2} \right)^2 \\
+ \sum_{x,t} \frac{1}{2\sigma_{x,t}^2} (\text{Trace}[\tilde{\rho}_x(t) O_x] - S_x(t))^2 \\
+ i \sum_t \int_0^\infty dt \text{Trace} \\
\left[ \tilde{\kappa}_x(t) \left( \frac{\partial \tilde{\rho}_x(t)}{\partial t} - i[\tilde{H} + H_x(t), \tilde{\rho}_x(t)] \right) \right].
\] (C1)

This is now a functional; in addition to the finite real values \( \tilde{h}_n \), it also takes as input any smooth matrix-valued functions \( \tilde{\rho}_x(t) \) and \( \tilde{\kappa}_x(t) \). This implies that all dependence of \( C \) on \( H \) is explicit; the dependence of \( \tilde{\rho}_x(t) \) [and \( \tilde{\kappa}_x(t) \)] will emerge by the principle of least action. The solution to our problem is given again by the minimum of the functional \( C \). Taking a functional derivative \( \delta C/\delta \tilde{\kappa}_x(t) = 0 \) yields the Schrödinger equation in its standard form
\[
\frac{\delta C}{\delta \tilde{\kappa}_x(t)} = 0 \rightarrow \frac{\partial \tilde{\rho}_x(t)}{\partial t} = i[\tilde{H} + H_x(t), \tilde{\rho}_x(t)],
\] (C2)

which yields the solution in the main text: \( \tilde{\rho}_x(t) = \tilde{U}_x(t,0) \rho_x \tilde{U}_x(t,0) \), Taking the functional derivative with respect to \( \tilde{\rho}_x(t) \) and setting this equal to 0 yields an update rule for \( \tilde{\kappa}_x(t) \):
\[
\frac{\partial \tilde{\kappa}_x(t)}{\partial t} = -i[\tilde{\kappa}_x(t),H] - i \sum_{x',t'} \frac{1}{\sigma_{x,t}^2} \delta(t - t') \\
\times (\text{Trace}[\tilde{\rho}_{x'}(t') O_{x'}] - S_{x'}(t')) O_{x}, \quad \kappa_x(+\infty) = 0.
\] (C3)

This takes the form of an external field \( \tilde{\kappa}_x(t) \) that propagates back in time and is perturbed in a nonunitary way by each measurement \( S_x(t') \) that does not completely match the predicted \( \tilde{S}_x(t') \). Substituting in the solution for \( \tilde{\rho}_x(t) \).
yields a solution for $\tilde{\kappa}_x(t)$:
\[
\kappa_x(t) = i \sum_{t > t'} \frac{1}{\sigma^2} \tilde{U}_{x,t,t'}(t) O_x \tilde{U}_{x,t,t'}(t')
\times (\text{Trace}[\hat{\rho}_x(t) O_x] - S_x(t)). \tag{C4}
\]

To recover the update rule, we then take the partial derivative of $C$ with respect to the parameters $h_n$ and set this to zero:
\[
\frac{\partial C}{\partial h_n} = \hat{h}_n - h_n^{(0)} + \sum_x \int_0^\infty dt \text{Trace}[\tilde{\kappa}_x(t) [V_n, \hat{\rho}_n(t)]].
\tag{C5}
\]

Substituting in Eq. (C4) yields Eq. (5) as required. Following a similar procedure to take second-order derivatives of $C$ with respect to $\hat{\rho}_n(t)$ and $\tilde{\kappa}_x(t)$ yields Eq. (6) after some rearrangement.

---


Learning models of quantum systems from experiments, 

[46] Rolando D. Somma and Sergio Boixo, Parameter estimation with mixed-state quantum computation, 

[47] Ashok Ajoy and Paola Cappellaro, Quantum Simulation via Filtered Hamiltonian Engineering: Application to Perfect Quantum Transport in Spin Networks, 

[48] Sheng-Tao Wang, Dong-Ling Deng, and Lu-Ming Duan, 
Hamiltonian tomography for quantum many-body systems with arbitrary couplings, 

[49] Eyal Bairey, Itai Arad, and Netanel H. Lindner, 
Learning a Local Hamiltonian from Local Measurements, 

[50] Tim J. Evans, Robin Harper, and Steven T. Flammia, 
Scalable Bayesian Hamiltonian learning, 

[51] Mark Srednicki, The approach to thermal equilibrium in quantized chaotic systems, 

[52] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Metal-insulator transition in a weakly interacting many-electron system with localized single-particle states, 

[53] Dmitry A. Abanin, Ehud Altman, Immanuel Bloch, and Maksym Serbyn, Many-body localization, 

[54] Maksym Serbyn, Zlatko Papić, and Dmitry A. Abanin, 
Local Conservation Laws and the Structure of the Many-Body Localized States, 

[55] Bela Bauer and Chetan Nayak, 
Area laws in a many-body localized state and its implications for topological order, 

[56] M. Pino, M. Srednicki, Non-ergodic metallic and insulating phases of Josephson junction chains, 

[57] B. L. Altshuler, E. Cuevas, L. B. Ioffe, and V. E. Kravtsov, Non-Ergodic Phases in Strongly Disordered Random Regular Graphs, 

[58] B. L. Altshuler, L. B. Ioffe, and V. E. Kravtsov, Non-ergodic delocalized phases on Bethe lattice and random regular graphs, 

[59] L. Faoro, M. V. Feigelman, and L. Ioffe, Non-ergodic extended phase of the quantum random energy model, 

[60] Nicolas Macé, Fabien Alet, and Nicolas Laflorencie, Multifractal Scalings Across the Many-Body Localization Transition, 

[61] E. Knill and R. Laflamme, Power of One Bit of Quantum Information, 

[62] Dries Sels, Hsem Dashi, Samia Mora, Olga Demler, and Eugene Demler, Quantum approximate Bayesian computation for NMR model inference, 

[63] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and D. G. Cory, Quantum Hamiltonian learning using imperfect quantum resources, 


